© Board of the Foundation of the Scandinavian Journal of Statistics 2006. Published by Blackwell Publishing Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA Vol 33: 83–104, 2006

Parameter Estimation for a Discretely Observed Integrated Diffusion Process

ARNAUD GLOTER

Département d'Economie, Université Bordeaux 4

ABSTRACT. We consider the estimation of unknown parameters in the drift and diffusion coefficients of a one-dimensional ergodic diffusion X when the observation is a discrete sampling of the integral of X 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 first prove limit theorems for functionals associated with our observations. We apply these results to obtain a contrast function. The associated minimum contrast estimators are shown to be consistent and asymptotically Gaussian with different rates for drift and diffusion coefficient parameters.

Key words: contrast function, diffusion processes, discrete time observations, non-Markovian process, parametric inference

1. Introduction

The statistics of one-dimensional diffusion processes with ergodic properties and when the sample path is discretely observed has been the subject of many recent papers. More precisely, let (X_t) be given by the stochastic differential equation:

$$\mathrm{d}X_t = b(X_t,\mu)\,\mathrm{d}t + a(X_t,\sigma)\,\mathrm{d}B_t, \quad X_0 = \eta \tag{1}$$

with B a standard Wiener process and η a random variable independent of B. Suppose that for some positive Δ, a sample $(X_{i\Delta}, i \le n)$ is observed and that it is required to estimate $(\mu,\sigma) \in \mathbb{R}^2$. The exact likelihood of such an observation being generally intractable, other methods have been developed to obtain explicit estimators. Under the assumption of fixed sampling interval, different kinds of estimating functions have been studied (see e.g. Barndorff-Nielsen & Sørensen, 1994; Kessler & Sørensen, 1999; Sørensen, 1999, 2000; Kessler, 2000; Bibby & Sørensen, 2001).

Another point of view which is also classical and complementary to the former one is to assume that the sampling interval $\Delta = \Delta_n$ tends to 0 as $n \to \infty$ and $n\Delta_n \to \infty$. In this framework, the likelihood of the Euler scheme of (1) is a contrast function and provides consistent and asymptotically Gaussian estimators. A noteworthy result is that drift and diffusion coefficient parameters are estimated with different rates, $(n\Delta_n)^{1/2}$ for the drift parameters and $n^{1/2}$ for diffusion coefficient parameters (see e.g. Dorogovtsev, 1976; Florens-Zmirou, 1989; Kessler, 1997).

In this paper, we consider a new type of observation. Our aim is to estimate the parameter (μ,σ) of (1) when we observe a discrete Δ-sampling of the integrated process $I_t = \int_0^t X_s\,\mathrm{d}s$.

Integrals of diffusion process have been recently considered in the field of finance in relation to stochastic volatility models (see e.g. Leblanc, 1996; Barndorff-Nielsen, 1998; Genon-Catalot *et al.*, 1998; Barndorff-Nielsen & Sheppard, 2002; Bollerslev & Zhou, 2002). Data may be obtained from option prices and their associated implied volatilities (see e.g. Pastorello *et al.*, 1994)

For fixed sampling interval Δ, the exact distribution of $(I_{i\Delta}, i \le n)$ is difficult to compute except for few models. For example, the case of X a stationary Ornstein–Uhlenbeck is treated in Gloter (2001) and the case of a Cox–Ingersoll–Ross model is studied in Ditlevsen & Sørensen (2004).

To deal with a general diffusion X, we shall assume that the sampling interval $\Delta=\Delta_n$ tends to 0. Now, let

$$\overline{X}_i^n=\Delta_n^{-1}\int_{i\Delta_n}^{(i+1)\Delta_n} X_s\,\mathrm{d}s=\Delta_n^{-1}(I_{(i+1)\Delta_n}-I_{i\Delta_n}). \tag{2}$$

We shall base our estimation on the sample $(\overline{X}_i^n)_{i\leq n-1}$ which is in one-to-one correspondence with the observation $(I_{i\Delta_n})_{i\leq n}$. Our starting idea is that, for small Δ_n, the law of $(\overline{X}_i^n, i\leq n-1)$ may be close to the law of $(X_{i\Delta_n}, i\leq n-1)$, so that methods available for the discrete sampling $(X_{i\Delta_n})$ apply for $(\overline{X}_i^n)$. In fact such a substitution fails mainly because $(\overline{X}_i^n)$ is not Markovian.

As preliminary steps, in Gloter (2000) we have obtained asymptotic expansions of $\overline{X}_i^n$ as $\Delta_n\to 0$. The results of Gloter (2000) hold without ergodicity assumptions on model (1).

In this paper, we assume that the diffusion (X_t) has ergodic properties with invariant probability $\mathrm{d}\nu_0(x)$. We first prove limit theorems concerning the variation and the quadratic variation of $(\overline{X}_i^n)$, which enlight the difference between $(\overline{X}_i^n)$ and the discrete sampling $(X_{i\Delta_n})$. These theorems enable us to construct a contrast by introducing the appropriate corrections on the Euler contrast of the diffusion.

The paper is organized as follows. Sections 2–4 are devoted to general limit theorems. In these sections, parameters are omitted and we set $a(x,\sigma)=a(x)$, $b(x,\mu)=b(x)$. Section 2 contains the assumptions and a recap of some expansions obtained in Gloter (2000). In section 3, we introduce the following functionals of the observed process (where $\overline{X}_i=\overline{X}_i^n$ to simplify notations):

$$\overline{\nu}_n(f)=n^{-1}\sum_{i=0}^{n-2}f(\overline{X}_i) \tag{3}$$

$$\overline{\mathcal{I}}_n(f)=(n\Delta_n)^{-1}\sum_{i=0}^{n-2}f(\overline{X}_i)(\overline{X}_{i+1}-\overline{X}_i-\Delta_n b(\overline{X}_i)) \tag{4}$$

$$\overline{\mathcal{Q}}_n(f)=(n\Delta_n)^{-1}\sum_{i=0}^{n-2}f(\overline{X}_i)(\overline{X}_{i+1}-\overline{X}_i)^2. \tag{5}$$

Section 3 contains convergence in probability results and section 4 some associated central limit theorems. The main result is that, under smoothness assumptions on f, g,

$$\left((n\Delta_n)^{1/2}\left(\overline{\mathcal{I}}_n(f)-\frac{1}{4}\overline{\mathcal{Q}}_n(f')\right), n^{1/2}\left(\frac{3}{2}\overline{\mathcal{Q}}_n(g)-\overline{\nu}_n(ga^2)\right)\right)$$

converges in distribution to a $\mathcal{N}(0,\nu_0(f^2a^2))\otimes\mathcal{N}(0,9/4\nu_0(g^2a^4))$ (where ν_0 is the invariant probability of the diffusion and $\nu_0(f)=\int f(x)\,\mathrm{d}\nu_0(x)$). This needs the additional (but classical, see e.g. Florens-Zmirou, 1989) condition $n\Delta_n^2\to 0$.

In section 5, we give examples of diffusion models satisfying the set of assumptions introduced in section 2.

Section 6 contains the statistical applications. Using results of sections 3 and 4, we define a contrast function, see (51), and study the associated minimum contrast estimator $\hat{\theta}_n=(\hat{\mu}_n,\hat{\sigma}_n)$. This estimator is shown to be consistent and $((n\Delta_n)^{1/2}(\hat{\mu}_n-\mu_0), n^{1/2}(\hat{\sigma}_n-\sigma_0))$ asymptotically normal ($\theta_0=(\mu_0,\sigma_0)$ denotes the true value of the parameter). We compare the asymptotic variances with those obtained for estimators based on a discrete sampling of the diffusion itself (by Kessler, 1997). The only difference is a slight increase in the asymptotic variance of the estimator of σ_0. In section 7, examples of parametric models are fully treated. We provide simulation results to see how our estimator behave on finite sample for small, but fixed, value of Δ_n.

Some technical results are given in the appendix.

To simplify notations and proofs we have chosen a one-dimensional parameter for the drift and for the diffusion coefficient, but we could easily extend this work to multidimensional parameters (see remark 5).

2. Assumptions and preliminary results

2.1. Model and assumptions

Let (X_t) be defined as the solution on a probability space $(\Omega, \mathcal{F}, P)$ of the stochastic differential equation:

$$dX_t = a(X_t)\,dB_t + b(X_t)\,dt, \quad X_0 = \eta \tag{6}$$

where $(B_t)_{t\geq 0}$ is a standard one-dimensional Brownian motion, η is a random variable independent of $(B_t)_{t\geq 0}$. We make now some classical assumptions on functions b and a ensuring that the solution of (6) is a positive recurrent diffusion on an interval $(l, r)(-\infty \leq l < r \leq \infty)$.

To keep general notations, we introduce two positive measurable functions $\mathcal{B}_l$ and $\mathcal{B}_r$ defined on (l, r) satisfying the following property: $\forall \alpha, \beta, \alpha', \beta', p \geq 0, \exists c, \forall x \in (l, r)$,

$$(\mathcal{B}_l^{\alpha}(x) + \mathcal{B}_r^{\beta}(x)) \times (\mathcal{B}_l^{\alpha'}(x) + \mathcal{B}_r^{\beta'}(x)) \leq c(\mathcal{B}_l^{\alpha+\alpha'}(x) + \mathcal{B}_r^{\beta+\beta'}(x)) \tag{7}$$

$$(\mathcal{B}_l^{\alpha}(x) + \mathcal{B}_r^{\beta}(x))^p \leq c(\mathcal{B}_l^{p\alpha}(x) + \mathcal{B}_r^{p\beta}(x)). \tag{8}$$

These functions are used below to bound the growth of other functions near the boundaries l, r of the state space. For example, if $l = 0, r = \infty$ we may take $\mathcal{B}_l(x) = 1 + x^{-1}, \mathcal{B}_r(x) = 1 + x$.

(A0) Equation (6) admits a unique strong solution such that $P(X_t \in (l, r), \forall t \geq 0) = 1$.
(A1) Function a and b are real valued, $\mathcal{C}^2$ on (l, r) and

$$\exists c, \alpha_1, \alpha_2, \beta_1, \beta_2 \geq 0, \forall x \in (l, r),$$
$$a(x) > 0, \quad |a(x)| + |b(x)| \leq c(1 + \mathcal{B}_r(x)),$$
$$|a'(x)| \leq c(\mathcal{B}_l^{\alpha_1}(x) + \mathcal{B}_r^{\alpha_1}(x)), \quad |a''(x)| \leq c(\mathcal{B}_l^{\alpha_2}(x) + \mathcal{B}_r^{\alpha_2}(x)),$$
$$|b'(x)| \leq c(\mathcal{B}_l^{\beta_1}(x) + \mathcal{B}_r^{\beta_1}(x)), \quad |b''(x)| \leq c(\mathcal{B}_l^{\beta_2}(x) + \mathcal{B}_r^{\beta_2}(x)).$$

Let $\mathcal{G}_t = \sigma(B_s, s \leq t; \eta)$.

(A2) There exists a positive constant K_l such that

$$\forall k \in [0, K_l), \exists c, \forall t > 0, E\left(\sup_{s\in[t,t+1]} \mathcal{B}_l^k(X_s) \,|\, \mathcal{G}_t\right) \leq c\mathcal{B}_l^k(X_t)$$
$$\forall k \in [0, \infty), \exists c, \forall t > 0, E\left(\sup_{s\in[t,t+1]} \mathcal{B}_r^k(X_s) \,|\, \mathcal{G}_t\right) \leq c\mathcal{B}_r^k(X_t).$$

For $x_0 \in (l, r)$, let $s(x) = \exp(-2\int_{x_0}^x (b(u)/a^2(u))\,du)$ denote the scale density and $m(x) = 1/(a^2(x)s(x))$ the speed density.

(A3) $\int_l^{x_0} s(x)\,dx = \int_{x_0}^r s(x)\,dx = \infty$, $\int_l^r m(x)\,dx = M < \infty$.
Let

$$\nu_0(dx) = \frac{1}{M} m(x) 1\!\!1_{\{x\in(l,r)\}}\,dx.$$

(A4) $\exists M_l, M_r > 0$ $\nu_0(\mathcal{B}_l^{M_l}) < \infty$, $\nu_0(\mathcal{B}_r^{M_r}) < \infty$.
(A5) $\sup_{t\geq 0} E(\mathcal{B}_l^{M_l}(X_t)) < \infty$, $\sup_{t\geq 0} E(\mathcal{B}_r^{M_r}(X_t)) < \infty$ (where M_l and M_r are defined in A4).

