



# Estimating functions for noisy observations of ergodic diffusions

Benjamin Favetto

► **To cite this version:**

Benjamin Favetto. Estimating functions for noisy observations of ergodic diffusions. MAP5 2010-30. 2010. <hal-00531096v2>

**HAL Id: hal-00531096**

**<https://hal.archives-ouvertes.fr/hal-00531096v2>**

Submitted on 16 Nov 2010

**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.

# Estimating functions for noisy observations of ergodic diffusions

Benjamin FAVETTO

November 16, 2010

## Abstract

In this article, general estimating functions for ergodic diffusions sampled at high frequency with noisy observations are presented. The theory is formulated in term of approximate martingale estimating functions based on local means of the observations, and simple conditions are given for rate optimality. The estimation of diffusion parameter is faster than the estimation of drift parameter, and the rate of convergence in the Central Limit Theorem is classical for the drift parameter but not classical for the diffusion parameter. The link with specific minimum contrast estimators is established, as an example.

**Key Words:** estimating functions, diffusion process, parametric inference, discrete time noisy observations, central limit theorem

## 1 Introduction

The aim of this article is to study estimating functions based on the observations of a noisy discretely observed one-dimensional diffusion inspired by Sørensen (2009). A one-dimensional diffusion process  $(X_t)$  is considered, solution of the stochastic differential equation

$$dX_t = b(X_t, \kappa)dt + \sigma(X_t, \lambda)dB_t, \quad X_0 = \eta \quad (1)$$

where  $(B_t)_{t \geq 0}$  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 sake of simplicity, in the proofs, we assume that  $d_1 = d_2 = 1$ . The true value of the parameter is denoted by  $\theta_0 = (\kappa_0, \lambda_0)$  and  $\theta_0 \in \overset{\circ}{\Theta}$ , where  $\Theta = \Theta_1 \times \Theta_2$  a product of compact subsets. The expectation under the probability distribution  $\mathbb{P}_\theta$  is denoted  $\mathbb{E}_\theta$ , for  $\theta \in \Theta$ , and  $\mathbb{E}_{\theta_0}$  for the expectation under the probability distribution  $\mathbb{P}_{\theta_0}$ .

At time  $t_i = i\delta_N, i = 0, \dots, N$ , with  $\delta_N$  the sampling time, the observation  $Y_{i\delta_N}$  is given by

$$Y_{i\delta_N} = X_{i\delta_N} + \rho_N \varepsilon_{i\delta_N} \quad (2)$$

where  $\rho_N$  is the standard deviation of the observation noise, and  $(\varepsilon_{i\delta_N})$  is a sequence of independent and identically distributed centered random variables, independent of the diffusion  $(X_t)$ , with unitary variance. Two cases are considered for the standard deviation of the observation noise: in the first case, denoted as **(B1)**,  $\rho_N = \rho > 0$  is constant, whereas in the second case **(B2)**,  $\rho_N \rightarrow 0$  as  $N \rightarrow \infty$ . Moreover, in the sequel,  $\rho_N$  is assumed to be known.

In Favetto (2010), minimum contrast estimators based on local means of the  $(Y_{i\delta_N})$  have been studied: dividing the  $N$  observations into  $k_N$  blocks of  $p_N$  observations, such that  $N = p_N k_N$ , and setting  $\Delta_N = p_N \delta_N$ , the local mean on the  $j$ th block (starting at time  $j\Delta_N$  and finishing at time  $(j+1)\Delta_N - \delta_N = j\Delta_N + (p_N - 1)\delta_N$ ) is

$$Y_{\bullet}^j = \frac{1}{p_N} \sum_{i=0}^{p_N-1} Y_{j\Delta_N + i\delta_N},$$

with the asymptotic framework:  $N \rightarrow \infty$ ,  $\delta_N \rightarrow 0$ ,  $t_N = N\delta_N = k_N\Delta_N \rightarrow \infty$  (high frequency sampling.) Moreover, the discretization step size and the number of observations in one block are linked by  $\delta_N = p_N^{-\alpha}$ , where  $\alpha \in (1, 2]$  is the *local mean size parameter*, which has to be tuned in practice. The case  $\alpha = 2$  is specific because the variance  $\rho_N^2$  has to be taken into account in the results (see Favetto (2010) for a discussion on  $\alpha$  based on simulations).

Our focus is on approximate martingale estimating functions based on the  $(Y_{\bullet}^j)$ . An exhaustive review on estimating functions for diffusion processes can be found in Sørensen (2010). The estimating functions  $G_{N,\alpha}$  considered here depend on  $\alpha \in (1, 2]$  and they are given by

$$G_{N,\alpha}(\theta) = \sum_{j=1}^{k_N-2} g_{\alpha}(\delta_N, Y_{\bullet}^{j+1} - Y_{\bullet}^j, Y_{\bullet}^{j-1}; \theta, \rho_N) \quad (3)$$

where the function  $g_{\alpha}(\delta, y, x; \theta, \rho)$ , valued in  $\mathbb{R}^2$ , is such that  $G_{N,\alpha}$  is approximately a martingale estimating function. Hence, the aim of this paper is to prove that the estimator  $\hat{\theta}_N$  given as the solution of  $G_{N,\alpha}(\hat{\theta}_N) = 0$  is consistent and asymptotically Gaussian. Notice the lag in (3) that must be introduced as  $(Y_{\bullet}^j)$  is not Markov.

The case of martingale estimating functions for discrete observations of diffusion processes has been treated in Bibby and Sørensen (1995), and the case of estimating functions that do not have the martingale property has been treated in Kessler (2000), with a fixed sampling time  $\Delta$ . Kessler and Sørensen (1999) introduce martingale estimating functions based on the eigenfunctions of the infinitesimal generator of the diffusion. Sørensen (2009) focuses on the high-frequency asymptotics for an ergodic diffusion.

The article is organized as follows: Section 2 is devoted to the presentation of the model and the assumptions, which are closed to those presented in

Favetto (2010). Section 3 contains the main results for the convergence in distribution of the variation and the quadratic variation of the local means, and the result of asymptotic normality for the estimators associated to the estimating functions. Section 4 is devoted to the examples, some concluding remarks are given in Section 5 and the proofs are gathered in Section 6.

## 2 Model and Assumptions

Consider the one-dimensional stochastic differential equation

$$dX_t = b(X_t, \kappa_0)dt + \sigma(X_t, \lambda_0)dB_t, \quad X_0 = \eta \quad (4)$$

where  $B$  is a standard Brownian motion,  $\theta_0 = (\kappa_0, \lambda_0)$  is the true value of the parameter, and  $\eta$  is a real valued random variable independent of  $B$ .

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 (4) 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  $x \in \mathbb{R}$ :

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

**(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  $\int_{-\infty}^{\infty} s(x)dx = \int_{-\infty}^{+\infty} s(x)dx = \infty$  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 (4) admits a unique strong solution on  $\mathbb{R}$ . Under **(A1)** and **(A2)**,  $\nu_0$  is the unique invariant probability of (4) and  $(X_t)$  is ergodic.

Furthermore, Assumptions **(A1)**-**(A3)** imply **(A4)** if  $\eta$  has distribution  $\nu_0$  or  $\eta$  is deterministic (for the latter case, see Gloter (2006), Proposition 3).

Conditions on the function  $g_\alpha$  for optimality are stated now. The two components of  $g_\alpha$  are denote by  $(g_{1,\alpha}, g_{2,\alpha})^T$ , and we assume that  $g_\alpha(\delta, y, x; \theta, \rho)$

satisfies the conditions of Sørensen (2009) recalled below. First, the condition for *rate optimality* is

$$\partial_y g_{2,\alpha}(0, 0, x; \theta, \rho) = 0 \quad (5)$$

for all  $x \in \mathbb{R}$ , all  $\rho > 0$  and all  $\theta \in \Theta$ , where  $\partial_y g_{2,\alpha}(0, 0, x; \theta, \rho) = 0$  means  $\partial_y g_{2,\alpha}(0, y, x; \theta, \rho) = 0$  evaluated at  $y = 0$ . This condition is also called *Jacobsen's condition* in Sørensen (2009), and corresponds to one of the conditions obtained in Jacobsen (2002).

For directly observed diffusion models, rate optimality is important because the diffusion coefficient parameter can be estimated at a higher rate than the drift parameter, and this result remains true for a diffusion observed with a noise.

Setting  $c(x, \lambda) = \sigma(x, \lambda)^2$ , the *second Jacobsen condition* is

$$\partial_y g_{1,\alpha}(0, 0, x; \theta, \rho) = \frac{\partial_\kappa b(x, \kappa)}{c(x, \lambda)} \quad \text{and} \quad \partial_{y^2}^2 g_{2,\alpha}(0, 0, x; \theta, \rho) = \frac{\partial_\lambda c(x, \lambda)}{c(x, \lambda)^2} \quad (6)$$

for all  $x \in \mathbb{R}$ , all  $\rho > 0$  and all  $\theta \in \Theta$ . In Sørensen (2009), this condition ensures the efficiency of the estimators in the case of direct observations of an ergodic diffusion.

The class  $\mathcal{C}_{p_1, p_2, p_3}(\mathbb{R}_+ \times \mathbb{R}^2 \times \Theta \times \mathbb{R}_+)$  is defined as the set of real functions  $f(t, y, x; \theta, \rho)$  satisfying

1.  $f$  is  $p_1$  times continuously differentiable in  $t$ ,  $p_2$  times continuously differentiable in  $y$  and  $p_3$  continuously differentiable in  $\kappa$  and  $\lambda$  ;
2.  $f$  and all partial derivatives of  $f$  are of polynomial growth uniformly for  $\theta \in \Theta$ .

The class  $\mathcal{C}_{p_1, p_2, p_3}(\mathbb{R}^2 \times \Theta \times \mathbb{R}_+)$  is defined in an similar way for functions  $f(y, x; \theta, \rho)$ . A function  $f(y, x; \theta, \rho)$  is said to be of polynomial growth in  $y$  and  $x$  uniformly for  $\theta \in \Theta$  (recall that  $\Theta$  is assumed to be a compact subset of  $\mathbb{R}^2$ ) if there exists a constant  $c > 0$  such that, for all  $x, y \in \mathbb{R}$  and all  $\rho > 0$ ,

$$\sup_{\theta \in \Theta} |f(y, x, ; \theta, \rho)| \leq c(1 + |x|^c + |y|^c).$$

From now on,  $R(\delta, y, x; \theta, \rho)$  denotes a generic function such that

$$|R(\delta, y, x; \theta, \rho)| \leq F(y, x; \theta, \rho)$$

with  $F$  of polynomial growth in  $y, x$  and  $\rho$ , uniformly in  $\theta$ . The function  $g_\alpha$  is assumed to belong to the class  $\mathcal{C}_{1,3,2}$  in the sequel.

Let  $\mathcal{H}_j^N$  be the  $\sigma$ -field defined by  $\mathcal{H}_j^N = \sigma(X_s, 0 \leq s \leq j\Delta_N) \vee \sigma(\varepsilon_{i\delta_N}, i\delta_N \leq (j-1)\Delta_N + (p_N-1)\delta_N)$ , such that  $Y_\bullet^j$  is  $\mathcal{H}_{j+1}^N$  measurable.

The function  $g(\delta, y, x; \theta, \rho)$  is assumed to satisfy the following condition **(D)**

1. For all  $\theta \in \Theta$ ,

$$\begin{aligned}\mathbb{E}_\theta(g(\delta_N, Y_\bullet^{j+1} - Y_\bullet^j, Y_\bullet^{j-1}; \theta, \rho_N) | \mathcal{H}_j^N) &= \delta_N^{2-\frac{2}{\alpha}} R(\delta_N, Y_\bullet^{j-1}, X_{j\Delta_N}; \theta, \rho_N) \\ &= \Delta_N^2 R(\delta_N, Y_\bullet^{j-1}, X_{j\Delta_N}; \theta, \rho_N),\end{aligned}$$

2. The function  $g(\delta, y, x; \theta, \rho)$  has an expansion in power of  $\delta$ : there exist functions  $g^{(1)}$  and  $g^{(\alpha)}$  such that

(a) if  $\alpha \in (1, \frac{3}{2}]$  or  $\alpha = 2$ , the expansion is

$$g(\delta, y, x; \theta, \rho) = g(0, y, x; \theta, \rho) + \delta^{1-\frac{1}{\alpha}} g^{(1)}(y, x; \theta, \rho) + \delta^{2-\frac{2}{\alpha}} R(\delta, y, x; \theta, \rho);$$

(b) if  $\alpha \in (\frac{3}{2}, 2)$ , the expansion is

$$\begin{aligned}g(\delta, y, x; \theta, \rho) &= g(0, y, x; \theta, \rho) + \delta^{1-\frac{1}{\alpha}} g^{(1)}(y, x; \theta, \rho) \\ &\quad + \delta^{\frac{1}{\alpha}} g^{(\alpha)}(y, x; \theta, \rho) + \delta^{2-\frac{2}{\alpha}} R(\delta, y, x; \theta, \rho)\end{aligned}$$

with  $R(\Delta, y, x; \theta, \rho)$  a generic remainder dominated by a function of polynomial growth in  $y$  and  $x$ , uniformly in  $\theta$ . (Remind that  $\delta_N = p_N^{-\alpha}$ , hence  $\Delta_N = p_N \delta_N = p_N^{1-\alpha} = \delta_N^{1-\frac{1}{\alpha}}$ , for a given  $\alpha \in (1, 2]$ .)

The condition on the expansion has to be precised: for  $\alpha \in (1, \frac{3}{2}]$ , the inequality  $\frac{1}{\alpha} \geq 2 - \frac{2}{\alpha}$  holds, and for  $\alpha = 2$ , we have  $1 - \frac{1}{\alpha} = \frac{1}{\alpha}$ , whereas for  $\alpha \in (\frac{3}{2}, 2)$ , the inequality  $2 - \frac{2}{\alpha} < \frac{1}{\alpha} < 1 - \frac{1}{\alpha}$  holds. Notice that  $\delta_N^{1-\frac{1}{\alpha}} = \Delta_N$  and  $\delta_N^{\frac{1}{\alpha}} = \Delta_N^{\frac{1}{\alpha-1}}$ , this term is needed to take into account the noise.

Finally, for any non-singular matrix  $M_N$ , the estimating functions  $G_{N,\alpha}$  and  $M_N G_{N,\alpha}$  give the same estimator, and the matrix  $M_N$  may depends on  $\delta_N$ . Therefore, a given version of an estimating function may not satisfy the above condition, but the point is that one version, up to a matrix  $M_N$  must exist and satisfy this condition. For example, it may be necessary to multiply one of the coordinates by  $\Delta_N = \delta_N^{1-\frac{1}{\alpha}}$ .

### 3 Rate-optimal estimating functions for local means

In this section, asymptotic results are proved for approximate martingale estimating functions based on local means. In particular, it is shown that the estimator of the parameter in the diffusion coefficient converges in distribution faster than the estimator of the parameter in the drift coefficient.

The infinitesimal generator of the diffusion  $(X_t)$  is

$$L_\theta = b(x, \kappa) \frac{d}{dx} + \frac{1}{2} c(x, \lambda) \frac{d^2}{dx^2}.$$

For a function  $h(y, x)$  of two variables,  $L_\theta h$  is defined by

$$L_\theta(h(\delta, \theta', \rho))(y, x) = b(x, \kappa)\partial_y h(\delta, y, x; \theta', \rho) + \frac{1}{2}c(x, \lambda)\partial_{y^2}^2 h(\delta, y, x; \theta', \rho).$$

The *modified* generator is also introduced as

$$\begin{aligned}\bar{L}_\theta &= b(x, \kappa)\frac{d}{dx} + \frac{1}{3}c(x, \lambda)\frac{d^2}{dx^2} \\ &= L_\theta - C_\theta\end{aligned}$$

with  $C_\theta = \frac{1}{6}c(x, \lambda)\frac{d^2}{dx^2}$ . Indeed, it comes from Proposition 2 (proved in Favetto (2010) and recalled in the Appendix) that, for  $f$  a twice continuously differentiable real function with bounded second derivative, with  $\Delta_N = p_N \delta_N = \delta_N^{1-\frac{1}{\alpha}}$ ,

$$\begin{aligned}\mathbb{E}_{\theta_0}(f(Y_\bullet^{j+1} - Y_\bullet^j) | \mathcal{H}_j^N) &= f(0) + f'(0)\mathbb{E}_{\theta_0}(Y_\bullet^{j+1} - Y_\bullet^j | \mathcal{H}_j^N) \\ &\quad + \frac{1}{2}f''(0)\mathbb{E}_{\theta_0}((Y_\bullet^{j+1} - Y_\bullet^j)^2 | \mathcal{H}_j^N) + \Delta_N o_P(1) \\ &= f(0) + \Delta_N(f'(0)b(X_{j\Delta_N}, \kappa_0) + \frac{1}{3}f''(0)c(X_{j\Delta_N}, \lambda_0)) \\ &\quad + \Delta_N^{\frac{1}{\alpha-1}}f''(0)\rho_N^2 + \Delta_N o_P(1)\end{aligned}$$

by Taylor's formula.

The following lemma provides an useful identity for the next computations.

**Lemma 1.** *Under Condition (D), for all  $x \in \mathbb{R}$ , all  $\theta \in \Theta$  and all  $\rho > 0$*

$$\begin{aligned}g(0, 0, x; \theta, \rho) &= 0, \\ g^{(1)}(0, x; \theta, \rho) &= -\bar{L}_\theta(g(0, \theta, \rho))(0, x) \text{ if } \alpha \in (1, 2), \\ g^{(\alpha)}(0, x; \theta, \rho) &= -\rho^2 \partial_{y^2}^2 g(0, 0, x; \theta, \rho) \text{ if } \frac{3}{2} < \alpha < 2, \\ g^{(1)}(0, x; \theta, \rho) &= -\bar{L}_\theta(g(0, \theta, \rho))(0, x) - \rho^2 \partial_{y^2}^2 g(0, 0, x, \theta, \rho) \text{ if } \alpha = 2.\end{aligned}$$

**Remark.** This lemma has to be compared with the result in Sørensen (2009), when the diffusion is directly observed. For  $\alpha \in (1, 2)$ , the result on  $g^{(1)}$  is the same except that it involves  $\bar{L}_\theta$  in place of  $L_\theta$ . For  $\alpha \in (\frac{3}{2}, 2)$ , the rate of sampling has to be taken into account, by the additional term  $g^{(\alpha)}$  which contains the variance  $\rho^2$  of the observation noise. In the case  $\alpha = 2$ , the time-step of a block of observations is  $\Delta_N = p_N^{-1}$  and an additional term is needed in  $g^{(1)}$ .

Setting  $\rho_\infty = \lim_{N \rightarrow \infty} \rho_N$ , two cases are distinguished for the limit theorems: for the first one, Assumption (B1) holds, or Assumption (B2) when  $\alpha \in (1, 2)$ , or  $\alpha = 2$  and  $\rho_\infty = 0$ , and in this case, the variance of the noise

does not appear in the limit theorems. For the second case, Assumption **(B1)** holds with  $\alpha = 2$ , and  $\rho_\infty^2 = \rho^2 > 0$  appears in the limit theorems.

Let  $\gamma(\theta, \theta_0, \rho_\infty)$  be defined as

$$\begin{aligned} \gamma(\theta, \theta_0, \rho_\infty) &= \nu_0((b(\cdot, \kappa_0) - b(\cdot, \kappa))\partial_y g(0, 0, \cdot; \theta, \rho_\infty)) \\ &\quad + \frac{1}{3}\nu_0((c(\cdot, \lambda_0) - c(\cdot, \lambda))\partial_{y^2}^2 g(0, 0, \cdot; \theta, \rho_\infty)), \end{aligned} \quad (7)$$

and the matrix

$$\bar{A}_\theta(\cdot, \rho) = \begin{pmatrix} \partial_\kappa b(\cdot, \kappa)\partial_y g_1(0, 0, \cdot; \theta, \rho) & \frac{1}{3}\partial_\lambda c(\cdot, \lambda)\partial_{y^2}^2 g_1(0, 0, \cdot; \theta, \rho) \\ \partial_\kappa b(\cdot, \kappa)\partial_y g_2(0, 0, \cdot; \theta, \rho) & \frac{1}{3}\partial_\lambda c(\cdot, \lambda)\partial_{y^2}^2 g_2(0, 0, \cdot; \theta, \rho) \end{pmatrix}.$$

Notice that  $\bar{A}_\theta(x, \rho) = \partial_\theta \bar{L}_\theta(g(0, \theta, \rho))(0, x) = \partial_\theta L_\theta(g(0, \theta, \rho))(0, x) - \partial_\theta C_\theta(g(0, \theta, \rho))(0, x)$ . Setting  $S = \nu_0(A_{\theta_0}(\cdot, \rho_\infty))$ , the following identity holds

$$\begin{aligned} \partial_\theta g^{(1)}(0, x; \theta, \rho) &= -\partial_\theta L_\theta(g(0, \theta, \rho))(0, x) - L_\theta(\partial_\theta g(0, \theta, \rho))(0, x) \\ &\quad + \partial_\theta C_\theta(g(0, \theta, \rho))(0, x) + C_\theta(\partial_\theta g(0, \theta, \rho))(0, x). \end{aligned}$$

With  $\phi(\theta, \theta_0, \rho_\infty)$  defined by

$$\phi(\theta, \theta_0, \rho_\infty) = \nu_0(\bar{L}_{\theta_0}(\partial_\theta g(0; \theta, \rho_\infty))(0, \cdot) - \bar{L}_\theta(\partial_\theta g(0; \theta, \rho_\infty))(0, \cdot)) - \nu_0(\bar{A}_\theta(\cdot, \rho_\infty)),$$

and  $\partial_\theta g$  the matrix of partial derivatives with respect to  $\kappa$  and  $\lambda$ , the following lemma contains the main statements for convergence in probability results (under  $\mathbb{P}_{\theta_0}$ ).

**Lemma 2.** *Under Assumptions **(A1)**-**(A5)**, for a local mean size parameter  $\alpha \in (1, 2]$  and Assumption **(B1)** or **(B2)**,*

$$\frac{1}{k_N \Delta_N} \sum_{j=1}^{k_N-2} g(\Delta_N, Y_\bullet^{j+1} - Y_\bullet^j, Y_\bullet^{j-1}; \theta, \rho_N) \xrightarrow{\mathbb{P}_{\theta_0}} \gamma(\theta, \theta_0, \rho_\infty), \quad (8)$$

$$\frac{1}{k_N \Delta_N} \sum_{j=1}^{k_N-2} \partial_\theta g(\Delta_N, Y_\bullet^{j+1} - Y_\bullet^j, Y_\bullet^{j-1}; \theta, \rho_N) \xrightarrow{\mathbb{P}_{\theta_0}} \phi(\theta, \theta_0, \rho_\infty) \quad (9)$$

uniformly in  $\theta \in \Theta$ .

Recall that, with  $b(\cdot) = b(\cdot, \kappa_0)$ , for a function  $f = f(\cdot, \theta_0)$ , the empirical mean  $\bar{\nu}_N(f)$ , the variation  $\bar{I}_N(f)$  and the quadratic variation  $\bar{Q}_N(f)$  of the local means of the observations are defined by

$$\begin{aligned} \bar{\nu}_N(f) &= \frac{1}{k_N} \sum_{j=0}^{k_N-1} f(Y_\bullet^j), \\ \bar{I}_N(f) &= \frac{1}{k_N \Delta_N} \sum_{j=1}^{k_N-2} f(Y_\bullet^{j-1})(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1})), \\ \bar{Q}_N(f) &= \frac{1}{k_N \Delta_N} \sum_{j=1}^{k_N-2} f(Y_\bullet^{j-1})(Y_\bullet^{j+1} - Y_\bullet^j)^2. \end{aligned}$$



The convergence in distribution of the functionals  $\bar{I}_N$  and  $\bar{Q}_N$ , which are useful to study the asymptotic normality of the estimator  $\hat{\theta}_N$ , is now precised.

**Theorem 1.** *Assume (A1)-(A5) and (B1) or (B2). If  $\alpha \in (1, 2]$  and  $N\delta_N^{3-\frac{2}{\alpha}} \rightarrow 0$ , then*

$$\sqrt{N\delta_N}\bar{I}_N(f) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu_0(f^2\sigma^2)).$$

A comment has to be done about the assumption  $N\delta_N^{3-\frac{2}{\alpha}} \rightarrow 0$ : when the diffusion is directly observed at time  $i\delta_N$ , a similar result holds under the condition  $N\delta_N^2 \rightarrow 0$ . Here, the condition  $k_N\Delta_N^3 \rightarrow 0$  is needed, and  $k_N\Delta_N^3 = N\delta_N^{3-\frac{2}{\alpha}}$ . When  $\alpha = 2$ , the usual condition  $N\delta_N^2 \rightarrow 0$  is obtained, but when  $\alpha \in (1, 2)$ , a stronger assumption is needed on the sampling rate than if the discretized diffusion process were observed directly.

The convergence in distribution for the quadratic variation is precised below.

**Theorem 2.** *Assume (A1)-(A5), and  $N\delta_N^{2-\frac{1}{\alpha}} \rightarrow 0$ .*

- *If (B1) or (B2) with  $\alpha \in (1, 2)$  or  $\alpha = 2$  and (B2) ( $\rho_N \rightarrow 0$ ), then*

$$\sqrt{N\delta_N^{\frac{1}{\alpha}}}(\bar{Q}_N(f) - \bar{\nu}_N(f(\frac{2}{3}\sigma^2 + 2\rho_N^2\Delta_N^{\frac{2-\alpha}{\alpha-1}}))) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu_0(f^2\sigma^4));$$

- *If  $\alpha = 2$  and (B1) ( $\rho_N = \rho$ ), then*

$$\sqrt{N\delta_N^{\frac{1}{\alpha}}}\left(\bar{Q}_N(f) - \bar{\nu}_N\left(f\left(\frac{2}{3}\sigma^2 + 2\rho^2\right)\right)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu_0(f^2(\sigma^4 + 4\sigma^2\rho^2 + 12\rho^4))).$$

In this second theorem, the condition  $N\delta_N^{2-\frac{1}{\alpha}} = k_N\Delta_N^2 \rightarrow 0$  is needed. This is more stringent than  $N\delta_N^{2-\frac{2}{\alpha}} \rightarrow 0$ :  $N\delta_N^{\frac{3}{2}} \rightarrow 0$  is needed in the case  $\alpha = 2$ , for example. Moreover, a distinction has to be done in the case  $\alpha = 2$ ,  $\rho_N = \rho$ , because the variance of the observation noise  $\rho^2$  appears in the asymptotic variance, which is increased significantly. Notice also that the rate  $k_N = N\delta_N^{\frac{1}{\alpha}}$  is not classical : when the diffusion is directly observed, the usual rate of convergence for a similar central limit theorem is  $N$ , whereas here it is slower. That explains why the condition  $N\delta_N^{2-\frac{1}{\alpha}}$  is more stringent in the case of diffusions observed with noise.