(A1) and (A3) imply (A0) but some results hold without (A3) under (A0)–(A2). Under (A1) and (A3), ν_0 is the unique invariant probability of model (6) and X satisfies the classical ergodic theorem

$$\forall f \in \mathbf{L}^1(\mathrm{d}\nu_0),\ \frac{1}{T}\int_0^T f(X_s)\,\mathrm{d}s \xrightarrow[a.s.]{T\to\infty} \nu_0(f).$$

Assumption (A4) means that some powers of $\mathcal{B}_l$ and $\mathcal{B}_r$ are in $\mathbf{L}^1(\mathrm{d}\nu_0)$ and hence all functions on (l,r) bounded by these powers of $\mathcal{B}_l$ and $\mathcal{B}_r$ will also be in $\mathbf{L}^1(\mathrm{d}\nu_0)$. Assumption (A5) follows immediately from (A4) if the initial distribution is ν_0 (the process X being strictly stationary). In section 5, we prove that (A5) follows again from (A4) when the initial condition is deterministic.

Assumption (A2) was already introduced in Kessler (1997) and in Gloter (2000), as a useful tool to control the behaviour of X near the endpoints l, r. In Gloter (2000), it is shown that (A2) holds for general diffusion processes on $\mathbb{R}$, and for some classical diffusions on $(0,\infty)$. The reason why (A2) is not symmetric appears in the examples of section 5.

2.2. Expansions for the observed process

Now, let Δ_n be a sequence of positive numbers with $\Delta_n \to 0$, as $n\to\infty$ and assume, for convenience, that $\Delta_n \le 1$, for all n. We set $\mathcal{G}_i^n = \mathcal{G}_{i\Delta_n}$.

We now recall the main properties of $\overline{X}_i^n = \overline{X}_i$ (see (2)) proved in Gloter (2000). In the following statements the constants c appearing never depend on i or n.

Proposition 1

Assume (A0)–(A2) and let $f \in \mathcal{C}^1(l,r)$ satisfy:

$$\exists\gamma \ge 0, \exists c>0, \forall x \in (l,r)\ |f'(x)| \le c(\mathcal{B}_l^{\gamma}(x) + \mathcal{B}_r^{\gamma}(x)).$$

1. *For all integer $k\ge 1$, such that $k\gamma < K_l$ (with K_l given in (A2)), there exists $c>0$ such that for all $i,n\ge 0$*

$$E\left(\sup_{v\in[i\Delta_n,\,(i+1)\Delta_n]} |f(X_v) - f(X_{i\Delta_n})|^k \mid \mathcal{G}_i^n\right) \le c\Delta_n^{k/2}(\mathcal{B}_l^{k\gamma}(X_{i\Delta_n}) + \mathcal{B}_r^{k(1+\gamma)}(X_{i\Delta_n})).$$

2. *For all $k\ge 1$, there exists $c>0$ such that for all $i,n\ge 0$,*

$$E\left(\left|\overline{X}_i - X_{i\Delta_n}\right|^k \mid \mathcal{G}_i^n\right) \le c\Delta_n^{k/2}(1+\mathcal{B}_r^k(X_{i\Delta_n}))$$
$$E\left(\left|\overline{X}_{i+1} - \overline{X}_i\right|^k \mid \mathcal{G}_i^n\right) \le c\Delta_n^{k/2}(1+\mathcal{B}_r^k(X_{i\Delta_n})).$$

Let us introduce

$$\xi_{i,n} = \Delta_n^{-3/2}\int_{i\Delta_n}^{(i+1)\Delta_n} (s - i\Delta_n)\,\mathrm{d}B_s \quad \text{for } i, \quad n\ge 0 \tag{9}$$

$$\xi'_{i+1,n} = \Delta_n^{-3/2}\int_{(i+1)\Delta_n}^{(i+2)\Delta_n} (i\Delta_n + 2\Delta_n - s)\,\mathrm{d}B_s \quad \text{for } i\ge -1, \quad n\ge 0 \tag{10}$$

$$U_{i,n} = \xi_{i,n} + \xi'_{i+1,n}. \tag{11}$$

For all $n\ge 0$, $(\xi_{i,n})_{i\ge 0}$, $(\xi'_{i+1,n})_{i\ge 0}$ and $(U_{i,n})_{i\ge 0}$ are Gaussian processes; $\xi_{i,n}$ is $\mathcal{G}_{i+1}^n$ measurable and independent of $\mathcal{G}_i^n$; $\xi'_{i+1,n}$ is $\mathcal{G}_{i+2}^n$ measurable and independent of $\mathcal{G}_{i+1}^n$. We easily compute the following expectations: $E\left(\xi_{i,n} \mid \mathcal{G}_i^n\right) = E\left(\xi'_{i+1,n} \mid \mathcal{G}_{i+1}^n\right) = 0$, $E\left(\xi_{i,n}^2 \mid \mathcal{G}_i^n\right) = E\left(\xi'^2_{i+1,n} \mid \mathcal{G}_{i+1}^n\right) = 1/3$, $E\left(\xi_{i,n}\xi'_{i,n} \mid \mathcal{G}_i^n\right) = 1/6$. We deduce that for $i\ge 0$, $\mathrm{var}(U_{i,n}) = 2/3$ and $\mathrm{cov}(U_{i,n}, U_{i+1,n}) = 1/6$, $\mathrm{cov}(U_{i,n}, U_{i+j,n}) = 0$ for $j\ge 2$. Hence $(U_{i,n})_{i\ge 0}$ has the covariance structure of an MA(1) process.

The following results hold.

Theorem 1

Assume (A0)–(A2).

1. *We have*

$$\overline{X}_i - X_{i\Delta_n} = \Delta_n^{1/2} a(X_{i\Delta_n})\xi'_{i,n} + e_{i,n},$$

with, if $2\alpha_1 < K_l, \forall i, n \geq 0$ $\quad \left|E\left(e_{i,n} \mid \mathcal{G}_i^n\right)\right| \leq \Delta_n c(1 + \mathcal{B}_r(X_{i\Delta_n}))$ (12)

$$E\left(e_{i,n}^2 \mid \mathcal{G}_i^n\right) \leq \Delta_n^2 c(\mathcal{B}_l^{2\alpha_1}(X_{i\Delta_n}) + \mathcal{B}_r^{2(1+\alpha_1)}(X_{i\Delta_n})). \quad (13)$$

2. *We have*

$$\overline{X}_{i+1} - \overline{X}_i - b(\overline{X}_i)\Delta_n = \Delta_n^{1/2} a(X_{i\Delta_n}) U_{i,n} + \varepsilon_{i,n}$$

where $\varepsilon_{i,n}$ *is* $\mathcal{G}_{i+2}^n$ *measurable, and if* $\beta_1 \vee 2\beta_2 \vee 4\alpha_1 < K_l$, $\forall i, n \geq 0$

$$\left|E\left(\varepsilon_{i,n} \mid \mathcal{G}_i^n\right)\right| \leq \Delta_n^2 c(\mathcal{B}_l^{\beta_1 \vee \beta_2}(X_{i\Delta_n}) + \mathcal{B}_r^{(1+\beta_1)\vee(2+\beta_2)}(X_{i\Delta_n})), \quad (14)$$

if $4\alpha_1 \vee 2\alpha_2 < K_l$, $\forall i, n \geq 0$

$$E\left(\varepsilon_{i,n}^2 \mid \mathcal{G}_i^n\right) \leq \Delta_n^2 c(\mathcal{B}_l^{2\alpha_1 \vee \alpha_2}(X_{i\Delta_n}) + \mathcal{B}_r^{3+2\alpha_1+\alpha_2}(X_{i\Delta_n})) \quad (15)$$

$$E\left(\varepsilon_{i,n}^4 \mid \mathcal{G}_i^n\right) \leq \Delta_n^4 c(\mathcal{B}_l^{4\alpha_1 \vee 2\alpha_2}(X_{i\Delta_n}) + \mathcal{B}_r^{6+4\alpha_1+2\alpha_2}(X_{i\Delta_n})) \quad (16)$$

Furthermore, if $4\alpha_1 \vee \alpha_2 \vee 4\beta_1 < K_l$, $\forall i, n \geq 0$

$$\left|E\left(\varepsilon_{i,n}\xi_{i,n} \mid \mathcal{G}_i^n\right)\right| \leq \Delta_n^{3/2} c(\mathcal{B}_l^{(\alpha_1+\beta_1)\vee\alpha_2}(X_{i\Delta_n}) + \mathcal{B}_r^{(1+\alpha_1+\beta_1)\vee(2+\alpha_2)}(X_{i\Delta_n})) \quad (17)$$

$$\left|E\left(\varepsilon_{i,n}\xi'_{i+1,n} \mid \mathcal{G}_i^n\right)\right| \leq \Delta_n^{3/2} c(\mathcal{B}_l^{(\alpha_1+\beta_1)\vee\alpha_2}(X_{i\Delta_n}) + \mathcal{B}_r^{(1+\alpha_1+\beta_1)\vee(2+\alpha_2)}(X_{i\Delta_n})). \quad (18)$$

3. Limit theorems for functionals of the observed process

In this section and the following one, we study the behaviour of functionals (3)–(5) for $f : (l, r) \mapsto \mathbb{R}$ satisfying some regularity assumptions. The conditions on f are expressed by the following.

Condition C_γ: $f \in \mathcal{C}^2(l, r)$ and for $\gamma \geq 0$,

$$\exists c, \forall x \in (l, r), \quad |f(x)| + |f'(x)| + |f''(x)| \leq c(\mathcal{B}_l^\gamma(x) + \mathcal{B}_r^\gamma(x)).$$

For statistical purposes, we also need to consider the functionals (3)–(5) for $f(x, \theta) : (l, r) \times \Theta \mapsto \mathbb{R}$ where Θ is a product of two compact intervals of $\mathbb{R}$. To obtain uniform convergences with respect to θ, the conditions on $f = f(x, \theta)$ are the following.

Condition CU_γ: $f : (l, r) \times \Theta \to \mathbb{R}$ satisfies $f(\cdot, \cdot) \in \mathcal{C}^2[(l, r) \times O]$ for some open set $O \supset \Theta$ and with $\gamma \geq 0$,

$$\exists c > 0, \forall x \in (l, r) \quad \sup_{\theta \in \Theta} |g(x, \theta)| \leq c(\mathcal{B}_l^\gamma(x) + \mathcal{B}_r^\gamma(x)), \quad \text{for } g = f, f'_x, f''_{x^2}, \nabla_\theta f, \nabla_\theta f'_x.$$

3.1. Empirical mean

As a first application of the expansions recalled in section 2, we give a mean theorem for the process $(\overline{X}_i)_{i \in \mathbb{N}}$.

Proposition 2

Assume (A1)–(A5), let f *satisfy* CU_γ *with* $\gamma < M_l$, $1 + \gamma < M_r$ *and* $2\gamma < K_l$, *then*

$$\overline{\nu}_n(f(\cdot, \theta)) \overset{n \to \infty}{\longrightarrow} \nu_0(f(\cdot, \theta)) \quad \text{uniformly in } \theta, \text{ in probability.}$$

Proof. By lemma A1 of the Appendix, we only have to prove the $\mathbf{L}^1$ convergence to zero of

$$\sup_{\theta\in\Theta} n^{-1}\sum_{i=0}^{n-2}\left|f(\overline{X}_i,\theta)-f(X_{i\Delta_n},\theta)\right|.$$

By Taylor's expansion and condition CU_γ on f, we get the bound

$$\sup_{\theta\in\Theta}\left|f(\overline{X}_i,\theta)-f(X_{i\Delta_n},\theta)\right|\leq c\sup_{s\in[i\Delta_n,(i+1)\Delta_n]}(\mathcal{B}_l^\gamma(X_s)+\mathcal{B}_r^\gamma(X_s))\left|\overline{X}_i-X_{i\Delta_n}\right|.$$

Now, the Cauchy–Schwarz inequality, (A2), proposition 1(2) and (7) yield

$$E\left(\sup_{\theta\in\Theta}\left|f(\overline{X}_i,\theta)-f(X_{i\Delta_n},\theta)\right|\mid\mathcal{G}_i^n\right)\leq c\Delta_n^{1/2}(\mathcal{B}_l^\gamma(X_{i\Delta_n})+\mathcal{B}_r^{1+\gamma}(X_{i\Delta_n})).$$

By assumption (A5), we deduce $E(\sup_{\theta\in\Theta}\left|f(\overline{X}_i,\theta)-f(X_{i\Delta_n},\theta)\right|)\leq c\Delta_n^{1/2}$. So, the proposition is proved.

3.2. Variation of the process

Our next result concerns the functional $\overline{\mathcal{I}}_n$ which involves the increments of the process $(\overline{X}_i)$ (see (4)).