And there is the multivariate version of the two last theorems.

**Theorem 3.** *Assume (A1)-(A5), and  $N\delta_N^{2-\frac{1}{\alpha}} \rightarrow 0$ . Then, for  $f$  a function which satisfies (C1), and  $\alpha \in (1, 2]$ ,*

$$\left( \begin{array}{c} \sqrt{N\delta_N} \bar{I}_N(f) \\ \sqrt{N\delta_N^{\frac{1}{\alpha}}} \left( \bar{Q}_N(f) - \bar{\nu}_N \left( f \left( \frac{2}{3}\sigma^2 + 2\Delta_N^{\frac{2-\alpha}{\alpha-1}} \rho_N^2 \right) \right) \right) \end{array} \right) \xrightarrow{\mathcal{L}} \mathcal{N}_2 \left( \left( \begin{array}{c} 0 \\ 0 \end{array} \right), \left( \begin{array}{cc} W_1(f) & 0 \\ 0 & W_2(f) \end{array} \right) \right) \quad (10)$$

with, in the case  $\alpha \in (1, 2)$  and **(B1)** or **(B2)**, or in the case  $\alpha = 2$  with **(B2)**,

$$W_1(f) = \nu_0(f^2\sigma^2) \text{ and } W_2(f) = \nu_0(f^2\sigma^4)$$

and in the case  $\alpha = 2$ , with **(B1)**,

$$W_1(f) = \nu_0(f^2\sigma^2) \text{ and } W_2(f) = \nu_0(f^2(\sigma^4 + 4\rho^2\sigma^2 + 12\rho^4)).$$

For the applications, the following corollary is needed.

**Corollary 1.** Assume **(A1)**-**(A5)**,  $N\delta_N^{2-\frac{1}{\alpha}} \rightarrow 0$  for an  $\alpha \in (1, 2]$ . Consider a sequence of functions  $f_N(x, \theta)$  and a function  $f(x, \theta)$  satisfying **(C1)**, a sequence  $v_N \rightarrow 0$  such that, for all  $x$ ,

$$\sup_{\theta \in \Theta} |f_N(x, \theta) - f(x, \theta)| \leq v_N(1 + |x|),$$

then

$$\left( \begin{array}{c} \sqrt{N\delta_N} \bar{I}_N(f_N) \\ \sqrt{N\delta_N^{\frac{1}{\alpha}}} \left( \bar{Q}_N(f_N) - \bar{\nu}_N \left( f_N \left( \frac{2}{3}\sigma^2 + 2\Delta_N^{\frac{2-\alpha}{\alpha-1}} \rho_N^2 \right) \right) \right) \end{array} \right) \xrightarrow{\mathcal{L}} \mathcal{N}_2 \left( \left( \begin{array}{c} 0 \\ 0 \end{array} \right), \left( \begin{array}{cc} W_1(f) & 0 \\ 0 & W_2(f) \end{array} \right) \right). \quad (11)$$

The main theorem about the estimation of  $\theta_0 = (\kappa_0, \lambda_0)$  and the asymptotic behaviour of the estimator  $\hat{\theta}_N = (\hat{\kappa}_N, \hat{\lambda}_N)$  is derived from the latter results.

With the *rate optimality* condition and the *second Jacobsen's condition*, the Central Limit Theorem gives the optimal rate of convergence for the estimator of the parameter involved in the diffusion coefficient.

**Theorem 4.** Assume **(A1)**-**(A5)**, **(D)**, and that the following *identifiability condition*

$$\begin{cases} \nu_0((b(\cdot, \kappa_0) - b(\cdot, \kappa))\partial_y g_1(0, 0, \cdot; \theta, \rho_\infty)) \neq 0 & \text{for } \kappa \neq \kappa_0 \\ \nu_0((c(\cdot, \lambda_0) - c(\cdot, \lambda))\partial_{y^2}^2 g_2(0, 0, \cdot; \theta, \rho_\infty)) \neq 0 & \text{for } \lambda \neq \lambda_0 \end{cases}$$

is satisfied. Define  $S = \nu_0(A_{\theta_0}(\cdot, \rho_\infty))$ . Assume that  $S$  is invertible,  $S_{1,1} \neq 0$ ,  $S_{2,2} \neq 0$ , and  $\partial_\kappa \partial_{y^2}^2 g_2(0, 0, x; \theta, \rho) = 0$  for all  $\theta$ ,  $x$  and  $\rho$ , then there exists a consistent estimator  $\hat{\theta}_N$  solution of  $G_{N,\alpha}(\hat{\theta}_N) = 0$ , is unique with a probability that goes to one as  $N \rightarrow \infty$ , and asymptotically Gaussian:

$$\left( \begin{array}{c} \sqrt{N\delta_N}(\hat{\kappa}_N - \kappa_0) \\ \sqrt{N\delta_N^{\frac{1}{\alpha}}}(\hat{\lambda}_N - \lambda_0) \end{array} \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left( \mathbf{0}_2, \left( \begin{array}{cc} \frac{W_1(\partial_y g_1(0, 0, \cdot; \theta_0, \rho_\infty))}{S_{1,1}^2} & 0 \\ 0 & \frac{W_2(\partial_{y^2}^2 g_2(0, 0, \cdot; \theta_0, \rho_\infty))}{4S_{2,2}^2} \end{array} \right) \right) \quad (12)$$

as  $N \rightarrow \infty$ , with  $N\delta_N \rightarrow \infty$  and  $N\delta_N^{2-\frac{1}{\alpha}} \rightarrow 0$ .

The assumption  $N\delta_N^{2-\frac{1}{\alpha}}$  in this theorem corresponds to  $k_N\Delta_N^2 \rightarrow 0$ . In the case of direct observations, the usual condition is  $N\delta_N^2 \rightarrow 0$  (see *e.g.* Sørensen (2009)).

One of the main differences between the estimating functions built on the local means  $(Y_{\bullet}^j)$  and those built on the direct observations  $(X_{j\Delta_N})$  is that the approximation  $Y_{\bullet}^{j+1} - Y_{\bullet}^j$  is  $\mathcal{H}_{j+2}^N$  measurable and involves the random variables  $\zeta_{j+1,N}$  and  $\zeta'_{j+2,N}$  (see the Appendix) which have an MA(1) structure, whereas the approximation  $X_{(j+1)\Delta_N} - X_{j\Delta_N}$  involves the random variable  $B_{(j+1)\Delta_N} - B_{j\Delta_N}$  and is  $\mathcal{G}_{j+1}^N$  measurable. To avoid cumbersome correlations between  $Y_{\bullet}^j$  and  $Y_{\bullet}^{j+1}$ , the estimating function for  $\lambda_0$

$$H_{N,\alpha}(\theta) = \frac{1}{k_N\Delta_N} \sum_{j=1}^{k_N-2} g_{2,\alpha}(\delta_N, Y_{\bullet}^{3j+1} - Y_{\bullet}^{3j}, Y_{\bullet}^{3j-1}; \theta, \rho_N), \quad (13)$$

is considered, for a sample  $X_{i\delta_N}, i = 0, \dots, 3N$  of size  $3N + 1$  in place of  $N + 1$ , with  $N = p_N k_N$ . The results of uniform convergence in probability under  $\mathbb{P}_{\theta_0}$  remain valid:

$$\begin{aligned} H_{N,\alpha}(\theta) &= \frac{1}{3} \nu_0((c(\cdot, \lambda_0) - c(\cdot, \lambda)) \partial_{y^2}^2 g_2(0, 0, \cdot; \theta, \rho_\infty)) + o_P(1), \\ \partial_\theta H_{N,\alpha}(\theta) &= \nu_0(\bar{L}_{\theta_0}(g_2(0, \theta, \rho_\infty)) - \bar{L}_\theta(g_2(0, \theta, \rho_\infty))) \\ &\quad - \frac{1}{3} \nu_0(\partial_\lambda c(\cdot, \lambda) \partial_{y^2}^2 g_2(0, 0, \cdot; \theta, \rho_\infty)) + o_P(1) \end{aligned}$$

uniformly in  $\theta \in \Theta$ . But for the convergence in distribution, the asymptotic variance decreases.

**Theorem 5.** *Assume that  $N\delta_N^{2-\frac{1}{\alpha}} \rightarrow 0$ . Under **(A1)**-**(A5)**, for a function  $f$  satisfying Condition **(C1)**,*

$$\begin{aligned} \sqrt{N\delta_N^{\frac{1}{\alpha}}} \left( \frac{1}{k_N\Delta_N} \sum_{j=1}^{k_N-2} f(Y_{\bullet}^{3j-1})(Y_{\bullet}^{3j+1} - Y_{\bullet}^{3j})^2 \right. \\ \left. - \frac{1}{k_N} \sum_{j=0}^{k_N-2} f(Y_{\bullet}^{3j}) \left( \frac{2}{3} \sigma(Y_{\bullet}^{3j})^2 + 2\rho_N^2 \Delta_N^{\frac{2-\alpha}{\alpha-1}} \right) \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, W_3(f)) \end{aligned}$$

where  $W_3(f) = \frac{8}{9} \nu_0(f^2 \sigma^4)$  if  $\alpha \in (1, 2)$  and **(B1)** or **(B2)**, or  $\alpha = 2$  and **(B2)**, and  $W_3(f) = \nu_0(f^2 (\frac{8}{9} \sigma^4 + 8\rho^4))$  if  $\alpha = 2$  and **(B1)**.

Then, with an estimator  $\hat{\lambda}_N$  of  $\lambda_0$  based on the estimating function  $H_{N,\alpha}$  and a sample of size  $3N + 1$ , the estimator  $\hat{\lambda}_N$  is consistent and asymptotically Gaussian:

$$\sqrt{N\delta_N^{\frac{1}{\alpha}}} (\hat{\lambda}_N - \lambda_0) \xrightarrow{\mathcal{L}} \mathcal{N} \left( 0, \frac{W_3(\partial_{y^2}^2 g_2(0, 0, \cdot; \theta_0, \rho_\infty))}{4S_{2,2}^2} \right) \quad (14)$$

as  $N \rightarrow \infty$ , with  $N\delta_N \rightarrow \infty$  and  $N\delta_N^{\frac{2-\alpha}{\alpha}} \rightarrow 0$ .

Then, for  $\alpha \in (1, 2)$ , the estimator of  $\lambda_0$  based on the estimating function  $H_{N,\alpha}$  has a better asymptotic variance. If the *second Jacobsen's condition* holds, the asymptotic variance is

$$\frac{W_3(\partial_{y_2}^2 g_2(0, 0, \cdot; \theta_0, \rho_\infty))}{4S_{2,2}^2} = 2 \left\{ \nu_0 \left( \frac{(\partial_\lambda c(\cdot, \lambda_0))^2}{c(\cdot, \lambda_0)^2} \right) \right\}^{-1}.$$

This is the same asymptotic variance as in the case of direct observations of the diffusion, and the estimator is efficient.

## 4 Applications and examples

The main application of these results about estimating functions is the asymptotic normality of the minimum contrast estimators built in Favetto (2010).

Consider the contrast

$$\mathcal{E}_N(\theta) = \sum_{j=1}^{k_N-2} \left\{ \frac{3}{2\Delta_N} \frac{(Y_{\bullet}^{j+1} - Y_{\bullet}^j - \Delta_N b(Y_{\bullet}^{j-1}, \kappa))^2}{c_N(Y_{\bullet}^{j-1}, \lambda)} + \log(c_N(Y_{\bullet}^{j-1}, \lambda)) \right\} \quad (15)$$

where  $c_N(\cdot, \lambda) = c(\cdot, \lambda) + 3\rho_N^2 \Delta_N^{\frac{2-\alpha}{\alpha-1}}$ . Then

$$\begin{aligned} \frac{\partial}{\partial \kappa} \mathcal{E}_N(\theta_0) &= -3 \sum_{j=1}^{k_N-2} \left( \frac{\partial_\kappa b(Y_{\bullet}^{j-1}, \kappa_0)}{c_N(Y_{\bullet}^{j-1}, \lambda_0)} (Y_{\bullet}^{j+1} - Y_{\bullet}^j - \Delta_N b(Y_{\bullet}^{j-1}, \kappa_0)) \right), \\ \frac{\partial}{\partial \lambda} \mathcal{E}_N(\theta_0) &= - \sum_{j=1}^{k_N-2} \left( \frac{\partial_\lambda c(Y_{\bullet}^{j-1}, \lambda_0)}{c_N(Y_{\bullet}^{j-1}, \lambda_0)^2} \right) \frac{3}{2\Delta_N} (Y_{\bullet}^{j+1} - Y_{\bullet}^j - \Delta_N b(Y_{\bullet}^{j-1}, \lambda_0))^2 \\ &\quad + \sum_{j=1}^{k_N-2} \left( \frac{\partial_\lambda c(Y_{\bullet}^{j-1}, \lambda_0)}{c_N(Y_{\bullet}^{j-1}, \lambda_0)^2} \right) (c(Y_{\bullet}^{j-1}, \lambda_0) + 3\Delta_N^{\frac{2-\alpha}{\alpha-1}}). \end{aligned}$$

Hence, for  $\alpha \in (1, 2]$ , the associated estimating function  $G_{N,\alpha}(\theta)$  is defined by

$$G_{N,\alpha}(\theta) = \frac{1}{k_N \Delta_N} \left( \begin{array}{c} \sum_{j=1}^{k_N-2} \frac{\partial_\kappa b(Y_{\bullet}^{j-1}, \kappa)}{c_N(Y_{\bullet}^{j-1}, \lambda)} (Y_{\bullet}^{j+1} - Y_{\bullet}^j - \Delta_N b(Y_{\bullet}^{j-1}, \kappa)) \\ \sum_{j=1}^{k_N-2} \frac{\partial_\lambda c(Y_{\bullet}^{j-1}, \lambda)}{c_N(Y_{\bullet}^{j-1}, \lambda)^2} \left( \frac{1}{2} (Y_{\bullet}^{j+1} - Y_{\bullet}^j - \Delta_N b(Y_{\bullet}^{j-1}, \kappa))^2 - \frac{\Delta_N}{3} c(Y_{\bullet}^{j-1}, \lambda) - \Delta_N^{\frac{1}{\alpha-1}} \rho_N^2 \right) \end{array} \right)$$

with  $\Delta_N = \delta_N^{1-\frac{1}{\alpha}}$ .

Setting  $c_\rho(x, \lambda) = c(x, \lambda)$  if  $1 < \alpha < 2$  and  $c_\rho(x, \lambda) = c(x, \lambda) + 3\rho^2$  if  $\alpha = 2$ , as  $|c_N(x, \lambda) - c_\rho(x, \lambda)| \leq 3\Delta_N^{\frac{2-\alpha}{\alpha-1}} \rho_N$  if  $\alpha \in (1, 2)$  or  $\alpha = 2$  and  $\rho_N \rightarrow 0$ ,

and vanishes if  $\alpha = 2$  and  $\rho_N = \rho$ , let  $G_{N,\alpha}^*(\theta)$  be the estimating function defined by

$$G_{N,\alpha}^*(\theta) = \frac{1}{k_N \Delta_N} \left( \begin{array}{c} \sum_{j=1}^{k_N-2} \frac{\partial_\kappa b(Y_\bullet^{j-1}, \kappa)}{c_\rho(Y_\bullet^{j-1}, \lambda)} (Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1}, \kappa)) \\ \sum_{j=1}^{k_N-2} \frac{\partial_\lambda c(Y_\bullet^{j-1}, \lambda)}{c_\rho(Y_\bullet^{j-1}, \lambda)^2} \left( \frac{1}{2} (Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1}, \kappa))^2 - \frac{\Delta_N}{3} c(Y_\bullet^{j-1}, \lambda) - \Delta_N^{\frac{1}{\alpha-1}} \rho_N^2 \right) \end{array} \right)$$

This definition is justified by Corollary 1. The estimator  $\hat{\theta}_N$  is defined by  $G_{N,\alpha}^*(\hat{\theta}_N) = 0$ .

Setting

$$\begin{aligned} g_{1,\alpha}(\delta, y, x; \theta, \rho) &= \frac{\partial_\kappa b(x, \kappa)}{c_\rho(x, \lambda)} (y - \delta^{1-\frac{1}{\alpha}} b(x, \kappa)), \\ g_{2,\alpha}(\delta, y, x; \theta, \rho) &= \frac{\partial_\lambda c(x, \kappa)}{c_\rho(x, \lambda)^2} \left\{ \frac{1}{2} (y - \delta^{1-\frac{1}{\alpha}} b(x, \kappa))^2 - \frac{\delta^{1-\frac{1}{3}}}{3} c(x, \lambda) - \delta^{\frac{1}{\alpha}} \rho^2 \right\}, \end{aligned}$$

we have

$$\begin{aligned} g_{1,\alpha}^{(1)}(y, x; \theta, \rho) &= -\frac{\partial_\kappa b(x, \kappa)}{c_\rho(x, \lambda)} b(x, \kappa), \\ g_{2,\alpha}^{(1)}(y, x; \theta, \rho) &= -\frac{1}{3} c(x, \lambda) \frac{\partial_\lambda c(x, \lambda)}{c_\rho(x, \lambda)^2} - y b(x, \kappa) \quad \text{if } 1 < \alpha < 2, \\ &= -\frac{1}{3} c_\rho(x, \lambda) \frac{\partial_\lambda c(x, \lambda)}{c_\rho(x, \lambda)^2} - y b(x, \kappa) \quad \text{if } \alpha = 2, \\ g_{1,\alpha}^{(\alpha)}(y, x; \theta, \rho) &= 0, \\ g_{2,\alpha}^{(\alpha)}(y, x; \theta, \rho) &= -\rho^2 \frac{\partial_\lambda c(x, \lambda)}{c_\rho(x, \lambda)^2} \end{aligned}$$

and  $g^{(\alpha)}$  is only needed if  $\alpha \in (\frac{3}{2}, 2]$ . Otherwise,  $\delta_N^{\frac{1}{\alpha}} = O(\delta_N^{2-\frac{2}{\alpha}})$ .

The estimating function  $G_{N,\alpha}^*$  satisfies the *rate optimality* condition:

$$\partial_y g_{2,\alpha}(0, 0, x; \theta, \rho) = 0$$

and the *second Jacobsen condition* if  $c_\rho = c$ , i.e. if  $\alpha \in (1, 2)$  or if  $\alpha = 2$  and **(B2)**. Hence, in this case, the following result holds.

**Theorem 6.** *Assume that  $\alpha \in (1, 2)$  and **(B1)** or **(B2)**, or  $\alpha = 2$  and **(B2)**. Then the estimator  $\hat{\theta}_N = (\hat{\kappa}_N, \hat{\lambda}_N)$  defined by  $G_{N,\alpha}^*(\hat{\theta}_N) = 0$  is consistent, and asymptotically Gaussian:*

$$\left( \begin{array}{c} \sqrt{N\delta_N}(\hat{\kappa}_N - \kappa_0) \\ \sqrt{N\delta_N^{\frac{1}{\alpha}}}(\hat{\lambda}_N - \lambda_0) \end{array} \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left( \mathbf{0}, \left( \begin{array}{cc} \left\{ \nu_0 \left( \frac{(\partial_\kappa b(\cdot, \kappa_0))^2}{c(\cdot, \lambda_0)} \right) \right\}^{-1} & 0 \\ 0 & \frac{9}{4} \left\{ \nu_0 \left( \frac{(\partial_\lambda c(\cdot, \lambda_0))^2}{c(\cdot, \lambda_0)^2} \right) \right\}^{-1} \end{array} \right) \right)$$

as  $N \rightarrow \infty$ , with  $N\delta_N \rightarrow \infty$  and  $N\delta_N^{2-\frac{1}{\alpha}} \rightarrow 0$ .

In the case  $\alpha = 2$  and  $\rho_N = \rho$ , with  $c_\rho(x, \lambda) = c(x, \lambda) + 3\rho^2$ , the second Jacobsen's condition is not satisfied. The estimator  $\hat{\theta}_N$  is still consistent, but we have

$$\begin{pmatrix} \sqrt{N\delta_N}(\hat{\kappa}_N - \kappa_0) \\ \sqrt{N\delta_N^{\frac{1}{\alpha}}}(\hat{\lambda}_N - \lambda_0) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N} \left( \mathbf{0}, \begin{pmatrix} \nu_0 \left( \frac{(\partial_\kappa b(\cdot, \kappa_0))^2 c(\cdot, \lambda_0)}{c_\rho(\cdot, \lambda_0)^2} \right) & 0 \\ \nu_0 \left( \frac{(\partial_\kappa b(\cdot, \kappa_0))^2}{c_\rho(\cdot, \lambda_0)^2} \right)^2 & \frac{9}{4} \frac{\nu_0 \left( \frac{(\partial_\lambda c(\cdot, \lambda_0))^2 (c^2 + 4c\rho^2 + 12\rho^4)}{c_\rho(\cdot, \lambda_0)^4} \right)}{\nu_0 \left( \frac{(\partial_\lambda c(\cdot, \lambda_0))^2}{c_\rho(\cdot, \lambda_0)^2} \right)^2} \end{pmatrix} \right)$$

as  $N \rightarrow \infty$ , with  $N\delta_N \rightarrow \infty$  and  $N\delta_N^{2-\frac{1}{\alpha}} \rightarrow 0$ .

An other estimating function can be considered:

$$G_{N,\alpha}^\diamond(\theta) = \frac{1}{k_N \Delta_N} \begin{pmatrix} \sum_{j=1}^{k_N-2} \frac{\partial_\kappa b(Y_\bullet^{j-1}, \kappa)}{c(Y_\bullet^{j-1}, \lambda)} (Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1}, \kappa)) \\ \sum_{j=1}^{k_N-2} \frac{\partial_\lambda c(Y_\bullet^{j-1}, \lambda)}{c(Y_\bullet^{j-1}, \lambda)^2} \left( \frac{1}{2} (Y_\bullet^{j+1} - Y_\bullet^j)^2 - \frac{\Delta_N}{3} c(Y_\bullet^{j-1}, \lambda) - \Delta_N^{\frac{1}{\alpha-1}} \rho^2 \right) \end{pmatrix}$$

but  $G_{N,\alpha}^\diamond$  is not a gradient: there is no function  $V$  such that  $G_{N,\alpha}^\diamond = \text{grad}(V)$ . Indeed, setting

$$\begin{aligned} g_{1,\alpha}^\diamond(\delta, y, x; \theta, \rho) &= \frac{\partial_\kappa b(x, \kappa)}{c(x, \lambda)} (y - \delta^{1-\frac{1}{\alpha}} b(x, \kappa)), \\ g_{2,\alpha}^\diamond(\delta, y, x; \theta, \rho) &= \frac{\partial_\lambda c(x, \lambda)}{c(x, \lambda)^2} \left( \frac{y^2}{2} - \delta^{1-\frac{1}{\alpha}} \frac{c(x, \lambda)}{3} c(x, \lambda) - \delta^{\frac{1}{\alpha}} \rho^2 \mathbf{1}_{\alpha \in (\frac{3}{2}, 2]} \right) \end{aligned}$$

it comes  $\partial_\lambda g_{1,\alpha}^\diamond \neq \partial_\kappa g_{2,\alpha}^\diamond$ .

However,  $g_\alpha^\diamond$  satisfies the *second Jacobsen's condition*. Setting  $\hat{\theta}_N^\diamond$  as the solution of  $G_{N,\alpha}^\diamond(\theta) = 0$ ,

$$\begin{pmatrix} \sqrt{N\delta_N}(\hat{\kappa}_N - \kappa_0) \\ \sqrt{N\delta_N^{\frac{1}{\alpha}}}(\hat{\lambda}_N - \lambda_0) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N} \left( \mathbf{0}, \begin{pmatrix} \left\{ \nu_0 \left( \frac{(\partial_\kappa b(\cdot, \kappa_0))^2}{c(\cdot, \lambda_0)} \right) \right\}^{-1} & 0 \\ 0 & \frac{9}{4} \frac{\nu_0 \left( \frac{(\partial_\lambda c(\cdot, \lambda_0))^2 (c^2 + 4c\rho^2 + 12\rho^4)}{c(\cdot, \lambda_0)^4} \right)}{\nu_0 \left( \frac{(\partial_\lambda c(\cdot, \lambda_0))^2}{c(\cdot, \lambda_0)^2} \right)^2} \end{pmatrix} \right)$$

as  $N \rightarrow \infty$ , with  $N\delta_N \rightarrow \infty$  and  $N\delta_N^{2-\frac{1}{\alpha}} \rightarrow 0$ .

Hence, the asymptotic variance obtained for the estimation of  $\kappa_0$  is the same that in the case of direct observations of the diffusion process.

#### 4.1 The Ornstein-Uhlenbeck diffusion

Let  $(X_t)$  be the solution of

$$dX_t = \kappa X_t dt + \lambda dB_t,$$

with  $\kappa < 0$ ,  $\lambda > 0$  and  $X_0$  deterministic or distributed with the stationary distribution of  $(X_t)$ . Explicit estimators  $\hat{\kappa}_N$  and  $\hat{\lambda}_N$  are given in Favetto (2010) and recalled here

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

For  $\alpha \in (1, 2)$  or  $\alpha = 2$  and  $\rho_N \rightarrow 0$ , these estimators are asymptotically Gaussian:

$$\begin{pmatrix} \sqrt{N\delta_N}(\hat{\kappa}_N - \kappa_0) \\ \sqrt{N\delta_N^{\frac{1}{\alpha}}}(\hat{\lambda}_N - \lambda_0) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 2|\kappa| & 0 \\ 0 & \frac{9}{4}\lambda^4 \end{pmatrix}\right)$$

as  $N \rightarrow \infty$ , with  $N\delta_N \rightarrow \infty$  and  $N\delta_N^{2-\frac{1}{\alpha}} \rightarrow 0$ .