Theorem 2

Assume (A1)–(A5), and let f satisfy CU_γ, *with* $(\gamma+2\alpha_1+\beta_1+\beta_2)\vee(2\gamma+\alpha_2)<M_l$, $(4+2\gamma+2\alpha_1+\alpha_2)\vee(2+\gamma+2\alpha_2+\beta_1+\beta_2)<M_r$ *and* $2\gamma\vee4\alpha_1\vee2\alpha_2\vee\beta_1\vee2\beta_2<K_l$,

$$\overline{\mathcal{I}}_n(f(\cdot,\theta))\overset{n\to\infty}{\longrightarrow}\frac{1}{6}\nu_0\left(f_x'(\cdot,\theta)a^2(\cdot)\right)\ \textit{uniformly in } \theta\textit{, in probability.} \tag{19}$$

Proof. The proof relies on the expansions of theorem 1. Set for the proof

$$V_{i,n}(\theta)=f(\overline{X}_i,\theta)(\overline{X}_{i+1}-\overline{X}_i-\Delta_n b(\overline{X}_i)). \tag{20}$$

Since $V_{i,n}(\theta)$ is $\mathcal{G}_{i+2}^n$ measurable, to deal with a triangular array of martingale increments, we split $\overline{\mathcal{I}}_n(f(\cdot,\theta))$ into the sum of terms with even index i and the sum of terms with odd index i. Now, it is enough to show that

$$(n\Delta_{2n})^{-1}\sum_{i=0}^{n-1}V_{2i,2n}(\theta)\overset{n\to\infty}{\longrightarrow}\frac{1}{6}\nu_0\left(f_x'(\cdot,\theta)a^2(\cdot)\right)\qquad\text{uniformly in probability} \tag{21}$$

(and that $(n\Delta_{2n})^{-1}\sum_{i=0}^{n-2}V_{2i+1,2n}(\theta)\overset{n\to\infty}{\longrightarrow}1/6\nu_0\left(f_x'(\cdot,\theta)a^2(\cdot)\right)$, but the proof is analogous).

By the Taylor formula, and theorem 1(2) (recall (9)–(11)), $V_{i,n}(\theta)=\sum_{j=0}^{5}v_{i,n}^{(j)}(\theta)$, with,

$$v_{i,n}^{(1)}(\theta)=\Delta_n^{1/2}U_{i,n}a(X_{i\Delta_n})f(X_{i\Delta_n},\theta) \tag{22}$$

$$v_{i,n}^{(2)}(\theta)=\Delta_n^{1/2}U_{i,n}(\overline{X}_i-X_{i\Delta_n})a(X_{i\Delta_n})f_x'(X_{i\Delta_n},\theta) \tag{23}$$

$$v_{i,n}^{(3)}(\theta)=\Delta_n^{1/2}U_{i,n}\frac{1}{2}(\overline{X}_i-X_{i\Delta_n})^2a(X_{i\Delta_n})f_{x^2}''(\hat{X}_i,\theta) \tag{24}$$

$$v_{i,n}^{(4)}(\theta)=\varepsilon_{i,n}f(X_{i\Delta_n},\theta) \tag{25}$$

$$v_{i,n}^{(5)}(\theta)=\varepsilon_{i,n}(\overline{X}_i-X_{i\Delta_n})f_x'(\tilde{X}_i,\theta) \tag{26}$$

where $\hat{X}_i,\tilde{X}_i\in[\overline{X}_i,X_{i\Delta_n}]$.

Now, define $\overline{\mathcal{I}}_n^{(j)}(\theta)=(n\Delta_{2n})^{-1}\sum_{i=0}^{n-1} v_{2i,2n}^{(j)}(\theta)$ for $j=1,\dots,5$.

Let us first study $\overline{\mathcal{I}}_n^{(2)}(\theta)$ and prove that, for each θ,

$$\overline{\mathcal{I}}_n^{(2)}(\theta)\overset{n\to\infty}{\underset{\mathbf{P}}{\longrightarrow}}\frac{1}{6}\nu_0\left(f_x'(\cdot,\theta)a^2(\cdot)\right). \tag{27}$$

By lemma A2 of the appendix, it is enough to show that:

$$(n\Delta_{2n})^{-1}\sum_{i=0}^{n-1}E\left(v_{2i,2n}^{(2)}(\theta)\,|\,\mathcal{G}_{2i}^{2n}\right)\overset{n\to\infty}{\underset{\mathbf{P}}{\longrightarrow}}\frac{1}{6}\nu_0\left(f_x'(\cdot,\theta)a^2(\cdot)\right), \tag{28}$$

$$(n\Delta_{2n})^{-2}\sum_{i=0}^{n-1}E\left((v_{2i,2n}^{(2)}(\theta))^2\,|\,\mathcal{G}_{2i}^{2n}\right)\overset{n\to\infty}{\underset{\mathbf{L}^1}{\longrightarrow}}0. \tag{29}$$

Using theorem 1(1) and $E(U_{2i,2n}\xi'_{2i,2n}\,|\,\mathcal{G}_{2i}^{2n})=1/6$, we get

$$E\left(v_{2i,2n}^{(2)}(\theta)\,|\,\mathcal{G}_{2i}^{2n}\right)=\frac{\Delta_{2n}}{6}a^2(X_{2i\Delta_{2n}})f_x'(X_{2i\Delta_{2n}},\theta)+r_{i,n} \tag{30}$$

with $r_{i,n}=E\left(\Delta_{2n}^{1/2}U_{2i,2n}f_x'(X_{2i\Delta_{2n}},\theta)a(X_{2i\Delta_{2n}})e_{i,n}\,|\,\mathcal{G}_{2i}^{2n}\right)$.

Now, (13), assumption (A1) and condition CU_γ yield

$$|r_{i,n}|\le\Delta_{2n}^{3/2}c(\mathcal{B}_l^{\gamma+\alpha_1}(X_{2i\Delta_{2n}})+\mathcal{B}_l^{2+\gamma+\alpha_1}(X_{2i\Delta_{2n}})). \tag{31}$$

Therefore, by (A5), $(n\Delta_{2n})^{-1}\sum_{i=0}^{n-1}r_{i,n}\overset{n\to\infty}{\underset{\mathbf{L}^1}{\longrightarrow}}0$.

Hence, an application of lemma A1 in the appendix to the first term of (30) yields (28).

Now, by proposition 1(2), $E\left((v_{2i,2n}^{(2)}(\theta))^2\,|\,\mathcal{G}_{2i}^{2n}\right)\le\Delta_{2n}^2c(\mathcal{B}_l^{2\gamma}(X_{2i\Delta_{2n}})+\mathcal{B}_r^{4+2\gamma}(X_{2i\Delta_{2n}}))$. This gives (29), and (27) follows.

To obtain uniformity with respect to θ we shall use proposition A1 of the appendix. We compute $\nabla_\theta v_{2i,2n}^{(2)}(\theta)=\Delta_{2n}^{1/2}U_{2i,2n}(\overline{X}_{2i}-X_{2i\Delta_{2n}})a(X_{2i\Delta_n})\nabla_\theta f_x'(X_{2i\Delta_{2n}},\theta)$. By condition CU_γ, we deduce $E(\sup_{\theta\in\Theta}|\nabla_\theta v_{2i,2n}^{(2)}(\theta)|\,|\,\mathcal{G}_{2i}^{2n})\le c\Delta_n(\mathcal{B}_l^{\gamma}(X_{2i\Delta_{2n}})+\mathcal{B}_r^{2+\gamma}(X_{2i\Delta_{2n}}))$. With assumption (A5), it implies $E(\sup_{\theta\in\Theta}|\nabla_\theta v_{2i,2n}^{(2)}(\theta)|)\le c\Delta_n$. Hence, $\sup_{n\in\mathbb{N}}E(\sup_{\theta\in\Theta}|\nabla_\theta\overline{\mathcal{I}}_n^{(2)}(\theta)|)<\infty$ and uniformity in (27) follows.

To end the proof of the theorem, it remains to show the uniform convergence to 0 for $\overline{\mathcal{I}}_n^{(1)}(\theta)$, $\overline{\mathcal{I}}_n^{(3)}(\theta)$, $\overline{\mathcal{I}}_n^{(4)}(\theta)$ and $\overline{\mathcal{I}}_n^{(5)}(\theta)$. This leads to rather long computations which are detailed in the appendix.

Remark 1. To enlighten the previous result, recall that

$$(n\Delta_n)^{-1}\sum_{i=0}^{n-1}f(X_{i\Delta_n})\;(X_{(i+1)\Delta_n}-X_{i\Delta_n}-b(X_{i\Delta_n})\Delta_n)\overset{n\to\infty}{\longrightarrow}0.$$

The difference with our result comes from the fact that $f(X_{i\Delta_n})$ and $X_{(i+1)\Delta_n}-X_{i\Delta_n}-b(X_{i\Delta_n})\Delta_n$ have a negligible correlation (of order Δ_n) whereas $f(\overline{X}_i)$ and $\overline{X}_{i+1}-\overline{X}_i-b(\overline{X}_i)\Delta_n$ have a correlation of order $\Delta_n^{1/2}$.

3.3. Quadratic variation of the observed process

The following result deals with the quadratic variation of $\overline{X}_i$ (recall (5)).

Theorem 3

Assume (A1)–(A5), let f satisfy CU_γ *with* $2\gamma \vee (\gamma + 2\alpha_1 + \alpha_2) < M_l$, $4 + 2\gamma + 2\alpha_1 + \alpha_2 < M_r$ *and* $2\gamma \vee 4\alpha_1 \vee 2\alpha_2 < K_l$, *then,*

$$\overline{\mathcal{Q}}_n(f(\cdot,\theta)) \overset{n\to\infty}{\longrightarrow} \frac{2}{3} v_0\left(f(\cdot,\theta)a^2(\cdot)\right) \text{ uniformly in } \theta\text{, in probability.} \tag{32}$$

Proof. First, we show the pointwise in θ convergence. Set $W_{i,n}(\theta) = \left(\overline{X}_{i+1} - \overline{X}_i\right)^2 f(\overline{X}_i, \theta)$. As $W_{i,n}(\theta)$ is $\mathcal{G}^n_{i+2}$ measurable, to prove the convergence of $(n\Delta_{2n})^{-1}\sum_{i=0}^n W_{i,n}(\theta)$, as in the previous theorem, we deal separately with the sum of even indexes and the one of odd indexes. And it is enough to show that:

$$(n\Delta_{2n})^{-1}\sum_{i=0}^{n-1} W_{2i,2n}(\theta) \overset{n\to\infty}{\longrightarrow} \frac{2}{3} v_0\left(f(\cdot,\theta)a^2(\cdot)\right) \text{ in probability.} \tag{33}$$

Using theorem 1(2) and Taylor's formula, we write: $W_{2i,2n}(\theta) = w^{(1)}_{2i,2n}(\theta) + w^{(2)}_{2i,2n}(\theta) + w^{(3)}_{2i,2n}(\theta) + w^{(4)}_{2i,2n}(\theta)$

$$\text{with} \quad w^{(1)}_{2i,2n}(\theta) = \Delta_{2n} U^2_{2i,2n} a^2(X_{2i\Delta_{2n}}) f(X_{2i\Delta_{2n}}, \theta) \tag{34}$$

$$w^{(2)}_{2i,2n}(\theta) = 2\Delta_{2n}^{1/2} U_{2i,2n} a(X_{2i\Delta_{2n}})(\varepsilon_{2i,2n} + \Delta_{2n} b(X_{2i\Delta_{2n}})) f(X_{2i\Delta_{2n}}, \theta) \tag{35}$$

$$w^{(3)}_{2i,2n}(\theta) = (\varepsilon_{2i,2n} + \Delta_{2n} b(X_{2i\Delta_{2n}}))^2 f(X_{2i\Delta_{2n}}, \theta) \tag{36}$$

$$w^{(4)}_{2i,2n}(\theta) = (\overline{X}_{2i+1} - \overline{X}_{2i})^2 (\overline{X}_{2i} - X_{2i\Delta_{2n}}) f'_x(\hat{X}_i, \theta), \quad \text{where } \hat{X}_i \in [\overline{X}_{2i}, X_{2i\Delta_{2n}}]. \tag{37}$$

We set $\overline{\mathcal{Q}}^{(j)}_n(\theta) = (n\Delta_{2n})^{-1}\sum_{i=0}^{n-1} w^{(j)}_{2i,2n}(\theta)$ for $j = 1, 2, 3, 4$.