## 4.2 The hyperbolic diffusion

Let  $(X_t)$  be the solution of

$$dX_t = \kappa X_t dt + \lambda \sqrt{1 + X_t^2} dB_t, \quad X_0 = \eta \in \mathbb{R}, \quad (16)$$

where  $\eta$  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

$$\nu_0(x) \propto \frac{1}{(1 + x^2)^{1 + \frac{|\kappa|}{\lambda^2}}}.$$

If  $X_0 = \eta$  has distribution  $\nu_0(x)dx$ , then,  $\sqrt{1 + \frac{2|\kappa|}{\lambda^2}}\eta$  has Student distribution. In this case,

$$\mathbb{E}_{\theta}(X_0) = 0 \text{ and } \mathbb{E}_{\theta}\left(\frac{X_0^2}{1 + X_0^2}\right) = \frac{3 + \frac{2|\kappa|}{\lambda^2}}{\left(\frac{2|\kappa|}{\lambda^2} + 1\right)^{3/2}}.$$

Then the estimators for  $\kappa_0$  and  $\lambda_0^2$  are

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

and they are asymptotically Gaussian

$$\begin{pmatrix} \sqrt{N\delta_N}(\hat{\kappa}_N - \kappa_0) \\ \sqrt{N\delta_N^{\frac{1}{\alpha}}}(\hat{\lambda}_N - \lambda_0) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \frac{\lambda^2 \left(\frac{2|\kappa|}{\lambda^2} + 1\right)^{3/2}}{3 + \frac{2|\kappa|}{\lambda^2}} & 0 \\ 0 & \frac{9}{4}\lambda^4 \end{pmatrix}\right)$$

as  $N \rightarrow \infty$ , with  $N\delta_N \rightarrow \infty$ ,  $\delta_N \rightarrow 0$  and  $N\delta_N^{2-\frac{1}{\alpha}} \rightarrow 0$ , in the case  $\alpha \in (1, 2)$  or  $\alpha = 2$  and  $\rho_N \rightarrow 0$ .

## 5 Concluding remarks

In this article, the asymptotic normality of the minimum contrast estimators of Favetto (2010) is proved in the context of estimating functions, which correspond to the gradient of the contrast function. Two different rates of convergence in distribution are obtained for the drift parameter and the diffusion coefficient parameter. The drift parameter estimator  $\hat{\kappa}_N$  is asymptotically Gaussian at rate  $\sqrt{N\delta_N}$ , which is the usual rate for directly observed diffusions, whereas the diffusion coefficient parameter estimator  $\hat{\lambda}_N$  is asymptotically Gaussian at rate  $\sqrt{N\delta_N^{\frac{1}{\alpha}}}$ : this is slower than the rate of convergence  $\sqrt{N}$  for directly observed diffusions, and depends on the local mean parameter  $\alpha$ . Finally, the results presented here show that even for non-Markovian observations, it is possible to introduce an estimating function, and obtain an estimator with good asymptical properties.

## Acknowledgements

The author wishes to adress many thanks to his PhD. advisor, Pr. Valentine Genon-Catalot, for her help, her advices and suggestions about this article.

## 6 Proofs

In the proofs, we use for sake of simplicity  $\Delta_N = \delta_N^{1-\frac{1}{\alpha}}$ .

**Proof of Lemma 1.** Setting

$$B_{j,N} = \mathbb{E}_\theta(g(\Delta_N, Y_\bullet^{j+1} - Y_\bullet^j, Y_\bullet^{j-1}; \theta, \rho_N) | \mathcal{H}_j^N),$$



we have, with Taylor's formula,

$$\begin{aligned}
B_{j,N} &= g(0, 0, Y_{\bullet}^{j-1}; \theta, \rho_N) + \partial_y g(0, 0, Y_{\bullet}^{j-1}; \theta, \rho_N) \mathbb{E}_{\theta}(Y_{\bullet}^{j+1} - Y_{\bullet}^j | \mathcal{H}_j^N) \\
&\quad + \frac{1}{2} \partial_{y^2}^2 g(0, 0, Y_{\bullet}^{j-1}, \theta, \rho_N) \mathbb{E}_{\theta}((Y_{\bullet}^{j+1} - Y_{\bullet}^j)^2 | \mathcal{H}_j^N) \\
&\quad + \Delta_N g^{(1)}(0, Y_{\bullet}^{j+1}; \theta, \rho_N) + \Delta_N^{\frac{1}{\alpha-1}} g^{(\alpha)}(0, Y_{\bullet}^{j-1}; \theta, \rho_N) \\
&\quad + \Delta_N^2 R(\Delta_N, Y_{\bullet}^{j-1}, X_{j\Delta_N}; \theta, \rho_N) \\
&= g(0, 0, Y_{\bullet}^{j-1}; \theta, \rho_N) + \partial_y g(0, 0, Y_{\bullet}^{j-1}; \theta, \rho_N) \Delta_N b(X_{j\Delta_N}, \kappa) \\
&\quad + \frac{1}{2} \partial_{y^2}^2 g(0, 0, Y_{\bullet}^{j-1}, \theta, \rho_N) (\Delta_N c(X_{j\Delta_N}, \lambda) + 2\Delta_N^{\frac{1}{\alpha-1}} \rho_N) \\
&\quad + \Delta_N g^{(1)}(0, Y_{\bullet}^{j-1}; \theta, \rho_N) + \Delta_N^{\frac{1}{\alpha-1}} g^{(\frac{1}{\alpha-1})}(0, Y_{\bullet}^{j-1}; \theta, \rho_N) \\
&\quad + \Delta_N^2 R(\Delta_N, Y_{\bullet}^{j-1}, X_{j\Delta_N}; \theta, \rho_N).
\end{aligned}$$

Then, replacing  $X_{j\Delta_N}$  by  $Y_{\bullet}^{j-1}$ ,

$$\begin{aligned}
B_{j,N} &= g(0, 0, Y_{\bullet}^{j-1}; \theta, \rho_N) \\
&\quad + \Delta_N (\bar{L}_{\theta}(g(0, \theta, \rho_N))(0, Y_{\bullet}^{j-1}) + g^{(1)}(0, Y_{\bullet}^{j-1}; \theta, \rho_N)) \\
&\quad + \Delta_N^{\frac{1}{\alpha-1}} (g^{(\frac{1}{\alpha-1})}(0, Y_{\bullet}^{j-1}; \theta, \rho_N) + \rho_N^2 \partial_{y^2}^2 g(0, 0, Y_{\bullet}^{j-1}, \theta, \rho_N)) \\
&\quad + \Delta_N^2 R(\Delta_N, Y_{\bullet}^{j-1}, X_{j\Delta_N}; \theta, \rho_N)
\end{aligned}$$

where the terms

$$\Delta_N \partial_y g(0, 0, Y_{\bullet}^{j-1}; \theta, \rho_N) (b(X_{j\Delta_N}, \kappa) - b(Y_{\bullet}^{j-1}, \kappa))$$

and

$$\frac{\Delta_N}{2} \partial_{y^2}^2 g(0, 0, Y_{\bullet}^{j-1}, \theta, \rho_N) (c(X_{j\Delta_N}, \lambda) - c(Y_{\bullet}^{j-1}, \lambda))$$

are  $\mathcal{H}_j^N$  measurable and negligible: their conditional expectation given  $\mathcal{H}_{j-1}^N$  is of order  $\Delta_N^2$ . The proof ends with

$$\mathbb{E}_{\theta}(g(\Delta_N, Y_{\bullet}^{j+1} - Y_{\bullet}^j, Y_{\bullet}^{j-1}; \theta, \rho_N) | \mathcal{H}_j^N) = O(\Delta_N^2)$$

and the identification of the terms.  $\square$

**Proof of Lemma 2.** Setting  $A_{j,N}(\theta, \theta_0) = \mathbb{E}_{\theta_0}(g(\Delta_N, Y_{\bullet}^{j+1} - Y_{\bullet}^j, Y_{\bullet}^{j-1}; \theta, \rho_N) | \mathcal{H}_j^N)$ ,

$$\begin{aligned}
A_{j,N}(\theta, \theta_0) &= \Delta_N (g^{(1)}(0, Y_{\bullet}^{j-1}; \theta, \rho_N) + \bar{L}_{\theta_0}(g(0, \theta))(0, Y_{\bullet}^{j-1}, \rho_N)) \\
&\quad + \Delta_N^{\frac{1}{\alpha-1}} (g^{(\alpha)}(0, Y_{\bullet}^{j-1}; \theta, \rho_N) + \rho_N^2 \partial_{y^2}^2 g(0, 0, Y_{\bullet}^{j-1}, \theta, \rho_N)) \\
&\quad + \Delta_N^2 R(\Delta_N, Y_{\bullet}^{j-1}, X_{j\Delta_N}; \theta, \rho_N) \\
&= \Delta_N (\bar{L}_{\theta_0}((g(0, \theta, \rho_N))(0, Y_{\bullet}^{j-1})) - \bar{L}_{\theta}(g(0, \theta, \rho_N))(0, Y_{\bullet}^{j-1})) \\
&\quad + \Delta_N^2 R(\Delta_N, Y_{\bullet}^{j-1}; \theta, \rho_N).
\end{aligned}$$

Splitting the sum  $\frac{1}{k_N \Delta_N} \sum_{j=1}^{k_N-2} A_{j,N}(\theta, \theta_0)$  in three parts, we conclude for the pointwise convergence with Lemma 4 in the Appendix. For the uniform

convergence in  $\theta$ , recall that  $\Theta$  is compact and the functionals  $\bar{I}_N(f(\cdot, \theta))$  and  $\bar{Q}_N(f(\cdot, \theta))$  studied in Favetto (2010) converge in probability uniformly in  $\theta$ .

Thus, with Taylor's formula, as  $g$  admits a developpement in powers of  $\Delta$ ,

$$\begin{aligned} g(\Delta, Y_{\bullet}^{j+1} - Y_{\bullet}^j, Y_{\bullet}^{j-1}; \theta, \rho_N) &= g(0, 0, Y_{\bullet}^{j-1}; \theta, \rho_N) \\ &\quad + (Y_{\bullet}^{j+1} - Y_{\bullet}^j) \partial_y g(0, 0, Y_{\bullet}^{j-1}, \theta, \rho_N) \\ &\quad + \frac{1}{2} (Y_{\bullet}^{j+1} - Y_{\bullet}^j)^2 \partial_{y^2}^2 g(0, 0, Y_{\bullet}^{j-1}; \theta, \rho_N) \\ &\quad + \Delta_N g^{(1)}(0, Y_{\bullet}^{j-1}; \theta, \rho_N) \\ &\quad + \Delta_N^{\frac{1}{\alpha-1}} g^{(\frac{1}{\alpha-1})}(0, Y_{\bullet}^{j-1}; \theta, \rho_N) \\ &\quad + \Delta_N^2 R(\Delta_N, Y_{\bullet}^{j+1} - Y_{\bullet}^j, Y_{\bullet}^{j-1}; \theta, \rho_N). \end{aligned}$$

Therefore, with  $g^{(1)}(0, Y_{\bullet}^{j-1}; \theta, \rho_N) = -\bar{L}_{\theta}(g(0, \theta, \rho_N))$ , the convergence of

$$\begin{aligned} &\frac{1}{k_N \Delta_N} \sum_{j=1}^{k_N-2} (\partial_y g(0, 0, Y_{\bullet}^{j-1}; \theta, \rho_N) (Y_{\bullet}^{j+1} - Y_{\bullet}^j - \Delta_N b(Y_{\bullet}^{j-1}, \kappa_0)) \\ &+ \partial_{y^2}^2 g(0, 0, Y_{\bullet}^{j-1}; \theta, \rho_N) (\frac{1}{2} (Y_{\bullet}^{j+1} - Y_{\bullet}^j)^2 - \frac{\Delta_N}{3} c(Y_{\bullet}^{j-1}, \lambda_0) - \Delta_N^{\frac{1}{\alpha-1}} g^{(\frac{1}{\alpha-1})}(0, Y_{\bullet}^{j-1}; \theta))) \end{aligned}$$

in probability is uniform in  $\theta \in \Theta$ .  $\square$

**Proof of Theorem 1.** In this proof, we set  $b(\cdot) = b(\cdot, \kappa_0)$  and  $\sigma(\cdot) = \sigma(\cdot, \lambda_0)$ . Considering a function  $f$  satisfying Condition **(C1)** and the random variables defined in Lemma 3, we have, for  $1 \leq j \leq k_N - 2$ ,

$$\begin{aligned} f(Y_{\bullet}^{j-1})(Y_{\bullet}^{j+1} - Y_{\bullet}^j - \Delta_N b(Y_{\bullet}^{j-1})) &= f(Y_{\bullet}^{j-1})(Y_{\bullet}^{j+1} - Y_{\bullet}^j - \Delta_N b(X_{j\Delta_N})) \\ &\quad + \Delta_N (b(X_{j\Delta_N}) - b(Y_{\bullet}^{j-1})). \end{aligned}$$

With Proposition 2,

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

Considering

$$\sqrt{k_N \Delta_N} \bar{I}_N(f) = \sum_{\ell=1}^4 \bar{R}_N^{(\ell)}$$

with

$$\begin{aligned}\bar{R}_N^{(1)} &= \frac{1}{\sqrt{k_N \Delta_N}} \sum_{j=1}^{k_N-2} f(Y_{\bullet}^{j-1}) \sigma(X_{j\Delta_N}) (\zeta_{j+1,N} + \zeta'_{j+2,N}), \\ \bar{R}_N^{(2)} &= \frac{1}{\sqrt{k_N \Delta_N}} \sum_{j=1}^{k_N-2} f(Y_{\bullet}^{j-1}) \rho_N(\varepsilon_{\bullet}^{j+1} - \varepsilon_{\bullet}^j), \\ \bar{R}_N^{(3)} &= \frac{1}{\sqrt{k_N \Delta_N}} \sum_{j=1}^{k_N-2} f(Y_{\bullet}^{j-1}) \tau_{j,N}, \\ \bar{R}_n^{(4)} &= \frac{1}{\sqrt{k_N \Delta_N}} \sum_{j=1}^{k_N-2} f(Y_{\bullet}^{j-1}) \Delta_N (b(X_{j\Delta_N}) - b(Y_{\bullet}^{j-1})).\end{aligned}$$

let  $r_{j,N}^{(1)}$  be defined by

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

where the terms have been rearranged, to have  $r_{j,N}^{(1)}$  be  $\mathcal{H}_{j+1}^N$  measurable.