We start by studying $\overline{\mathcal{Q}}^{(1)}_n(\theta)$. Using $E\left(U^2_{2i,2n} \mid \mathcal{G}^{2n}_{2i}\right) = 2/3$ and $E\left(U^4_{2i,2n} \mid \mathcal{G}^{2n}_{2i}\right) = 4/3$ we obtain:

$$E\left(w^{(1)}_{2i,2n}(\theta) \mid \mathcal{G}^{2n}_{2i}\right) = \frac{2\Delta_{2n}}{3} f(X_{i\Delta_{2n}}, \theta) a^2(X_{i\Delta_{2n}})$$

$$E\left((w^{(1)}_{2i,2n}(\theta))^2 \mid \mathcal{G}^{2n}_{2i}\right) = \frac{4\Delta^2_{2n}}{3} f^2(X_{i\Delta_{2n}}, \theta) a^4(X_{i\Delta_{2n}}).$$

First, applying lemma A1 we get:

$$(n\Delta_{2n})^{-1}\sum_{i=0}^{n-1} E\left(w^{(1)}_{2i,2n}(\theta) \mid \mathcal{G}^{2n}_{2i}\right) \overset{n\to\infty}{\longrightarrow} \frac{2}{3} v_0\left(f(\cdot,\theta)a^2(\cdot)\right) \text{ in probability.}$$

Second, using (A5) we get

$$E\left|E\left((w^{(1)}_{2i,2n}(\theta))^2 \mid \mathcal{G}^{2n}_{2i}\right)\right| \leq c\Delta^2_{2n},$$

and therefore, $(n\Delta_{2n})^{-2}\sum_{i=0}^{n-1} E\left((w^{(1)}_{2i,2n}(\theta))^2 \mid \mathcal{G}^{2n}_{2i}\right) \underset{\mathbf{L}^1}{\overset{n\to\infty}{\longrightarrow}} 0.$

Hence, by lemma A2, we deduce $\overline{\mathcal{Q}}^{(1)}_n(\theta) \overset{n\to\infty}{\longrightarrow} 2/3 v_0\left(f(\cdot,\theta)a^2(\cdot)\right)$ in probability.

To establish the pointwise in θ convergence for (32), it remains to show that $\overline{\mathcal{Q}}^{(i)}_n(\theta) \overset{n\to\infty}{\longrightarrow} 0$ in probability for $i = 2, 3, 4$. This is detailed in the appendix. To obtain the uniformity, we use proposition A1 of the appendix together with the easily obtained following bound:

$$\sup_{n\geq 1}(n\Delta_n)^{-1}\sum_{i=0}^{n-1} E\left[\left(\overline{X}_{i+1} - \overline{X}_i\right)^2 \sup_{\theta\in\Theta}\left|\nabla_\theta f(\overline{X}_i, \theta)\right|\right] < \infty.$$

Remark 2. Comparing theorem 3 with the well known result

$$(n\Delta_n)^{-1}\sum_{i=0}^{n-2} f(X_{i\Delta_n})(X_{(i+1)\Delta_n}-X_{i\Delta_n})^2 \overset{n\to\infty}{\longrightarrow} \nu_0\left(fa^2\right),$$

the main difference comes from the fact that the variance of $U_{i,n}$ is 2/3, whereas $\Delta_n^{-1/2}(B_{(i+1)\Delta_n}-B_{i\Delta_n})$ has variance 1.

The quadratic variation $\overline{\mathcal{Q}}_n(f)$ based on $(\overline{X}_i)$ is therefore a biased estimator of the quadratic variation $\nu_0(fa^2)$. An analogous result was obtained in Delattre & Jacod (1997) for the quadratic variation based on discrete observations of the diffusion with round-off errors.

4. Associated central limit theorems

Now, we study some related central limit theorems. We need no more uniformity in θ.

Theorem 4

Assume (A1)–(A5) and $n\Delta_n^2 \overset{n\to\infty}{\longrightarrow} 0$. *Let* f *satisfies* C_γ *with* $(2\gamma+3\alpha_1+\alpha_2+\beta_1+\beta_2)\vee 4\gamma < M_l$, $(4+2\gamma+3\alpha_1+\alpha_2+\beta_1+\beta_2)\vee(4+4\gamma) < M_r$ *and* $4\gamma\vee 4\alpha_1\vee 2\alpha_2\vee\beta_1\vee 2\beta_2 < K_l$ *then,*

$$\overline{N}_n(f) := \sqrt{n\Delta_n}(\overline{\mathcal{I}}_n(f) - \frac{1}{4}\overline{\mathcal{Q}}_n(f')) \underset{\mathcal{D}}{\overset{n\to\infty}{\longrightarrow}} \mathcal{N}\left(0, \nu_0\left(f^2a^2\right)\right). \tag{38}$$

Proof. In this proof, we use the notations (20), (22)–(26) introduced in theorem 2. Furthermore, here, we set ${v'_{i,n}}^{(2)} = -1/4(\overline{X}_{i+1}-\overline{X}_i)^2 f'(\overline{X}_i)$.

Define $\overline{N}_n^{(j)} = (n\Delta_n)^{-1/2}\sum_{i=0}^{n-1} v_{i,n}^{(j)}$ for $j=1,3,4,5$ and $\overline{N}_n^{(2)} = (n\Delta_n)^{-1/2}\sum_{i=0}^{n-1}(v_{i,n}^{(2)} + {v'_{i,n}}^{(2)})$. With these notations $\overline{N}_n(f) = \sum_{l=1}^{5}\overline{N}_n^{(l)}$.

First, we study $\overline{N}_n^{(1)} = n^{-1/2}\sum_{i=0}^{n-1}\alpha(X_{i\Delta_n})U_{i,n}$ with $\alpha(x) = f(x)a(x)$. In order to apply a martingale central limit theorem, we first have to reorder terms (recall (9)–(11)).

$$\overline{N}_n^{(1)} = n^{-1/2}\sum_{i=1}^{n-1} s_{i,n}^{(1)} + n^{-1/2}\left(\alpha(X_0)\xi_{0,n} + \alpha(X_{(n-1)\Delta_n})\xi'_{n-1,n}\right), \tag{39}$$

with

$$s_{i,n}^{(1)} = \alpha(X_{i\Delta_n})\xi_{i,n} + \alpha(X_{(i-1)\Delta_n})\xi'_{i,n}. \tag{40}$$

We now have the conditional centring, $E\left(s_{i,n}^{(1)} \mid \mathcal{G}_i^n\right) = 0$, and compute the conditional variance $E\left((s_{i,n}^{(1)})^2 \mid \mathcal{G}_i^n\right) = 1/3\{\alpha^2(X_{i\Delta_n}) + \alpha^2(X_{(i-1)\Delta_n}) + \alpha(X_{i\Delta_n})\alpha(X_{(i-1)\Delta_n})\}$. An application of lemma A1, and proposition 1(1) yields: $n^{-1}\sum_{i=0}^{n-1} E\left((s_{i,n}^{(1)})^2 \mid \mathcal{G}_i^n\right) \underset{\mathbf{P}}{\overset{n\to\infty}{\longrightarrow}} \nu_0\left(f^2a^2\right)$. We easily bound $E\left((s_{i,n}^{(1)})^4 \mid \mathcal{G}_i^n\right)$ and show $n^{-2}\sum_{i=0}^{n-1} E\left((s_{i,n}^{(1)})^4 \mid \mathcal{G}_i^n\right) \underset{\mathbf{L}^1}{\overset{n\to\infty}{\longrightarrow}} 0$. Using theorem 3.2 (p. 58) in Hall & Heyde (1980), these two conditions are sufficient to imply

$$n^{-1/2}\sum_{i=0}^{n-1} s_{i,n}^{(1)} \underset{\mathcal{D}}{\overset{n\to\infty}{\longrightarrow}} \mathcal{N}\left(0, \nu_0\left(f^2a^2\right)\right). \tag{41}$$

Clearly, by (39), we deduce $\overline{N}_n^{(1)} \underset{\mathcal{D}}{\overset{n\to\infty}{\longrightarrow}} \mathcal{N}\left(0, \nu_0\left(f^2a^2\right)\right)$.

Second, we show the convergence to 0 of $\overline{N}_n^{(2)}$. Computations based on theorem 1(1) and proposition 1 show that $E\left({v'_{2i,2n}}^{(2)} \mid \mathcal{G}_i^n\right) = -(\Delta_{2n}/6)a^2(X_{2i\Delta_{2n}})f'(X_{2i\Delta_{2n}}) + r'_{i,n}$ with,

$$\left|r'_{i,n}\right| \le c\Delta_{2n}^{3/2}(\mathcal{B}_l^{\gamma+3\alpha_1\vee\gamma+2\alpha_1+\alpha_2}(X_{2i\Delta_{2n}}) + \mathcal{B}_r^{4+\gamma+3\alpha_1+\alpha_2}(X_{2i\Delta_{2n}})). \tag{42}$$

Recalling (30), we get $E\left(v^{(2)}_{2i,2n}+{v'}^{(2)}_{2i,2n}\,|\,\mathcal{G}^{2n}_{2i}\right)=r_{i,n}+r'_{i,n}$. Now, by (31), (42) and (A5) we deduce,

$$E\left|E\left(v^{(2)}_{2i,2n}+{v'}^{(2)}_{2i,2n}\,|\,\mathcal{G}^{2n}_{2i}\right)\right|\le c\Delta^{3/2}_{2n}.$$

This implies $E\left|(n\Delta_{2n})^{-1/2}\sum_{i=0}^{n-1}E\left(v^{(2)}_{2i,2n}+{v'}^{(2)}_{2i,2n}\,|\,\mathcal{G}^{2n}_{2i}\right)\right|\le c(n\Delta^2_{2n})^{1/2}$, which tends to 0 using now the condition $n\Delta^2_n\to 0$. As for (29), we get $(n\Delta_{2n})^{-1}\sum_{i=0}^{n-1}E\left((v^{(2)}_{2i,2n}+{v'}^{(2)}_{2i,2n})^2\,|\,\mathcal{G}^{2n}_{2i}\right)=o_{\mathbf{P}}(1)$. This implies the convergence for the sum of terms with even indexes:

$$(n\Delta_{2n})^{-1/2}\sum_{i=0}^{n-1}\{v^{(2)}_{2i,2n}+{v'}^{(2)}_{2i,2n}\}=o_{\mathbf{P}}(1).$$

Analogously we show $(n\Delta_{2n+1})^{-1/2}\sum_{i=0}^{n-1}\{v^{(2)}_{2i+1,2n}+{v'}^{(2)}_{2i+1,2n}\}=o_{\mathbf{P}}(1)$. Hence $\overline{N}^{(2)}_n\to 0$.

To end the proof, we show the convergence to zero for $\overline{N}^{(3)}_n$, $\overline{N}^{(4)}_n$ and $\overline{N}^{(5)}_n$ using that $n\Delta^2_n\to 0$ (the proof is a repetition of the proof of convergence for the corresponding terms, $\overline{\mathcal{I}}^{(3)}_n$, $\overline{\mathcal{I}}^{(4)}_n$ and $\overline{\mathcal{I}}^{(5)}_n$, in theorem 2).

Remark 3.

1. Comparing with the classic convergence limit,

$$(n\Delta_n)^{-1/2}\sum_{i=0}^{n-1}(X_{(i+1)\Delta_n}-X_{i\Delta_n}-\Delta_n b(X_{i\Delta_n}))f(X_{i\Delta_n})\overset{n\to\infty}{\longrightarrow}\mathcal{N}\left(0,v_0\left(f^2a^2\right)\right),$$

   it appears that, if we just replace $X_{i\Delta_n}$ by $\overline{X}_i$ above, then the expression may tend to ∞, in probability, because of the non-negligible correlation between $f(\overline{X}_i)$ and $\overline{X}_{i+1}-\overline{X}_i-b(\overline{X}_i)\Delta_n$. So we have to introduce the appropriate correction.
2. We cannot replace in the statement of the previous theorem $1/4\overline{\mathcal{Q}}_n(f')$ by the term $1/6v_0\left(f'a^2\right)$ since we can show that

$$(n\Delta_n)^{-1/2}\left\{\sum_{i=0}^{n-2}\frac{1}{4}f'(\overline{X}_i)\left(\overline{X}_{i+1}-\overline{X}_i\right)^2-\frac{1}{6}v_0\left(f'a^2\right)\right\}$$

   does not tend to zero when $n\to\infty$, hence theorem 4 does not provide an exact central limit theorem for theorem 2.
3. The condition $n\Delta^2_n\to 0$ is classical (see Florens-Zmirou, 1989). This condition imposes that the discretization step decreases to zero fast enough, to ensure that the contribution in $\overline{N}_n(f)$ of the error terms $\varepsilon_{i,n}$ tends to 0 as $n\to\infty$.

Let us now state a central limit theorem related with the functional $\overline{\mathcal{Q}}_n(g)$ (see (5)).