The first quantity of interest is

$$\bar{R}_N^{(1)} = \frac{1}{\sqrt{k_N \Delta_N}} \sum_{j=1}^{k_N-2} f(Y_{\bullet}^{j-1}) \sigma(X_{j\Delta_N}) (\zeta_{j+1,N} + \zeta'_{j+2,N}),$$

and

$$\bar{R}_N^{(1)} = \frac{1}{\sqrt{k_N \Delta_N}} \sum_{j=2}^{k_N-2} r_{j,N}^{(1)} + o_P(1).$$

To prove the convergence in distribution of  $\bar{R}_N^{(1)}$ , a Central Limit Theorem for martingale arrays (Theorem 3.2 in Hall and Heyde (1980)) is used. The conditional centering  $\mathbb{E}(r_{j,N}^{(1)} | \mathcal{H}_j^N) = 0$  is ensured and the conditional variance is

$$\begin{aligned}\mathbb{E}((r_{j,N}^{(1)})^2 | \mathcal{H}_j^N) &= \Delta_N \left( f(Y_{\bullet}^{j-1})^2 \sigma(X_{j\Delta_N})^2 \Delta_N \left( \frac{1}{3} + \frac{1}{2p_N} + \frac{1}{6p_N^2} \right) \right. \\ &\quad + f(Y_{\bullet}^{j-2})^2 \sigma(X_{(j-1)\Delta_N})^2 \left( \frac{1}{3} - \frac{1}{2p_N} + \frac{1}{6p_N^2} \right) \\ &\quad \left. + f(Y_{\bullet}^{j-2}) \sigma(X_{(j-1)\Delta_N}) f(Y_{\bullet}^{j-1}) \sigma(X_{j\Delta_N}) \frac{1}{3} \left( 1 - \frac{1}{p_N^2} \right) \right).\end{aligned}$$

Hence

$$\frac{1}{k_N \Delta_N} \sum_{j=2}^{k_N-1} \mathbb{E}((r_{j,N}^{(1)})^2 | \mathcal{H}_j^N) \xrightarrow{\mathbb{P}} \nu_0(f^2 \sigma^2).$$

We easily bound  $\mathbb{E}((r_{j,N}^{(1)})^4|\mathcal{H}_j^N)$  with Proposition 2 and derive

$$\frac{1}{k_N^2 \Delta_N^2} \sum_{j=2}^{k_N-1} \mathbb{E}((r_{j,N}^{(1)})^4|\mathcal{H}_j^N) \xrightarrow{\mathbb{P}} 0.$$

Then,

$$\bar{R}_N^{(1)} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu_0(f^2 \sigma^2)).$$

We now prove that  $\bar{R}_N^{(\ell)} = o_P(1)$  for  $\ell = 2, 3$  and apply Slutsky's lemma to conclude. Dealing firstly with  $\bar{R}_N^{(3)}$ ,

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

Then, the tools are similar to the pointwise convergence: as  $\tau_{j,N}$  is  $\mathcal{H}_{j+2}^N$  measurable,  $\bar{R}_N^{(3)}$  is split in three sums, in order to prove that it converges in probability to 0 with Lemma 4. With the condition  $k_N \Delta_N^3 \rightarrow 0$ ,

$$\frac{1}{k_N \Delta_N} \sum_{2 \leq 3j \leq k_N-2} \sqrt{k_N \Delta_N} \mathbb{E}(f(Y_{\bullet}^{3j-1})\tau_{3j,N}|\mathcal{H}_{3j}^N) \xrightarrow{\mathbb{P}} 0.$$

It comes

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

Hence, as  $N \rightarrow \infty$ ,

$$\frac{1}{k_N \Delta_N} \sum_{j=1}^{k_N-2} \mathbb{E}(f(Y_{\bullet}^{3j-1})^2 \tau_{3j,N}^2|\mathcal{H}_{3j}^N) \longrightarrow 0$$

in probability.

From now on,  $\bar{R}_N^{(2)}$  is considered. Let the  $\mathcal{H}_{j+1}^N$  measurable random variable

$$r_{j,N}^{(2)} = (f(Y_{\bullet}^{j-2}) - f(Y_{\bullet}^{j-1}))\rho \varepsilon_{\bullet}^j$$

and then

$$\bar{R}_N^{(2)} = \frac{1}{\sqrt{k_N \Delta_N}} \sum_{j=2}^{k_N-2} r_{j,N}^{(2)} + o_P(1).$$

Therefore  $\mathbb{E}(r_{j,N}^{(2)}|\mathcal{H}_j^N) = 0$  and

$$\mathbb{E}((r_{j,N}^{(2)})^2|\mathcal{H}_j^N) = (f(Y_{\bullet}^{j-2}) - f(Y_{\bullet}^{j-1}))^2 \rho_N^2 \mathbb{E}((\varepsilon_{\bullet}^j)^2).$$

As  $\Delta_N = p_N^{1-\alpha}$ ,

$$\frac{1}{k_N \Delta_N} \sum_{j=2}^{k_N-2} \mathbb{E}((r_{j,N}^{(2)})^2 | \mathcal{H}_j^N) = \frac{1}{k_N} \sum_{j=2}^{k_N-2} \frac{\rho_N^2}{p_N^{2-\alpha}} (f(Y_{\bullet}^{j-2}) - f(Y_{\bullet}^{j-1}))^2$$

and this quantity vanishes in probability, for  $\alpha \in (1, 2]$ , with Assumption **(B1)** or **(B2)**. Finally,

$$\bar{R}_n^{(4)} = \frac{1}{\sqrt{k_N \Delta_N}} \sum_{j=1}^{k_N-2} f(Y_{\bullet}^{j-1}) \Delta_N (b(X_{j\Delta_N}) - b(Y_{\bullet}^{j-1})).$$

For  $2 \leq j \leq k_N - 2$ ,

$$\begin{aligned} f(Y_{\bullet}^{j-1})(b(X_{j\Delta_N}) - b(Y_{\bullet}^{j-1})) &= f(Y_{\bullet}^{j-2})(b(X_{j\Delta_N}) - b(Y_{\bullet}^{j-1})) \\ &\quad + f(Y_{\bullet}^{j-2})(b(X_{(j-1)\Delta_N}) - b(Y_{\bullet}^{j-1})) \\ &\quad + (f(Y_{\bullet}^{j-1}) - f(Y_{\bullet}^{j-2}))(b(X_{j\Delta_N}) - b(Y_{\bullet}^{j-1})). \end{aligned}$$

With Proposition 1 in the Appendix and the Cauchy Schwarz inequality, it comes

$$\begin{aligned} |\mathbb{E}(f(Y_{\bullet}^{j-2})(b(X_{j\Delta_N}) - b(Y_{\bullet}^{j-1})) | \mathcal{H}_{j-1}^N)| &\leq \frac{cf(Y_{\bullet}^{j-2})(\Delta_N(1 + X_{(j-1)\Delta_N}^2) + \rho_N^2 \sqrt{\mathbb{E}((\varepsilon_{\bullet}^{j-1})^4)})}{|\mathbb{E}(f(Y_{\bullet}^{j-2})(b(X_{(j-1)\Delta_N}) - b(Y_{\bullet}^{j-1})) | \mathcal{H}_{j-1}^N)|} \\ &\leq \frac{cf(Y_{\bullet}^{j-2})(\Delta_N(1 + X_{(j-1)\Delta_N}^2) + \rho_N^2 \sqrt{\mathbb{E}((\varepsilon_{\bullet}^{j-1})^4)})}{|\mathbb{E}((f(Y_{\bullet}^{j-1}) - f(Y_{\bullet}^{j-2}))(b(X_{j\Delta_N}) - b(Y_{\bullet}^{j-1})) | \mathcal{H}_{j-1}^N)|} \\ &\leq c(1 + X_{j\Delta_N}^2 + \rho_N^2 \mathbb{E}((\varepsilon_{\bullet}^{j-1})^2))(\Delta_N(1 + X_{j\Delta_N}^4) + \rho_N^2 \sqrt{\mathbb{E}((\varepsilon_{\bullet}^{j-1})^4)}). \end{aligned}$$

With  $k_N \Delta_N^3 = N \delta_N^{3-\frac{2}{\alpha}} \rightarrow 0$ ,  $\bar{R}_N^{(4)} = o_P(1)$ .  $\square$

**Proof of Theorem 2.** With the notations of Proposition 2,

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

Setting  $U_N(f) = \sqrt{k_N}(\bar{Q}_N(f) - \bar{\nu}_N(f(\frac{2}{3}\sigma^2 + 2\rho_N^2\Delta_N^{\frac{2-\alpha}{\alpha-1}})))$ , we define

$$\begin{aligned}
u_{j,N}^{(1)} &= f(Y_{\bullet}^{j-1})\sigma(X_{j\Delta_N})^2\left(\frac{1}{\Delta_N}(\zeta_{j+1,N} + \zeta'_{j+2,N})^2 - \frac{2}{3}\right) \\
u_{j,N}^{(2)} &= \frac{2}{3}f(Y_{\bullet}^{j-1})(\sigma(X_{j\Delta_N})^2 - \sigma(Y_{\bullet}^j)^2) \\
u_{j,N}^{(3)} &= f(Y_{\bullet}^{j-1})\rho_N^2\left(\frac{1}{\Delta_N}(\varepsilon_{\bullet}^{j+1} - \varepsilon_{\bullet}^j)^2 - 2\Delta_N^{\frac{2-\alpha}{\alpha-1}}\right) \\
u_{j,N}^{(4)} &= 2f(Y_{\bullet}^{j-1})\sigma(X_{j\Delta_N})\rho_N\frac{1}{\Delta_N}(\zeta_{j+1,N} + \zeta'_{j+2,N})(\varepsilon_{\bullet}^{j+1} - \varepsilon_{\bullet}^j) \\
u_{j,N}^{(5)} &= f(Y_{\bullet}^{j-1})(\Delta_N b(X_{j\Delta_N}) + \tau_{j,N})^2 \\
u_{j,N}^{(6)} &= 2f(Y_{\bullet}^{j-1})(\Delta_N b(X_{j\Delta_N}) + \tau_{j,N})\sigma(X_{j\Delta_N})(\zeta_{j+1,N} + \zeta'_{j+2,N}) \\
u_{j,N}^{(7)} &= 2f(Y_{\bullet}^{j-1})\rho_N(\Delta_N b(X_{j\Delta_N}) + \tau_{j,N})(\varepsilon_{\bullet}^{j+1} - \varepsilon_{\bullet}^j)
\end{aligned}$$

so that

$$U_N(f) = \frac{1}{\sqrt{k_N}} \sum_{j=1}^{k_N-2} \sum_{\ell=1}^7 u_{j,N}^{(\ell)}.$$

Recall that  $\zeta_{j+1,N}$  is  $\mathcal{H}_{j+1}^N$  measurable and  $\zeta'_{j+2,N}$  is  $\mathcal{H}_{j+2,N}^N$  measurable, so some terms have to be reordered to use a Central Limit Theorem for martingale array. Then,

$$U_N^{(1)} = \frac{1}{\sqrt{k_N}} \sum_{j=1}^{k_N-2} u_{j,N}^{(1)} = \frac{1}{\sqrt{k_N}} \sum_{j=2}^{k_N-2} s_{j,N}^{(1)} + \frac{1}{\sqrt{k_N}} \sum_{j=2}^{k_N-2} \tilde{s}_{j,N}^{(1)} + o_P(1)$$

with

$$\begin{aligned}
s_{j,N}^{(1)} &= f(Y_{\bullet}^{j-1})\sigma(X_{j\Delta_N})^2 \left( \frac{\zeta_{j+1,N}^2}{\Delta_N} - m_N \right) \\
&\quad + f(Y_{\bullet}^{j-2})\sigma(X_{(j-1)\Delta_N})^2 \left( \frac{\zeta'_{j+1,N}{}^2}{\Delta_N} - m'_N \right) \\
&\quad + 2f(Y_{\bullet}^{j-2})\sigma(X_{(j-1)\Delta_N})^2 \frac{\zeta_{j,N}\zeta'_{j+1,N}}{\Delta_N}, \\
\tilde{s}_{j,N}^{(1)} &= f(Y_{\bullet}^{j-1})\sigma(X_{j\Delta_N})^2 \left( \frac{1}{2p_N} + \frac{1}{6p_N^2} \right) \\
&\quad + f(Y_{\bullet}^{j-2})\sigma(X_{(j-1)\Delta_N})^2 \left( -\frac{1}{2p_N} + \frac{1}{6p_N^2} \right)
\end{aligned}$$

where  $m_N = \frac{1}{3} + \frac{1}{2p_N} + \frac{1}{6p_N^2}$ ,  $m'_N = \frac{1}{3} - \frac{1}{2p_N} + \frac{1}{6p_N^2}$  and  $\chi_N = \frac{1}{6}(1 - \frac{1}{p_N^2})$  such that  $\zeta_{j+1,N} \sim \mathcal{N}(0, m_N\Delta_N)$ ,  $\zeta'_{j+1,N} \sim \mathcal{N}(0, m'_N\Delta_N)$ , and  $\text{Cov}(\zeta_{j+1,N}, \zeta'_{j+1,N}) = \chi_N\Delta_N$ . As  $p_N^{-1} = \Delta_N^{\frac{1}{\alpha-1}} \leq \Delta_N$ , with  $k_N\Delta_N^2 = N\delta_N^{2-\frac{1}{\alpha}} \rightarrow 0$ , we deduce