Theorem 5

Assume (A1)–(A5) and $n\Delta^2_n\overset{n\to\infty}{\longrightarrow}0$. *Let g satisfies* C_γ *with* $4\gamma\vee(2\gamma+4\alpha_1+2\alpha_2+\beta_1)<M_l$, $(8+4\gamma)\vee(6+2\gamma+4\alpha_1+2\beta_2+\beta_1)<M_r$ *and* $4\gamma\vee(2\gamma+4\alpha_1+2\alpha_2)\vee 4\beta_2<K_l$. *Then,*

$$\overline{M}_n(g):=n^{1/2}\left(\frac{3}{2}\overline{\mathcal{Q}}_n(g)-\overline{v}_n(a^2g)\right)\underset{\mathcal{D}}{\overset{n\to\infty}{\longrightarrow}}\mathcal{N}\left(0,\frac{9}{4}v_0(g^2a^4)\right).\qquad(43)$$

Proof. In this proof we use notations (34)–(37) of theorem 3. Further, we set $\beta(x)=a^2(x)g(x)$, ${w'}^{(1)}_{i,n}=-2/3\beta(X_{i\Delta_n})$ and ${w'}^{(5)}_{i,n}=2/3(\beta(X_{i\Delta_n})-\beta(\overline{X}_i))$.

Define $\overline{M}^{(l)}_n=3/2n^{-1/2}\Delta_n^{-1}\sum_{i=0}^{n-2}w^{(l)}_{i,n}$, for $l=2,3,4,5$ and $\overline{M}^{(1)}_n=3/2n^{-1/2}\Delta_n^{-1}\sum_{i=0}^{n-2}(w^{(1)}_{i,n}+{w'}^{(1)}_{i,n})$. With these notations, $\overline{M}_n(g)=\sum_{l=1}^{5}\overline{M}^{(l)}_n$.

First, we study $\overline{M}_n^{(1)} = n^{-1/2}\sum_{i=0}^{n-1}\beta(X_{i\Delta_n})\{3/2(U_{i,n})^2 - 1\}$. Reordering terms in $\overline{M}_n^{(1)}$ to obtain a triangular array of martingale increments, we get (recall (11)),

$$\overline{M}_n^{(1)} = \frac{3}{2n^{1/2}}\left\{\sum_{i=1}^{n-1} s_{i,n}^{(2)} + \left(\xi_{0,n}^2 - \frac{1}{3}\right)\beta(X_0) + \left(\xi_{n,n}'^2 - \frac{1}{3}\right)\beta(X_{(n-1)\Delta_n}) + 2\xi_{n-1,n}\xi'_{n,n}\beta(X_{(n-1)\Delta_n})\right\}$$

where,

$$s_{i,n}^{(2)} = \left(\xi_{i,n}^2 - \frac{1}{3}\right)\beta(X_{i\Delta_n}) + \left(\xi_{i,n}'^2 - \frac{1}{3}\right)\beta(X_{(i-1)\Delta_n}) + 2\xi_{i-1,n}\xi'_{i,n}\beta(X_{(i-1)\Delta_n}). \tag{44}$$

But, now, $s_{i,n}^{(2)}$ is $\mathcal{G}_{i+1}^n$ measurable and centred conditionally to $\mathcal{G}_i^n$. Furthermore, using the expression for the covariance structure of $(\xi_{i,n}, \xi'_{i,n})$ given in section 2, we deduce:

$$E\left((s_{i,n}^{(2)})^2 \mid \mathcal{G}_i^n\right) = \frac{2}{9}\beta^2(X_{i\Delta_n}) + \frac{2}{9}\beta^2(X_{(i-1)\Delta_n}) + \frac{1}{9}\beta(X_{(i-1)\Delta_n})\beta(X_{i\Delta_n}) + \frac{4}{3}\xi_{i-1,n}^2\beta^2(X_{(i-1)\Delta_n}). \tag{45}$$

By lemmas A1 and A2 (in the appendix) we show the convergence for the following array of martingale increments, $n^{-1}\sum_{i=0}^{n-1}\xi_{i-1,n}^2\beta^2(X_{(i-2)\Delta_n}) \xrightarrow[\mathbf{P}]{n\to\infty} 1/3\nu_0(\beta^2)$ (recall $\beta(x) = a^2(x)g(x)$). Using proposition 1(1) we deduce $n^{-1}\sum_{i=0}^{n-1}\xi_{i-1,n}^2\beta^2(X_{(i-1)\Delta_n}) \xrightarrow[\mathbf{P}]{n\to\infty} 1/3\nu_0(\beta^2)$. Thus we have, by (45), $9/4n^{-1}\sum_{i=1}^{n-1}E\left((s_{i,n}^{(2)})^2 \mid \mathcal{G}_i^n\right) \xrightarrow[\mathbf{P}]{n\to\infty} 9/4\nu_0(\beta^2)$. Using the bound on β^4, we see that we can apply theorem 3.2 in Hall & Heyde (1980) and get, $3/2n^{-1/2}\sum_{i=0}^{n-1}s_{i,n}^{(2)} \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, 9/4\nu_0(\beta^2)\right)$.

Finally, using that $n\Delta_n^2 \to 0$, we can show $\overline{M}_n^{(l)} \xrightarrow{n\to\infty} 0$ in probability for $l = 2, \ldots, 5$ and, as a consequence, obtain (43).

Remark 4. Comparing with the usual property

$$n^{-1/2}\sum_{i=0}^{n-1}\left\{g(X_{i\Delta_n})\frac{\left(X_{(i+1)\Delta_n} - X_{i\Delta_n}\right)^2}{\Delta_n} - g(X_{i\Delta_n})a^2(X_{i\Delta_n})\right\} \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, 2\nu_0(g^2a^4)\right)$$

we see that when we replace $X_{i\Delta_n}$ by $\overline{X}_i$ in the equation above the variance of the limit increases a little.

Finally, we have the following theorem.

Theorem 6

Let f and g be two functions satisfying, respectively, C_γ and $C_{\gamma'}$. Assume that the assumptions of theorem 4 are valid for f, and those of theorem 5 are valid for g. Suppose, furthermore, that $2\gamma + 2\gamma' + \alpha_1 < M_l$, $6 + 2\gamma + 2\gamma' + \alpha_1 < M_r$. Then,

$$(\overline{N}_n(f), \overline{M}_n(g)) \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, \nu_0(f^2a^2)\right) \otimes \mathcal{N}\left(0, \frac{9}{4}\nu_0(g^2a^4)\right). \tag{46}$$

Proof. We have shown, in theorems 4 and 5, that $\overline{N}_n(f) - n^{-1/2}\sum_{i=1}^{n-1}s_{i,n}^{(1)}$ and $\overline{M}_n(g) - 3/2n^{-1/2}\sum_{i=1}^{n-1}s_{i,n}^{(2)}$ tend to zero in probability. As we deal with martingale arrays, it suffices to prove

$$n^{-1}\sum_{i=1}^{n-1} E\left(s_{i,n}^{(1)}s_{i,n}^{(2)}\mid \mathcal{G}_i^n\right) \underset{\mathbf{P}}{\xrightarrow{n\to\infty}} 0. \tag{47}$$

But using (40) and (44), we get (recall $\alpha(x)=a(x)g(x)$, $\beta(x)=a^2(x)g(x)$):

$$E\left(s_{i,n}^{(1)}s_{i,n}^{(2)}\mid \mathcal{G}_i^n\right) = \xi_{i-1,\,n}\beta(X_{(i-1)\Delta_n})\left(\frac{1}{3}\alpha(X_{i\Delta_n}) + \frac{2}{3}\alpha(X_{(i-1)\Delta_n})\right).$$

Then (47) can be obtained by using proposition 1 and lemma A2.

5. Checking assumptions

In this section we give examples of models for which our assumptions hold. Actually, only (A2) and (A5) have to be studied. The examples below show that they are not restrictive conditions.

The condition (A2) was studied in Gloter (2000) and was shown to hold for all models below.

5.1. Diffusion on $\mathbb{R}$

Here $(l,r)=(-\infty,\infty)$; $\mathcal{B}_{-\infty}(x)=1$ and $\mathcal{B}_{\infty}(x)=1+|x|$; (X_t) is the solution of

$$\mathrm{d}X_t = a(X_t)\,\mathrm{d}B_t + b(X_t)\,\mathrm{d}t, \quad X_0=\eta. \tag{48}$$

Let us assume that assumptions (A1) and (A3) are satisfied, and the stationary distribution v_0 has finite moments of every order. This means that (A4) is satisfied for any positive constants $M_{-\infty}$ and M_{∞}. By Gloter (2000), assumption (A2) is satisfied with $K_{-\infty}=\infty$.

It remains to check (A5). If η has distribution $\mathrm{d}v_0$, then (A5) follows from (A4) by stationarity. The next proposition shows that (A5) follows again from (A4) if η is deterministic.

Proposition 3

Let X be solution of (48) starting from $X_0=y\in\mathbb{R}$. Assume that (A1), (A3) and (A4) hold with any positive constants $M_{-\infty}$ and M_{∞}, then

$$\forall p\geq 0,\ \sup_{t\geq 0} E(|X_t|^p)<\infty \text{ (and hence (A5) holds).}$$

Proof. Define the probability measure $\mathrm{d}\pi(x)=\mathrm{d}v_0(x)v_0\left([y,\infty)\right)^{-1}\mathbb{1}_{\{x\geq y\}}$. Using that the stochastic differential equation admits strong solutions we can define on $(\Omega,\mathcal{F},P)$ the processes $(X_t^x), x\in\mathbb{R}$ and (X_t^π), solutions of:

$$\mathrm{d}X_t^x = a(X_t^x)\,\mathrm{d}B_t + b(X_t^x)\,\mathrm{d}t, \quad X_0^x = x$$
$$\mathrm{d}X_t^\pi = a(X_t^\pi)\,\mathrm{d}B_t + b(X_t^\pi)\,\mathrm{d}t, \quad X_0^\pi = \Pi$$

with Π a random variable with distribution $\mathrm{d}\pi(x)$, independent of B.

Now, since $X_0=X_0^y \leq X_0^\pi$ a.s., using that $X=X^y$ and X^π coincide after the time $T=\inf\{s\geq 0;\ X_s^y=X_s^\pi\}$ we get $P(\forall t\geq 0, X_s\leq X_s^\pi)=1$. So, for $p\in\mathbb{N}$ (with the notation $x^+=x\vee 0$):

$$E\left((X_t{}^+)^p\right)\leq E\left((X_t^{\pi+})^p\right) = \int_{\mathbb{R}} E\left((X_t^{x+})^p\right)\,\mathrm{d}\pi(x)$$
$$\leq v_0\left([y,\infty)\right)^{-1}\int_{\mathbb{R}} E\left((X_t^{x+})^p\right)\mathrm{d}v_0(x) = v_0\left([y,\infty)\right)^{-1}\int_{\mathbb{R}} (x^+)^p\,\mathrm{d}v_0(x).$$

Since by assumption, v_0 has finite moments of every orders, we have $\sup_{t\geq 0} E\big(\big(X_t^+\big)^p\big) < \infty$. We analogously show: $\sup_{t\geq 0} E\big(\big(X_t^-\big)^p\big) < \infty$ with $x^- = (-x) \vee 0$ and get the proposition.

As a result, the work of sections 3 and 4 encompasses a large class of diffusion models on $\mathbb{R}$, among them we can quote the Ornstein–Uhlenbeck process (for $b(x) = \mu x$ and $a(x) = \sigma$) or the hyperbolic process (for $b(x) = \mu x(1 + x^2)^{-1/2}$ and $a(x) = \sigma$).

5.2. The Cox–Ingersoll–Ross process

Here $(l, r) = (0, \infty)$; $\mathcal{B}_0(x) = 1 + 1/x$, $\mathcal{B}_\infty(x) = 1 + x$ and

$$\mathrm{d}X_t = (\mu X_t + \mu')\,\mathrm{d}t + \sigma\sqrt{X_t}\,\mathrm{d}B_t, \quad X_0 = \eta, \tag{49}$$

with $\mu < 0$, $\sigma, \mu' > 0$. We set $c_0 = 2\mu'/\sigma^2$, $\lambda = 2\,|\mu|\,/\sigma^2$ and suppose that $c_0 > 1$.

Assumption (A1) is clear and $c_0 > 1$ implies assumption (A3), with the stationary probability measure, $\mathrm{d}v_0(x) = \lambda^{c_0}\Gamma(c_0)^{-1}x^{c_0-1}\mathrm{e}^{-\lambda x}\,\mathbb{1}_{\{x>0\}}\,\mathrm{d}x$.

Using this expression for the stationary probability, we easily check (A4) with any $M_0 < c_0$. It is shown in Gloter (2000) that (A2) holds with $K_0 = c_0 - 1$ (actually this model justifies the introduction of the constant K_l in (A2)).

If η has distribution v_0, (A5) holds. If η is deterministic, we show (A5) as in proposition 3.

6. Statistical applications. Minimum contrast estimation

Let (X_t) be the unique solution of the equation:

$$\mathrm{d}X_t = a(X_t, \sigma_0)\,\mathrm{d}B_t + b(X_t, \mu_0)\,\mathrm{d}t, \quad X_0 = \eta, \tag{50}$$

where $(B_t)_{t\geq 0}$ is a standard one-dimensional Brownian motion, η is a random variable independent of $(B_t)_{t\geq 0}$; b and a are two real-valued functions, respectively, defined on $\mathbb{R} \times \Theta_1$ and $\mathbb{R} \times \Theta_2$, where Θ_1 and Θ_2 are two compact intervals of $\mathbb{R}$. We denote by $\theta = (\mu, \sigma)$ the elements of $\Theta = \Theta_1 \times \Theta_2$, furthermore we suppose that $\theta_0 = (\mu_0, \sigma_0) \in \overset{\circ}{\Theta}$.

In the case of the observation of $X_{i\Delta_n}$ a contrast may be constructed by approximating $X_{(i+1)\Delta_n} - X_{i\Delta_n}$ by an $\mathcal{N}(b(X_{i\Delta_n}, \mu_0)\Delta_n, a^2(X_{i\Delta_n}, \sigma_0)\Delta_n)$. This leads to the Euler contrast (see Kessler, 1997).

To obtain a contrast based on the observation of $\overline{X}_i$, we correct the Euler contrast by taking into account the factor 2/3 in theorem 3, and compensate the effect of the correlation between $\overline{X}_i$ and $\overline{X}_{i+1} - \overline{X}_i - \Delta_n b(\overline{X}_i, \mu_0)$ (theorem 2).

This leads to the following contrast

$$\mathcal{L}_n(\theta) = \sum_{i=0}^{n-2}\left\{\frac{3}{2\Delta_n}\left(\frac{\left(\overline{X}_{i+1} - \overline{X}_i - b(\overline{X}_i, \mu)\Delta_n\right)^2}{a^2(\overline{X}_i, \sigma)} + \frac{\Delta_n}{2}h(\overline{X}_i, \theta)\left(\overline{X}_{i+1} - \overline{X}_i\right)^2\right) + \log a^2(\overline{X}_i, \sigma)\right\} \tag{51}$$

with $h(x, \theta) = (\partial/\partial x)(b(x, \mu)/a^2(x, \sigma))$. Let $\hat{\theta}_n = \operatorname{arginf}_{\theta\in\Theta}\mathcal{L}_n(\theta)$ be a minimum contrast estimator.

Observe that $\mathcal{L}_n$ is not obtained as a pseudo-likelihood function suggested by the expansions of theorem 1. Indeed, it seems difficult to deduce a pseudo-likelihood function from these expansions, in particular because $U_{i,n}$ in theorem 1(2) is not Markovian.

We suppose that the diffusion X satisfies assumptions (A1)–(A5) (where $a(x)$ stands for $a(x, \sigma_0)$ and $b(x)$ for $b(x, \mu_0)$). Furthermore, to keep proofs on the behaviour of $\hat{\theta}_n$ tractable, we make the additional assumption that K_l in assumption (A2) is equal to ∞ and that assumptions (A4)–(A5) hold for any M_l and M_r (and hence these constants can be chosen as large as we need for the application of results of sections 3 and 4). We have seen in section 5.1 that this is true for a large class of diffusion processes. For models that do not satisfy this additional assumption, we will give a specific proof of the consistency and normality of $\hat{\theta}_n$ (see section 7.2).

We suppose that the following identifiability assumption holds:

(S1) $a(x,\sigma)=a(x,\sigma_0)$ $\mathrm{d}\nu_0(x)$ almost everywhere implies $\sigma=\sigma_0$,
$b(x,\mu)=b(x,\mu_0)$ $\mathrm{d}\nu_0(x)$ almost everywhere implies $\mu=\mu_0$.

We need an assumption on the smoothness of $a(x,\sigma)$ and $b(x,\mu)$ with respect to the parameter.

(S2) a and b are the restrictions of functions defined on an open subset of $\mathbb{R}^2$, on which they are differentiable up to order 6, furthermore they satisfy: $\exists c\geq 0, \exists k\geq 0$ such that $\forall i,\ j\in\{0,\dots,3\}^2, \forall x\in(l,r)$:

$$\sup_{\mu\in\Theta_1}\left|\frac{\partial^{i+j}}{\partial\mu^i x^j}b(x,\mu)\right| + \sup_{\sigma\in\Theta_2}\left|\frac{\partial^{i+j}}{\partial\sigma^i x^j}a(x,\sigma)\right| + \sup_{\sigma\in\Theta_2}\left|\frac{\partial^{i+j}}{\partial\sigma^i x^j}a^{-1}(x,\sigma)\right| \leq c(\mathcal{B}_l^k(x)+\mathcal{B}_r^k(x)).$$

By the previous assumption, all functions appearing below satisfy CU_γ for some $\gamma\geq 0$ and hence limit theorems of sections 3 and 4 apply for these functions.

We can now prove consistency and normality for $\hat{\theta}_n$. To maintain formulae short we denote $\partial_\sigma f=(\partial/\partial\sigma)f$, $\partial_\mu f=(\partial/\partial\mu)f$, $\partial^2_{\sigma^2}f=(\partial^2/\partial\sigma^2)f$, $\partial^2_{\sigma\mu}f=(\partial^2/\partial\sigma\mu)f,\dots$.

Theorem 7

The estimator $\hat{\theta}_n$ is consistent,

$$\hat{\theta}_n\overset{n\to\infty}{\longrightarrow}\theta_0 \quad \textit{in probability.}$$

Proof. Following the proof of Kessler's (1999) theorem 1, by assumption (S1), it is enough to show that, uniformly in θ,

$$n^{-1}\mathcal{L}_n(\theta)\overset{n\to\infty}{\longrightarrow}\nu_0\left(\frac{a^2(\cdot,\sigma_0)}{a^2(\cdot,\sigma)}+\log a^2(\cdot,\sigma)\right) \quad \text{in probability.} \tag{52}$$

This will ensure the convergence of $\hat{\sigma}_n$ to σ_0. Then, we prove that, uniformly in $\theta=(\mu,\sigma)$,

$$(n\Delta_n)^{-1}(\mathcal{L}_n(\mu,\sigma)-\mathcal{L}_n(\mu_0,\sigma))\overset{n\to\infty}{\longrightarrow}\frac{3}{2}\nu_0\left(\frac{\big(b(\cdot,\mu)-b(\cdot,\mu_0)\big)^2}{a^2(\cdot,\sigma)}\right) \quad \text{in probability.} \tag{53}$$

This enables to obtain the convergence of $\hat{\mu}_n$ to μ_0 (for more details on why (52)–(53) imply consistency, see Kessler, 1997).

We start the proof by (52). With the notations (3)–(5), the contrast (divided by n) writes, after easy computations,

$$\begin{aligned} n^{-1}\mathcal{L}_n(\theta)=&\frac{3}{2}\overline{\mathcal{Q}}_n\Big(a^{-2}(\cdot,\sigma)\Big)+\bar{\nu}_n\Big(\log a^2(\cdot,\sigma)\Big)-3\Delta_n\overline{\mathcal{I}}_n\Big(a^{-2}(\cdot,\sigma)b(\cdot,\mu)\Big)+\frac{3\Delta_n}{4}\overline{\mathcal{Q}}_n\Big(h(\cdot,\theta)\Big)\\ &+\frac{3\Delta_n}{2}\bar{\nu}_n\Big(a^{-2}(\cdot,\sigma)\{b^2(\cdot,\mu)-2b(\cdot,\mu)b(\cdot,\mu_0)\}\Big). \end{aligned} \tag{54}$$

Using proposition 2, theorems 2, 3 and $\Delta_n\to 0$, we easily obtain (52).

For the proof of (53), by the expression of the contrast (divided by n) above, we get

$$(n\Delta_n)^{-1}(\mathcal{L}_n(\mu,\sigma)-\mathcal{L}_n(\mu_0,\sigma))=3\overline{\mathcal{I}}_n\left(\frac{b}{a^2}(\cdot,\mu_0,\sigma)-\frac{b}{a^2}(\cdot,\mu,\sigma)\right)$$
$$-\frac{3}{4}\overline{\mathcal{Q}}_n\left(h(\cdot,\mu_0,\sigma)-h(\cdot,\mu,\sigma)\right)$$
$$+\frac{3}{2}\overline{\nu}_n\left(\frac{\left(b(\cdot,\mu)-b(\cdot,\mu_0)\right)^2}{a^2(\cdot,\sigma)}\right).$$

Now, we apply theorems 2, 3 (recall $h(x,\theta)=\partial_x(b/a^2)(x,\theta)$ too) and proposition 2 to get (53), and the theorem follows.

Remark that for the consistency of the estimator we only need $\Delta_n\to 0$. We now prove that, under the additional condition $n\Delta_n^2\to 0$, the estimator $\hat{\theta}_n$ is asymptotically normal. The scheme of the proof is classical.

Theorem 8

If $n\Delta_n^2\overset{n\to\infty}{\longrightarrow}0$, *then* $\left((n\Delta_n)^{1/2}(\hat{\mu}_n-\mu_0),n^{1/2}(\hat{\sigma}_n-\sigma_0)\right)$ *converges in law to an*

$$\mathcal{N}\left(0,\left\{\nu_0\left(\frac{(\partial_\mu b)^2(\cdot,\mu_0)}{a^2(\cdot,\sigma_0)}\right)\right\}^{-1}\right)\otimes\mathcal{N}\left(0,\frac{9}{16}\left\{\nu_0\left(\frac{(\partial_\sigma a)^2(\cdot,\sigma_0)}{a^2(\cdot,\sigma_0)}\right)\right\}^{-1}\right).$$

Proof. A Taylor's formula around $\hat{\theta}_n$ shows: $\int_0^1\mathcal{C}_n(\theta_0+u(\hat{\theta}_n-\theta_0))\,\mathrm{d}u\;\mathcal{E}_n=\mathcal{D}_n$ where,

$$\mathcal{C}_n(\theta)=\begin{bmatrix}(n\Delta_n)^{-1}\dfrac{\partial^2}{\partial\mu^2}\mathcal{L}_n(\theta) & n^{-1}\Delta_n^{-1/2}\dfrac{\partial^2}{\partial\sigma\mu}\mathcal{L}_n(\theta)\\ n^{-1}\Delta_n^{-1/2}\dfrac{\partial^2}{\partial\mu\sigma}\mathcal{L}_n(\theta) & n^{-1}\dfrac{\partial^2}{\partial\sigma^2}\mathcal{L}_n(\theta)\end{bmatrix},$$

$$\mathcal{E}_n=\begin{bmatrix}(n\Delta_n)^{1/2}(\hat{\mu}_n-\mu_0)\\ n^{1/2}(\hat{\sigma}_n-\sigma_0)\end{bmatrix},\quad \mathcal{D}_n=\begin{bmatrix}-(n\Delta_n)^{-1/2}\dfrac{\partial}{\partial\mu}\mathcal{L}_n(\theta_0)\\ -n^{-1/2}\dfrac{\partial}{\partial\sigma}\mathcal{L}_n(\theta_0)\end{bmatrix}.$$

Now the proof of $\mathcal{E}_n\overset{n\to\infty}{\longrightarrow}\mathcal{N}\left(0,\left\{\nu_0\left(\frac{(\partial_\mu b)^2(\cdot,\mu_0)}{a^2(\cdot,\sigma_0)}\right)\right\}^{-1}\right)\otimes\mathcal{N}\left(0,\frac{9}{16}\left\{\nu_0\left(\frac{(\partial_\sigma a)^2(\cdot,\sigma_0)}{a^2(\cdot,\sigma_0)}\right)\right\}^{-1}\right)$ consists in showing the two following points.