$$\frac{1}{\sqrt{k_N}} \sum_{j=2}^{k_N-2} \tilde{s}_{j,N}^{(1)} = o_P(1).$$

The conditional centering condition holds:  $\mathbb{E}(s_{j,N}^{(1)}|\mathcal{H}_j^N) = 0$ . With

$$\begin{aligned}
(s_{j,N}^{(1)})^2 &= f(Y_{\bullet}^{j-1})^2 \sigma(X_{j\Delta_N})^4 \left( \frac{\zeta_{j+1,N}^2}{\Delta_N} - m_N \right)^2 \\
&\quad + f(Y_{\bullet}^{j-2})^2 \sigma(X_{(j-1)\Delta_N})^4 \left( \frac{\zeta'_{j+1,N}}{\Delta_N} - m'_N \right)^2 \\
&\quad + 4f(Y_{\bullet}^{j-2})^2 \sigma(X_{(j-1)\Delta_N})^4 \frac{\zeta_{j,N}^2 \zeta'_{j+1,N}}{\Delta_N^2} \\
&\quad + 2f(Y_{\bullet}^{j-2}) \sigma(X_{(j-1)\Delta_N})^2 f(Y_{\bullet}^{j-1}) \sigma(X_{j\Delta_N})^2 \left( \frac{\zeta_{j+1,N}^2}{\Delta_N} - m_N \right) \left( \frac{\zeta'_{j+1,N}}{\Delta_N} - m'_N \right) \\
&\quad + 4f(Y_{\bullet}^{j-2}) \sigma(X_{(j-1)\Delta_N})^2 f(Y_{\bullet}^{j-1}) \sigma(X_{j\Delta_N})^2 \left( \frac{\zeta_{j+1,N}^2}{\Delta_N} - m_N \right) \frac{\zeta_{j,N} \zeta'_{j+1,N}}{\Delta_N} \\
&\quad + 4f(Y_{\bullet}^{j-2})^2 \sigma(X_{(j-1)\Delta_N})^4 \left( \frac{\zeta'_{j+1,N}}{\Delta_N} - m'_N \right) \frac{\zeta_{j,N} \zeta'_{j+1,N}}{\Delta_N}
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}((s_{j,N}^{(1)})^2|\mathcal{H}_j^N) &= 2f(Y_{\bullet}^{j-1})^2 \sigma(X_{j\Delta_N})^4 m_N^2 \\
&\quad + 2f(Y_{\bullet}^{j-2})^2 \sigma(X_{(j-1)\Delta_N})^4 m'_N{}^2 \\
&\quad + 4f(Y_{\bullet}^{j-2})^2 \sigma(X_{(j-1)\Delta_N})^4 \frac{\zeta_{j,N}^2}{\Delta_N} m'_N \\
&\quad + 4f(Y_{\bullet}^{j-2}) \sigma(X_{(j-1)\Delta_N})^2 f(Y_{\bullet}^{j-1}) \sigma(X_{j\Delta_N})^2 \chi_N^2,
\end{aligned}$$

Lemma 4 and (32) imply

$$\frac{1}{k_N} \sum_{j=2}^{k_N-2} \mathbb{E}((s_{j,N}^{(1)})^2|\mathcal{H}_j^N) \longrightarrow \nu_0(f^2\sigma^4)$$

in probability. The fourth conditional moment  $\mathbb{E}((s_{j,N}^{(1)})^4|\mathcal{H}_j^N)$  is bounded with Proposition 2 and

$$\frac{1}{k_N^2} \sum_{j=2}^{k_N-2} \mathbb{E}((s_{j,N}^{(1)})^4|\mathcal{H}_j^N) \longrightarrow 0$$

in probability. Theorem 3.2 in Hall and Heyde (1980) can be applied to obtain

$$U_N^{(1)} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu_0(f^2\sigma^4)).$$

Considering now  $U_N^{(2)} = \frac{1}{\sqrt{k_N}} \sum_{j=1}^{k_N-2} u_{j,N}^{(2)}$ , with Proposition 1,

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

and with  $k_N \Delta_N^2 = N \delta_N^{2-\frac{1}{\alpha}}$ ,  $U_N^{(2)} = o_P(1)$ .

The third term is

$$U_N^{(3)} = \frac{1}{\sqrt{k_N}} \sum_{j=1}^{k_N-2} u_{j,N}^{(3)} = \frac{1}{\sqrt{k_N}} \sum_{j=2}^{k_N-2} s_{j,N}^{(3)} + o_P(1)$$

where

$$s_{j,N}^{(3)} = (f(Y_{\bullet}^{j-2}) + f(Y_{\bullet}^{j-1})) \rho_N^2 \left( \frac{(\varepsilon_{\bullet}^j)^2}{\Delta_N} - \Delta_N^{\frac{2-\alpha}{\alpha-1}} \right) - 2f(Y_{\bullet}^{j-2}) \rho_N^2 \frac{\varepsilon_{\bullet}^{j-1} \varepsilon_{\bullet}^j}{\Delta_N}$$

is  $\mathcal{H}_{j+1}^N$  measurable. The conditional centering holds:  $\mathbb{E}(s_{j,N}^{(3)} | \mathcal{H}_j^N) = 0$  and

$$\mathbb{E}((s_{j,N}^{(3)})^2 | \mathcal{H}_j^N) = 2(f(Y_{\bullet}^{j-2}) + f(Y_{\bullet}^{j-1}))^2 \frac{\rho_N^4}{\Delta_N^2 p_N^2} + 4f(Y_{\bullet}^{j-1})^2 \rho_N^4 \frac{(\varepsilon_{\bullet}^{j-1})^2}{\Delta_N^2 p_N}.$$

Then, two cases have to be considered:

1. If  $\alpha \in (1, 2)$  or  $\rho_N \rightarrow 0$  and  $\alpha = 2$ , from the previous computation

$$\frac{1}{k_N} \sum_{j=2}^{k_N-2} \mathbb{E}((s_{j,N}^{(3)})^2 | \mathcal{H}_j^N) = o_P(1),$$

2. If  $\alpha = 2$  and  $\rho_N = \rho$ ,

$$\frac{1}{k_N} \sum_{j=2}^{k_N-2} \mathbb{E}((s_{j,N}^{(3)})^2 | \mathcal{H}_j^N) = 12\nu_0(f^2 \rho^4) + o_P(1).$$

In the case 1, Lemma 4 implies that  $U_N^{(3)} = o_P(1)$ , but in the case 2, a Central Limit Theorem for martingale array is necessary: using

$$\frac{1}{k_N^2} \sum_{j=2}^{k_N-2} \mathbb{E}((s_{j,N}^{(3)})^4 | \mathcal{H}_j^N) = o_P(1),$$

we derive

$$U_N^{(3)} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 12\nu_0(f^2 \rho^4)).$$

Notice that, in this case,  $\mathbb{E}(s_{j,N}^{(1)} s_{j,N}^{(3)} | \mathcal{H}_j^N) = 0$ , and then  $U_N^{(1)}$  and  $U_N^{(3)}$  are asymptotically uncorrelated.

Setting

$$\begin{aligned} s_{j,N}^{(4)} &= 2f(Y_{\bullet}^{j-2}) \rho_N \sigma(X_{(j-1)\Delta_N}) \frac{1}{\Delta_N} (\zeta_{j,N} \varepsilon_{\bullet}^j + \zeta'_{j+1,N} \varepsilon_{\bullet}^j - \zeta'_{j+1,N} \varepsilon_{\bullet}^{j-1}) \\ &\quad - 2f(Y_{\bullet}^{j-1}) \rho_N \sigma(X_{j\Delta_N}) \frac{\varepsilon_{\bullet}^j \zeta_{j+1,N}}{\Delta_N} \end{aligned}$$



where  $r_{j,N}^{(4)}$  is  $\mathcal{H}_{j+1}^N$  measurable, with  $\mathbb{E}(r_{j,N}^{(4)}|\mathcal{H}_j^N) = 0$ , the fourth term is reordered:

$$U_N^{(4)} = \frac{1}{\sqrt{k_N}} \sum_{j=1}^{k_N-2} u_{j,N}^{(4)} = \frac{1}{\sqrt{k_N}} \sum_{j=2}^{k_N-2} s_{j,N}^{(4)} + o_P(1).$$

The conditional variance is

$$\begin{aligned} \mathbb{E}\left((s_{j,N}^{(4)})^2 \middle| \mathcal{H}_j^N\right) &= 4f(Y_{\bullet}^{j-2})^2 \rho_N^2 \sigma(X_{(j-1)\Delta_N})^2 \left( \frac{\zeta_{j,N}^2}{\Delta_N} + m'_N + \frac{(\varepsilon_{\bullet}^{j-1})^2}{\Delta_N} m'_N \right) \\ &\quad + 4f(Y_{\bullet}^{j-1})^2 \rho_N^2 \sigma(X_{j\Delta_N})^2 m_N \\ &\quad - 8f(Y_{\bullet}^{j-1}) \sigma(X_{j\Delta_N}) f(Y_{\bullet}^{j-2}) \sigma(X_{(j-1)\Delta_N}) \rho_N^2 \chi_N. \end{aligned}$$

Hence, the same two situations have to be considered:

1. If  $\alpha \in (1, 2)$  or  $\rho_N \rightarrow 0$  and  $\alpha = 2$ ,

$$\frac{1}{k_N} \sum_{j=2}^{k_N-2} \mathbb{E}\left((s_{j,N}^{(4)})^2 \middle| \mathcal{H}_j^N\right) = o_P(1).$$

2. If  $\alpha = 2$  and  $\rho_N = \rho$ ,

$$\frac{1}{k_N} \sum_{j=2}^{k_N-2} \mathbb{E}\left((s_{j,N}^{(4)})^2 \middle| \mathcal{H}_j^N\right) = 4\nu_0(f^2 \rho^2 \sigma^2) + o_P(1).$$

Besides, in case 1, using Lemma 4,  $U_N^{(4)} = o_P(1)$ , whereas in case 2,

$$\frac{1}{k_N^2} \sum_{j=2}^{k_N-2} \mathbb{E}\left((s_{j,N}^{(4)})^4 \middle| \mathcal{H}_j^N\right) = o_P(1)$$

and the Central Limit Theorem for martingale array implies

$$U_N^{(4)} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\nu_0(f^2 \rho^2 \sigma^2)).$$

Finally,

$$\mathbb{E}(s_{j,N}^{(1)} s_{j,N}^{(4)} | \mathcal{H}_j^N) = 0, \quad \mathbb{E}(s_{j,N}^{(2)} s_{j,N}^{(4)} | \mathcal{H}_j^N) = 0$$

and  $U_N^{(1)}, U_N^{(3)}, U_N^{(4)}$  are asymptotically uncorrelated.

To finish the proof, let  $U_N^{(\ell)} = \frac{1}{\sqrt{k_N}} \sum_{j=1}^{k_N-2} u_{j,N}^{(\ell)}$  for  $\ell = 5, 6, 7$ . With Propositions 1, 2, and the Cauchy Schwarz inequality,

$$\begin{aligned} \mathbb{E}(|u_{j,N}^{(5)}| | \mathcal{H}_j^N) &\leq c|f(Y_{\bullet}^{j-1})| \Delta_N (1 + X_{j\Delta_N}^2 + \rho_N^2 \mathbb{E}((\varepsilon_{\bullet}^j)^2)) \\ &\quad \times (\Delta_N (1 + X_{j\Delta_N}^4 + \rho_N^2 \sqrt{\mathbb{E}((\varepsilon_{\bullet}^j)^4)}), \\ \mathbb{E}(|u_{j,N}^{(6)}| + |u_{j,N}^{(7)}| | \mathcal{H}_j^N) &\leq c|f(Y_{\bullet}^{j-1})| \Delta_N (1 + |X_{j\Delta_N}|^2 + \rho_N \sqrt{\mathbb{E}((\varepsilon_{\bullet}^j)^2)}) \\ &\quad \times (\sqrt{\Delta_N} (1 + X_{j\Delta_N}^2 + \rho_N \mathbb{E}((\varepsilon_{\bullet}^j)^4))^{\frac{1}{4}}). \end{aligned}$$

With  $k_N \Delta_N^2 \rightarrow 0$ , we have  $U_N^{(\ell)} = o_P(1)$  for  $\ell = 5, 6, 7$ .

□

**Proof of Theorem 3.** As we deal with martingale arrays, with the notations of Theorems 5.1 and 5.2, it is sufficient to remark that

$$\frac{1}{k_N} \sum_{j=1}^{k_N-2} \mathbb{E}(r_{j,N}^{(1)} s_{j,N}^{(1)} | \mathcal{H}_j^N) = o_P(1).$$

□

**Proof of Corollary 1.** As  $f_N(x) - f(x) = v_N(1 + |x|)$ , we have

$$\begin{aligned} |\bar{v}_N(f_N) - \bar{v}_N(f)| &\leq \frac{1}{k_N} \sum_{j=0}^{k_N-1} |f_N(Y_{\bullet}^j) - f(Y_{\bullet}^j)| \\ &\leq v_N \times \frac{1}{k_N} \sum_{j=0}^{k_N-1} (1 + |Y_{\bullet}^j|) \end{aligned}$$

and the last quantity tends to 0 in probability. It is also sufficient to consider the case  $f = 0$ .