1. We have the convergence in law:

$$\mathcal{D}_n\underset{\mathcal{D}}{\overset{n\to\infty}{\longrightarrow}}\mathcal{N}\left(0,\begin{bmatrix}9\nu_0\left(\dfrac{(\partial_\mu b)^2(\cdot,\mu_0)}{a^2(\cdot,\sigma_0)}\right) & 0\\ 0 & 9\nu_0\left(\dfrac{(\partial_\sigma a)^2(\cdot,\sigma_0)}{a^2(\cdot,\sigma_0)}\right)\end{bmatrix}\right).$$

2. We have the uniform (with respect to θ) convergence in probability:

$$\mathcal{C}_n(\theta)\overset{n\to\infty}{\longrightarrow}\begin{bmatrix}\mathcal{C}_{1,1}(\theta) & 0\\ 0 & \mathcal{C}_{2,2}(\theta)\end{bmatrix}$$

with

$$\mathcal{C}_{1,1}(\theta)=3\nu_0\left(\frac{(\partial_\mu b)^2(\cdot,\mu)}{a^2(\cdot,\sigma)}+\frac{\partial^2_{\mu^2}b}{a^2}(\cdot,\theta)(b(\cdot,\mu)-b(\cdot,\mu_0))\right)$$

$$\mathcal{C}_{2,2}(\theta)=\nu_0\left((\partial_\sigma a)^2(\cdot,\sigma)\left(\frac{6a^2(\cdot,\sigma_0)}{a^4(\cdot,\sigma)}-\frac{2}{a^2(\cdot,\sigma)}\right)\right)$$
$$+\nu_0\left(2\partial^2_{\sigma^2}a(\cdot,\sigma)\left(\frac{1}{a(\cdot,\sigma)}-\frac{a^2(\cdot,\sigma_0)}{a^3(\cdot,\sigma)}\right)\right).$$

Indeed, this second point immediately implies, using the consistency of $\hat{\theta}_n$,

$$\int_0^1 \mathcal{C}_n(\theta_0+u(\hat{\theta}_n-\theta_0))\,\mathrm{d}u \xrightarrow[\mathbf{P}]{n\to\infty} \begin{bmatrix} 3v_0\left(\dfrac{(\partial_\mu b)^2(\cdot,\mu_0)}{a^2(\cdot,\sigma_0)}\right) & 0 \\ 0 & 4v_0\left(\dfrac{(\partial_\sigma a)^2(\cdot,\sigma_0)}{a^2(\cdot,\sigma_0)}\right)\end{bmatrix}.$$

For (1) we remark that $\bar{v}_n$, $\overline{\mathcal{I}}_n$ and $\overline{\mathcal{Q}}_n$ are linear functionals, therefore we can differentiate the expression of $n^{-1}\mathcal{L}_n$ given in the proof of theorem 7 with respect to the parameter μ and get (recall $h(x,\theta)=\partial_x(b/a^2)(x,\theta)$),

$$(n\Delta_n)^{-1/2}\partial_\mu\mathcal{L}_n(\theta_0)=-3\overline{N}_n\left(\frac{\partial_\mu b}{a^2}(\cdot,\theta_0)\right). \tag{55}$$

Analogously, we get

$$\begin{aligned} n^{-1/2}\partial_\sigma\mathcal{L}_n(\theta_0)= &-2\overline{M}_n\left(\frac{\partial_\sigma a}{a^3}(\cdot,\theta_0)\right)\\ &+\sqrt{n}\Delta_n\left\{3\overline{\mathcal{I}}_n\left(\partial_\sigma\left(\frac{b}{a^2}\right)(\cdot,\theta_0)\right)+\frac{3}{4}\overline{\mathcal{Q}}_n\left(\partial_\sigma h(\cdot,\theta_0)\right)\right.\\ &\left.-\frac{3}{2}v_n\left(\partial_\sigma(a^{-2})(\cdot,\sigma_0)b^2(\cdot,\mu_0)\right)\right\}. \end{aligned}$$

By proposition 2, theorems 2, 3 and $n\Delta_n^2\to 0$, this yields

$$n^{-1/2}\partial_\sigma\mathcal{L}_n(\theta_0)=-2\overline{M}_n\left(\frac{\partial_\sigma a}{a^3}(\cdot,\theta_0)\right)+o_{\mathbf{P}}(1). \tag{56}$$

Now, (1) follows from (55)–(56) and theorem 6.

To obtain (2), we differentiate twice $\mathcal{L}_n$ and use results of section 3.

Remark 5.

1. As for estimation based on $(X_{i\Delta_n})_{0\le i\le n-1}$, the rate of convergence is different for $\hat{\mu}_n$ and $\hat{\sigma}_n$. The drift term is estimated with rate $(n\Delta_n)^{1/2}$ and the diffusion term is estimated with rate $n^{1/2}$. Comparing with the asymptotic variance of the estimator based on the Euler contrast (Kessler, 1997; when $X_{i\Delta_n}$ itself is observed), we notice a slight increase in the variance of the estimator of the diffusion term (the constant 9/16 instead of 1/2 for Kessler, 1997). The estimation of μ is asymptotically efficient since $v_0\left(\frac{(\partial_\mu b)^2}{a^2}(\cdot,\theta_0)\right)$ is the Fisher information of the continuous time model.
2. In the case of multidimensional parameters $\mu\in\mathbb{R}^d$, $\sigma\in\mathbb{R}^{d'}$, under the additional assumption that the matrices I_θ, J_θ defined by,

$$I_\theta=\left[v_0\left(\frac{(\partial_{\mu_i}b)(\cdot,\mu)(\partial_{\mu_j}b)(\cdot,\mu)}{a^2(\cdot,\sigma)}\right)\right]_{1\le i,j\le d},\ J_\theta=\left[v_0\left(\frac{(\partial_{\sigma_i}a)(\cdot,\sigma)(\partial_{\sigma_j}a)(\cdot,\sigma)}{a^2(\cdot,\sigma)}\right)\right]_{1\le i,j\le d'}$$

are non-singular at the true value of the parameter, we can prove that theorem 8 holds with the limiting law $\mathcal{N}\left(0,I_\theta^{-1}\right)\otimes\mathcal{N}\left(0,9/16J_\theta^{-1}\right)$.

7. Examples of parametric models

We apply our statistical results to some classical models.

7.1. Example 1: Ornstein–Uhlenbeck process

The diffusion solves $\mathrm{d}X_t = \mu X_t\,\mathrm{d}t + \sigma\,\mathrm{d}B_t$, with $\mu < 0, \sigma > 0$ and X_0 is deterministic or has for distribution the stationary probability of X.

Here, we can compute explicitly the estimator $\hat{\theta}_n$ by minimizing the contrast (51). We find,

$$\hat{\sigma}_n^2 = \frac{3}{2}(n\Delta_n)^{-1}\sum_{i=0}^{n-2}\left(\overline{X}_{i+1} - \overline{X}_i\right)^2$$

$$\hat{\mu}_n = \frac{\Delta_n^{-1}\sum_{i=0}^{n-2}\left(\overline{X}_{i+1} - \overline{X}_i\right)\overline{X}_i}{\sum_{i=0}^{n-2}\left(\overline{X}_i\right)^2} - \frac{1}{4}\frac{\Delta_n^{-1}\sum_{i=0}^{n-2}\left(\overline{X}_{i+1} - \overline{X}_i\right)^2}{\sum_{i=0}^{n-2}\left(\overline{X}_i\right)^2}$$

(we have suppressed negligible terms in $\hat{\sigma}_n^2$).

Using results of section 5.1, (A1)–(A5) are valid with $K_{-\infty} = \infty$, and any positive constants $M_{-\infty}$, M_{∞}. Thus, results of section 6 apply:

$$(\hat{\mu}_n, \hat{\sigma}_n^2) \xrightarrow[\mathbf{P}]{n\to\infty} (\mu, \sigma^2)$$

$$\text{and if } n\Delta_n^2 \to 0, \begin{bmatrix} (n\Delta_n)^{1/2}(\hat{\mu}_n - \mu) \\ n^{1/2}(\hat{\sigma}_n^2 - \sigma^2) \end{bmatrix} \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, \begin{bmatrix} 2|\mu| & 0 \\ 0 & \frac{9}{4}\sigma^4 \end{bmatrix}\right).$$

Using numerical simulations, we see that our estimator gives good results on finite samples (see Table 1, where we give mean and variance of $\hat{\theta}_n$ for different values of n and $T = n\Delta_n$ when $-\mu = \sigma^2 = 1$; we have used 500 replications). It is worth noticing that if Δ is not small enough then we underestimate both σ^2 and $|\mu|$. This is similar to the behaviour of the estimator based on the Euler scheme when $X_{i\Delta_n}$ is observed (see Bibby & Sørensen, 1995).

Table 1. *Simulation results for an Ornstein–Uhlenbeck process.*

	$\hat{\sigma}_n^2$	$\hat{\mu}_n$
$\Delta = 1/100,\ n = 10,000,\ T = 100$		
Mean	0.99	−1.01
n.Var	2.26	
T.Var		2.19
$\Delta = 1/10,\ n = 10,000,\ T = 1000$		
Mean	0.93	−0.96
n.Var	1.84	
T.Var		1.76
$\Delta = 1/2,\ n = 1000,\ T = 500$		
Mean	0.70	−0.82
n.Var	1.21	
T.Var		1.54
$\Delta = 1/100,\ n = 1000,\ T = 10$		
Mean	0.99	−1.20
n.Var	2.24	
T.Var		3.22
$\Delta = 1/10,\ n = 1000,\ T = 100$		
Mean	0.92	−0.98
n.Var	2.19	
T.Var		2.16
$\Delta = 1/2,\ n = 200,\ T = 100$		
Mean	0.69	−0.83
n.Var	1.06	
T.Var		1.75

7.2. Example 2: The Cox–Ingersoll–Ross process

The diffusion solves (49) and starts either from the stationary distribution or from a deterministic variable. We find the following explicit expressions for the estimator $\hat{\theta}_n = (\hat{\mu}_n, \hat{\mu}'_n, \hat{\sigma}^2_n)$ by minimizing the contrast (51):

$$\begin{bmatrix} \Delta_n \sum_{i=0}^{n-2} \overline{X}_i & n\Delta_n \\ n\Delta_n & \Delta_n \sum_{i=0}^{n-2} \overline{X}_i^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mu}_n \\ \hat{\mu}'_n \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{n-2} \{\overline{X}_{i+1} - \overline{X}_i\} \\ \sum_{i=0}^{n-2} \left\{ \overline{X}_i^{-1} (\overline{X}_{i+1} - \overline{X}_i) + \frac{1}{4} \overline{X}_i^{-2} (\overline{X}_{i+1} - \overline{X}_i)^2 \right\} \end{bmatrix}$$

$$\hat{\sigma}^2_n = \frac{3}{2} (n\Delta_n)^{-1} \sum_{i=0}^{n-2} \overline{X}_i^{-1} \left(\overline{X}_{i+1} - \overline{X}_i \right)^2$$

(we have suppressed negligible terms in $\hat{\sigma}^2_n$).

Results of section 6 do not apply since here, $K_0 = c_0 - 1 = 2\mu'/\sigma^2 - 1 < \infty$ and we can only choose $M_0 < c_0$ (see section 5.2). We directly show the convergence of $\hat{\theta}_n$ using its expression and applying results of sections 3 and 4. For this we have to take care that K_0 and M_0 are large enough. After some easy computation we get the following theorem.

Theorem 9

- *If* $c_0 > 9$, *then* $\hat{\theta}_n \xrightarrow[\mathbf{P}]{n\to\infty} \theta_0$.
- *If* $c_0 > 13$ *and* $n\Delta_n^2 \to 0$, *then*

$$\begin{bmatrix} (n\Delta_n)^{1/2}(\hat{\mu}_n - \mu) \\ (n\Delta_n)^{1/2}(\hat{\mu}'_n - \mu') \\ n^{1/2}(\hat{\sigma}^2_n - \sigma^2) \end{bmatrix} \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, \begin{bmatrix} 2|\mu| & \sigma^2 - 2\mu' & 0 \\ \sigma^2 - 2\mu' & (2\mu' - \sigma^2)\frac{\mu'}{|\mu|} & 0 \\ 0 & 0 & \frac{9}{4}\sigma^4 \end{bmatrix}\right).$$

8. Discussion

Investigations in this paper show that even for non-Markovian observations it is possible to introduce an explicit contrast function and obtain an estimator with good asymptotical properties. However, the question of efficiency for the estimation of the diffusion parameter remains opened. In the specific case where X is an Ornstein–Uhlenbeck process, we can improve the estimation of σ^2 and find an estimator $\tilde{\sigma}^2_n$ such that $n^{1/2}(\tilde{\sigma}^2_n - \sigma^2) \xrightarrow{n\to\infty} \mathcal{N}(0, 2\sigma^4)$. Comparing with section 7.1, we see that $\hat{\theta}^2_n$ is not optimal in this case. As suggested by an anonymous referee, an improvement in the estimation of σ seems feasible more generally, by constructing directly a contrast function as a consequence of theorem 1(2). This idea yields to technical difficulties above those of this paper and we were not able to obtain satisfactory results.

The final words are about applicability of the method. Two issues appear. First, our crucial assumption $\Delta_n \to 0$ is not as realistic as a constant sampling interval, however it is suitable for the 'high-frequency' data commonly encountered in finance. Second, in stochastic volatility models the integrated volatility itself may be estimated with some errors which could yield to serious biases.

Acknowledgements

I would like to thank the Editor, an Associate Editor, and the referees for their detailed reading of the paper and for comments which resulted in an improved version of the paper.

References

Barndorff-Nielsen, O. E. (1998). Processes of normal inverse Gaussian type. *Finance Stoch.* **2**, 41–68.

Barndorff-Nielsen, O. E. & Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. Roy. Statist. Soc. Ser. B* **64**, 253–280.

Barndorff-Nielsen, O. E. & Sørensen, M. (1994). A review of some aspects of asymptotic likelihood theory for stochastic processes. *Int. Statist. Rev.* **62**, 133–165.

Bibby, B. M. & Sørensen, M. (1995). Martingale estimations functions for discretely observed diffusion processes. *Bernoulli* **1**, 17–39.