Then, considering the proof of Theorem 1, and keeping the same notations for  $\bar{R}_N^{(\ell)}$ ,  $\ell = 1 \dots 4$ , we consider

$$\bar{R}_N^{(1)} = \frac{1}{\sqrt{k_N \Delta_N}} \sum_{j=1}^{k_N-2} f_N(Y_{\bullet}^{j-1}) \sigma(X_{j \Delta_N}) (\zeta_{j+1,N} + \zeta'_{j+2,N}).$$

Splitting the sum into three parts, using Lemma 4 and

$$\mathbb{E}(f(Y_{\bullet}^{3j-1})^2 \sigma(X_{j \Delta_N})^2 (\zeta_{j+1,N} + \zeta'_{j+2,N})^2 | \mathcal{H}_{3j}^N) \leq cv_N (1 + |Y_{\bullet}^{j-1}|^2) \sigma(X_{j \Delta_N})^2 \Delta_N,$$

$\bar{R}_N^{(1)}$  tends to 0 in probability. The same argument is used for the other terms in  $\sqrt{k_N \Delta_N} \bar{I}_N(f_N)$  and  $\sqrt{k_N \Delta_N} \bar{Q}_N(f_N)$ .

□

**Proof of Theorem 4.** Consider the estimating function

$$G_N(\theta) = \frac{1}{k_N \Delta_N} \sum_{j=1}^{k_N-2} g(\delta_N, Y_{\bullet}^{j+1} - Y_{\bullet}^j, Y_{\bullet}^{j-1}; \theta, \rho_N).$$

By Lemma 2, we have  $G_N(\theta_0) \rightarrow 0$  and  $\partial_{\theta} G_N(\theta) \rightarrow \phi(\theta, \theta_0, \rho_{\infty})$  in probability, under  $\mathbb{P}_{\theta_0}$ , uniformly in  $\theta \in \Theta$ . As  $\phi(\theta_0, \theta_0, \rho_{\infty}) = -S$  is invertible, this implies the existence and the consistency of  $\hat{\theta}_N$ , with standard arguments (see Bibby and Sørensen (1995), Theorem 3.2).

A Taylor's formula around  $\hat{\theta}_N$  shows:

$$G_N(\hat{\theta}_N) - G_N(\theta_0) = \left( \int_0^1 \partial_{\theta} G_N(\theta_0 + u(\hat{\theta}_N - \theta_0)) du \right) (\hat{\theta}_N - \theta_0).$$

As  $\partial_\theta G_N(\theta)$  converges uniformly in probability to  $\phi(\theta, \theta_0, \rho_\infty)$  and  $\hat{\theta}_N$  converges in probability to  $\theta_0$ ,

$$\int_0^1 \partial_\theta G_N(\theta_0 + u(\hat{\theta}_N - \theta_0)) du \xrightarrow{\mathbb{P}_{\theta_0}} \phi(\theta_0, \theta_0, \rho_\infty) = -S.$$

With a second Taylor's formula around  $Y_\bullet^{j+1} - Y_\bullet^j$ :

$$\begin{aligned} g(\delta_N, Y_\bullet^{j+1} - Y_\bullet^j, Y_\bullet^{j-1}; \theta_0, \rho_N) &= (Y_\bullet^{j+1} - Y_\bullet^j) \partial_y g(0, 0, Y_\bullet^{j-1}; \theta_0, \rho_N) \\ &+ \frac{1}{2} (Y_\bullet^{j+1} - Y_\bullet^j)^2 \partial_{y^2}^2 g(0, 0, Y_\bullet^{j-1}; \theta_0, \rho_N) \\ &+ \Delta_N g^{(1)}(0, Y_\bullet^{j-1}, \theta_0, \rho_N) \\ &+ \Delta_N^{\frac{1}{\alpha-1}} g^{(\alpha)}(0, Y_\bullet^{j-1}; \theta_0, \rho_N) \\ &+ \Delta_N^2 R(\Delta_N, Y_\bullet^{j+1} - Y_\bullet^j, Y_\bullet^{j-1}; \theta_0, \rho_N). \end{aligned}$$

Then, with Lemma 1,

$$\begin{aligned} g(\delta_N, Y_\bullet^{j+1} - Y_\bullet^j, Y_\bullet^{j-1}; \theta_0, \rho_N) &= (Y_\bullet^{j+1} - Y_\bullet^j) \partial_y g(0, 0, Y_\bullet^{j-1}; \theta_0, \rho_N) \\ &- \Delta_N b(Y_\bullet^{j-1}, \kappa_0) \partial_y g(0, 0, Y_\bullet^{j-1}; \theta_0, \rho_N) \\ &+ \frac{1}{2} (Y_\bullet^{j+1} - Y_\bullet^j)^2 \partial_{y^2}^2 g(0, 0, Y_\bullet^{j-1}, \theta_0, \rho_N) \\ &- \frac{\Delta_N}{3} \partial_{y^2}^2 g(0, 0, Y_\bullet^{j-1}, \theta_0, \rho_N) \\ &- \Delta_N^{\frac{1}{\alpha-1}} \rho_N^2 \partial_{y^2}^2 g(0, 0, Y_\bullet^{j-1}, \theta_0, \rho_N) \\ &+ \Delta_N^2 R(\Delta_N, Y_\bullet^{j+1} - Y_\bullet^j, Y_\bullet^{j-1}; \theta_0, \rho_N). \end{aligned}$$

Thus,

$$\begin{aligned} &g(\delta_N, Y_\bullet^{j+1} - Y_\bullet^j, Y_\bullet^{j-1}; \theta_0, \rho_N) = \\ &\left( \begin{array}{c} \partial_y g_1(0, 0, Y_\bullet^{j-1}; \theta_0, \rho_N) (Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1}, \kappa_0)) \\ 0 \end{array} \right) \\ &+ \left( \begin{array}{c} \partial_{y^2}^2 g_1(0, 0, Y_\bullet^{j-1}; \theta_0, \rho_N) (\frac{1}{2} (Y_\bullet^{j+1} - Y_\bullet^j)^2 - \frac{\Delta_N}{3} c(Y_\bullet^{j-1}, \lambda_0) - \Delta_N^{\frac{1}{\alpha-1}} \rho_N^2) \\ \partial_{y^2}^2 g_2(0, 0, Y_\bullet^{j-1}; \theta_0, \rho_N) (\frac{1}{2} (Y_\bullet^{j+1} - Y_\bullet^j)^2 - \frac{\Delta_N}{3} c(Y_\bullet^{j-1}, \lambda_0) - \Delta_N^{\frac{1}{\alpha-1}} \rho_N^2) \\ + \Delta_N^2 R(\Delta_N, Y_\bullet^{j+1} - Y_\bullet^j, Y_\bullet^{j-1}; \theta_0, \rho_N). \end{array} \right) \end{aligned}$$

The results comes from Theorem 3.  $\square$

**Proof of Theorem 5.** With the notations of the proof of Theorem 5.2, it is sufficient to remark that  $u_{3j,N}^{(1)}$  is  $\mathcal{H}_{3j+2}^N$  measurable, conditionally centered, and

$$\mathbb{E}((u_{3j,N}^{(1)})^2 | \mathcal{H}_{3j}^N) = \frac{8}{9} \nu_0 (f^2 \sigma^4) + o_P(1).$$

Then, the correlation between  $\zeta_{j+1,N}$  and  $\zeta'_{j+1,N}$  is avoided, and the asymptotic variance is reduced in the Central Limit Theorem.  $\square$

## References

- Bibby, B. M. and Sørensen, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, 1(1-2):17–39.
- Favetto, B. (2010). Consistent parameter estimation by contrast minimization for noisy discrete observations of a hidden diffusion process. MAP5 2010-13.
- 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.
- 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.
- Jacobsen, M. (2002). Optimality and small  $\Delta$ -optimality of martingale estimating functions. *Bernoulli*, 8(5):643–668.
- Kessler, M. (2000). Simple and explicit estimating functions for a discretely observed diffusion process. *Scand. J. Statist.*, 27(1):65–82.
- Kessler, M. and Sørensen, M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli*, 5(2):299–314.
- Sørensen, M. (2009). Efficient estimation for ergodic diffusions sampled at high frequency. Preprint.
- Sørensen, M. (2010). Estimating functions for diffusion-type processes. Preprint.

## Appendix

Consider the following random variables which will appear in some expansions,

$$\xi_{j,N} = \frac{1}{\Delta_N^{3/2}} = \int_{j\Delta_N}^{(j+1)\Delta_N} (s - j\Delta_N) dB_s, \quad (17)$$

$$\xi'_{j+1,N} = \frac{1}{\Delta_N^{3/2}} = \int_{(j+1)\Delta_N}^{(j+2)\Delta_N} ((j+2)\Delta_N - s) dB_s. \quad (18)$$

Analogously, define:

$$\xi_{i,j,N} = \frac{1}{\delta_N^{3/2}} \int_{j\Delta_N+i\delta_N}^{j\Delta_N+(i+1)\delta_N} (s - j\Delta_N - i\delta_N) dB_s, \quad (19)$$

$$\xi'_{i+1,j,N} = \frac{1}{\delta_N^{3/2}} \int_{j\Delta_N+(i+1)\delta_N}^{j\Delta_N+(i+2)\delta_N} (j\Delta_N + (i+2)\delta_N - s) dB_s. \quad (20)$$

Some basic properties of these random variables are summarized in Lemma 5 in the Appendix.

Define the following random variables:

$$\zeta_{j+1,N} = \frac{1}{p_N} \sum_{i=0}^{p_N-1} \int_{j\Delta_N+i\delta_N}^{(j+1)\Delta_N} dB_s, \quad (21)$$

$$\zeta'_{j+2,N} = \frac{1}{p_N} \sum_{i=0}^{p_N-1} \int_{(j+1)\Delta_N}^{(j+1)\Delta_N+i\delta_N} dB_s. \quad (22)$$

Notice that

$$\zeta_{j+1,N} = \frac{1}{p_N} \sum_{k=0}^{p_N-1} (k+1) \int_{j\Delta_N+k\delta_N}^{j\Delta_N+(k+1)\delta_N} dB_s, \quad (23)$$

$$\zeta'_{j+2,N} = \frac{1}{p_N} \sum_{k=0}^{p_N-1} (p_N - 1 - k) \int_{(j+1)\Delta_N+k\delta_N}^{(j+1)\Delta_N+(k+1)\delta_N} dB_s, \quad (24)$$

and that

$$\frac{1}{p_N} \sum_{k=0}^{p_N-1} (k+1)\delta_N = \frac{\Delta_N}{2} + \frac{\delta_N}{2p_N}, \quad (25)$$

$$\frac{1}{p_N} \sum_{k=0}^{p_N-1} (p_N - k - 1)\delta_N = \frac{\Delta_N}{2} - \frac{\delta_N}{2p_N}. \quad (26)$$

**Lemma 3.** *The random variables  $\zeta_{j+1,N}$  and  $\zeta'_{j+1,N}$  are  $\mathcal{G}_{(j+1)\Delta_N}$  measurable, and  $\zeta'_{j+2,N}$  is independent of  $\mathcal{G}_{(j+1)\Delta_N}$ . Moreover, we have*

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

**Lemma 4.** Let  $\chi_j^N, U$  be random variables, with  $\chi_j^N$  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  $\sum_{j=0}^{k_N-1} \chi_j^N \xrightarrow{\mathbb{P}} U$ .

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

$$Y_{\bullet}^j - X_{j\Delta_N} = \sigma(X_{j\Delta_N}) \sqrt{\Delta_N} \xi'_{j,N} + e'_{j,N} + \rho_N \varepsilon_{\bullet}^j \quad (27)$$

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  $j \leq k_N - 1$ :

$$\mathbb{E}\left(|Y_{\bullet}^j - X_{j\Delta_N}|^k | \mathcal{H}_j^N\right) \leq C \left(\Delta_N^{k/2}(1 + |X_{j\Delta_N}|^k) + \rho_N^k \mathbb{E}\left(|\varepsilon_{\bullet}^j|^k\right)\right). \quad (28)$$

**Lemma 5.** 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:

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

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

**Theorem 7.** 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

$$\bar{Q}_N(f(\cdot, \theta)) = \frac{1}{k_N \Delta_N} \sum_{j=1}^{k_N-2} f(Y_{\bullet}^{j-1}, \theta) (Y_{\bullet}^{j+1} - Y_{\bullet}^j)^2 \xrightarrow{\mathbb{P}} \frac{2}{3} \nu_0(f(\cdot, \theta) \sigma^2), \quad (29)$$

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

$$\bar{Q}_N(f(\cdot, \theta)) \xrightarrow{\mathbb{P}} \frac{2}{3} \nu_0(f(\cdot, \theta) \sigma^2) + 2\rho^2 \nu_0(f(\cdot, \theta)), \quad (30)$$

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

$$\bar{Q}_N(f(\cdot, \theta)) \xrightarrow{\mathbb{P}} \frac{2}{3} \nu_0(f(\cdot, \theta) \sigma^2), \quad (31)$$

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  $N\delta_N \rightarrow \infty$ .

**Lemma 6.** 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:

$$\frac{1}{k_N} \sum_{j=0}^{k_N-1} f(X_{j\Delta_N}, \theta) \xrightarrow[k_N \rightarrow \infty]{} \nu_0(f(\cdot, \theta)) \quad (32)$$

uniformly in  $\theta$ , in probability.

In Favetto (2010), several properties of  $Y_\bullet^j$  and some results of convergence in probability for functionals of the blocks have been established. Recall that the random variables  $\zeta_{j,N}$  and  $\zeta'_{j+1,N}$  are defined in (4.8), and Lemma 4 is used to establish convergence in probability results.

**Proposition 2.** Under Assumptions **(A1)**-**(A2)** and **(A5)**, we have

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

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

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