Bibby, B. M. & Sørensen, M. (2001). Simplified estimating functions for diffusion models with a high-dimensional parameter. *Scand. J. Statist.* **28**, 99–112.

Bollerslev, T. & Zhou, H. (2002). Estimating volatility diffusion using conditional moments of integrated volatility. *J. Econometrics* **109**, 33–65.

Delattre, S. & Jacod, J. (1997). A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round-off errors. *Bernoulli* **3**, 1–28.

Ditlevsen, S. & Sørensen, M. (2004). Inference for observations of integrated diffusion processes. *Scand. J. Statist.* **31**, 417–429.

Dorogovtsev, A. (1976). The consistency of an estimate of a parameter of a stochastic differential equation. *Theory Probab. Math. Statist.* **10**, 73–82.

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

Genon-Catalot, V. & Jacod, J. (1993). On the estimation of the diffusion coefficient for multidimensional diffusion processes. *Ann. Inst. H. Poincaré Probab. Statist.* **29**, 119–151.

Genon-Catalot, V., Jeantheau, T. & Laredo, C. (1998). Parameter estimation for discretely observed stochastic volatility models. *Bernoulli* **4**, 283–303.

Gloter, A. (2000). Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM Probab. Stat.* **4**, 205–224.

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

Hall, P. & Heyde, C. C. (1980). *Martingale limit theory and its application*. Academic Press. New York-London.

Ibragimov, I. A. & Khas'minskii, R. Z. (1981). *Statistical estimation, asymptotic theory*. Springer-Verlag, New York.

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

Kessler, M. (2000). Simple and explicit estimating functions for a discretely observed diffusion process. *Scand. J. Statist.* **27**, 65–82.

Kessler, M. & Sørensen, M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli* **5**, 299–314.

Leblanc, B. (1996). Une aproche unifiée pour une forme exacte du prix d'une option dans les différents modèles a volatilité stochastique. *Stoch. Stoch. Rep.* **57**, 1–35.

Pastorello, S., Renault, E. & Touzi, N. (1994). Statistical inference for random variance option pricing. *Southern European Economics Discussion Series*, D.P.136.

Sørensen, M. (1999). On asymptotics of estimating functions. *Braz. J. Probab. Stat.* **13**, 111–136.

Sørensen, M. (2000). Prediction-based estimating functions. *Econ. J.* **3**, 123–147.

Received April 2000, in final form April 2005

Arnaud Gloter, GRAPE, UMR CNRS 5113, Université Bordeaux IV, Avenue Léon Duguit, 33608 PESSAC, France.
E-mail: gloter@u-bordeaux4.fr

Appendix

Technical lemmas

The following lemma precises lemma 8 in Kessler (1997).

Lemma A1
Let $f \in \mathcal{C}^1((l,r)\times\Theta)$*, satisfy*

$$\sup_{\theta\in\Theta}\{|f(x,\theta)| + |f'_x(x,\theta)| + |\nabla_\theta f(x,\theta)|\} \le c(\mathcal{B}_l^{\gamma}(x) + \mathcal{B}_r^{\gamma'}(x)),$$

with $\gamma \le M_l$*,* $1+\gamma' \le M_r$ *and* $\gamma < K_l$ *then:*

$$n^{-1}\sum_{i=0}^{n-1} f(X_{i\Delta_n},\theta) \overset{n\to\infty}{\longrightarrow} \nu_0\left(f(\cdot,\theta)\right) \text{ uniformly in } \theta\text{, in probability.} \tag{57}$$

Proof. We use assumption (A4) and the ergodic theorem for X to obtain the convergence $(n\Delta_n)^{-1}\int_0^{n\Delta_n} f(X_s,\theta)\,\mathrm{d}s \underset{\mathbf{P}}{\overset{n\to\infty}{\longrightarrow}} \nu_0\left(f(\cdot,\theta)\right)$.

Denote for the proof, $D_n(\theta) = (n\Delta_n)^{-1}\sum_{i=0}^{n-1}\int_{i\Delta_n}^{(i+1)\Delta_n}\{f(X_s,\theta) - f(X_{i\Delta_n},\theta)\}\,\mathrm{d}s$. Using proposition 1(1) and (A5) yields $\sup_{s\in[i\Delta_n,(i+1)\Delta_n]} E(|f(X_s,\theta) - f(X_{i\Delta_n},\theta)|) \le c\sqrt{\Delta_n}$. We deduce that $D_n(\theta) \overset{n\to\infty}{\longrightarrow} 0$ in $\mathbf{L}^1$, which gives the convergence in (57) for all θ. To get the uniformity in θ, by proposition A1, it suffices to show: $\sup_{n\ge 0} E(\sup_\theta |n^{-1}\sum_{i=0}^{n-1}\nabla_\theta f(X_{i\Delta_n},\theta)|) < \infty$. Using $\sup_\theta |\nabla_\theta f|(x,\theta) \le c(B_l^{M_l}(x) + \mathcal{B}_r^{M_r}(x))$ and (A5) yields the result.

Proposition A1
Let $S_n(\omega,\theta)$ *be a sequence of measurable real valued functions defined on* $\Omega\times\Theta$*, where* $(\Omega,\mathcal{F},P)$ *is a probability space, and* Θ *is product of compact intervals of* $\mathbb{R}$*. We assume that* $S_n(\cdot,\theta)$ *converges to zero in probability for all* $\theta\in\Theta$*; and that there exists an open neighbourhood of* Θ *on which* $S_n(\omega,\cdot)$ *is continuously differentiable for all* $\omega\in\Omega$*. Furthermore, we suppose that* $\sup_{n\in\mathbb{N}} E\left(\sup_{\theta\in\Theta}|\nabla_\theta S_n(\theta)|\right) < \infty$*, then*

$$S_n(\theta) \overset{n\to\infty}{\longrightarrow} 0 \text{ uniformly in } \theta\text{, in probability.}$$

Proof. Let $\epsilon > 0$ and $\eta > 0$, let us show that for n large enough: $P(\sup_{\theta\in\Theta}|S_n(\theta)| > \eta) < \epsilon$. Denote $Z_n = \sup_{\theta\in\Theta}|\nabla_\theta S_n(\theta)|$, and let M such that $\sup_{n\in\mathbb{N}} E\left(Z_n\right) M^{-1} < \epsilon/3$.

Using that Θ is compact we can find an integer d and $(\alpha_1,\dots,\alpha_d) \in \Theta^d$, such that for all θ in Θ: $\inf_{i\in\{1,\dots,d\}}|\alpha_i - \theta| < \eta/(2M)$. Define n_0 such that, $n \ge n_0$ implies $P\left(|S_n(\alpha_i)| > \eta/2\right) < \epsilon/(3d)$. Using that Θ is convex: $\sup_{\theta\in\Theta}|S_n(\theta)| \le Z_n\eta/(2M) + \sup_{i=1,\dots,d}|S_n(\alpha_i)|$. We deduce, $P(\sup_{\theta\in\Theta}|S_n(\theta)| \ge \eta) \le P(Z_n\eta/(2M) \ge \eta/2) + \sum_{i=1}^d P(|S_n(\alpha_i)| \ge \eta/2)$, and conclude using the Bienayme–Tchebychev inequality.

We recall, the useful lemma which is given in Genon-Catalot & Jacod (1993).

Lemma A2
Let χ_i^n*,* U *be random variables, with* χ_i^n *being* $\mathcal{G}_i^n$*-measurable. The following two conditions imply* $\sum_{i=1}^n \chi_i^n \overset{\mathbf{P}}{\to} U$*:*

$$\sum_{i=1}^n E(\chi_i^n \mid \mathcal{G}_{i-1}^n) \overset{\mathbf{P}}{\to} U,$$

$$\sum_{i=1}^n E((\chi_i^n)^2 \mid \mathcal{G}_{i-1}^n) \overset{\mathbf{P}}{\to} 0.$$

Details on the proof of theorem 2

First, we study $\overline{\mathcal{I}}_n^{(1)}(\theta)$. By (9) and (10), we have

$$E\left(v_{2i,2n}^{(1)}(\theta)\,|\,\mathcal{G}_{2i}^{2n}\right)=0 \text{ and } E\left((v_{2i,2n}^{(1)}(\theta))^2\,|\,\mathcal{G}_{2i}^{2n}\right)=\frac{2\Delta_{2n}}{3}a^2(X_{2i\Delta_{2n}})f^2(X_{2i\Delta_{2n}},\theta).$$

We deduce, using $n\Delta_n\to\infty$, $2\gamma\le M_l$, $2+2\gamma\le M_r$ and (A5) that

$$(n\Delta_{2n})^{-2}\sum_{i=0}^{n-1}E\left((v_{2i,2n}^{(1)}(\theta))^2\,|\,\mathcal{G}_{2i}^{2n}\right)=(n\Delta_{2n})^{-1}\left(\frac{2}{3}n^{-1}\sum_{i=0}^{n-1}a^2(X_{2i\Delta_{2n}})f^2(X_{2i\Delta_{2n}},\theta)\right)\xrightarrow{\mathbf{L^1}}0.$$

Now lemma A2, again, yields the convergence $\overline{\mathcal{I}}_n^{(1)}(\theta)\xrightarrow{\mathbf{P}}0$ for each θ in Θ. To prove that this convergence is uniform we cannot use proposition A1 because $E\left(\left|v_{2i,2n}^{(1)}(\theta)\right|\,|\,\mathcal{G}_{2i}^{2n}\right)$ is of order $\sqrt{\Delta_{2n}}$. We will instead use theorem 20 in appendix 1 of Ibragimov & Khas'minskii (1981). It is enough to show that there exist two constants $M\ge 0$, and $\epsilon>0$ such that:

$$\forall\theta,n,\quad E\left(\left|\overline{\mathcal{I}}_n^{(1)}(\theta)\right|^{2+\epsilon}\right)\le M\quad\text{and}\quad\forall\theta,\theta',n,\quad D_n(\theta,\theta')\le M\left|\theta-\theta'\right|^{2+\epsilon}\tag{58}$$

with $D_n(\theta,\theta')=E\left(\left|\overline{\mathcal{I}}_n^{(1)}(\theta)-\overline{\mathcal{I}}_n^{(1)}(\theta')\right|^{2+\epsilon}\right)$.

We only prove the first inequality (the second one is similar). Using Rosenthal's inequality for martingales (see Hall & Heyde, 1980, p. 23), we get for any $\epsilon>0$,

$$E\left(\left|\overline{\mathcal{I}}_n^{(1)}\right|^{2+\epsilon}\right)\le(n\Delta_{2n})^{-2-\epsilon}E\left(\left|\sum_{i=0}^{n-1}E\left((v_{2i,2n}^{(1)})^2\,|\,\mathcal{G}_{2i}^{2n}\right)\right|^{1+\epsilon/2}\right)$$
$$+(n\Delta_{2n})^{-2-\epsilon}\sum_{i=0}^{n-1}E\left(\left|v_{2i,2n}^{(1)}\right|^{2+\epsilon}\right).$$

By the classical inequality, for $p=1+\epsilon/2$, $(\sum_{i=0}^{n-1}|a_i|)^p\le n^{p-1}\sum_{i=0}^{n-1}|a_i|^p$, we have

$$E\left(\left|\sum_{i=0}^{n-1}E\left((v_{2i,2n}^{(1)})^2\,|\,\mathcal{G}_{2i}^{2n}\right)\right|^{1+\epsilon/2}\right)\le n^{\epsilon/2}\sum_{i=0}^{n-1}E\left(\left|E\left((v_{2i,2n}^{(1)})^2\,|\,\mathcal{G}_{2i}^{2n}\right)\right|^{1+\epsilon/2}\right)$$

But, if ϵ is small enough, $2\gamma(1+\epsilon)\le M_l$ and $(2\gamma+1)(1+\epsilon)\le M_r$ and by (A5) we deduce

$$\sup_{i,n}E\left(\left|E\left((v_{2i,2n}^{(1)})^2\,|\,\mathcal{G}_{2i}^{2n}\right)\right|^{1+\epsilon/2}\right)\le c\Delta_{2n}^{1+\epsilon/2},\quad\sup_{i,n}E\left(\left|v_{2i,2n}^{(1)}\right|^{2+\epsilon}\right)\le c\Delta_{2n}^{1+\epsilon/2}.$$

Hence, $E\left(\left|\overline{\mathcal{I}}_n^{(1)}\right|^{2+\epsilon}\right)\le c\{(n\Delta_{2n})^{-1-\epsilon/2}+(n\Delta_{2n})^{-1-\epsilon/2}n^{-\epsilon/2}\}$. As $(n\Delta_{2n})^{-1}$ is bounded, we obtain (58). So, $\overline{\mathcal{I}}_n^{(1)}(\theta)\underset{\mathbf{P}}{\xrightarrow{n\to\infty}}0$ uniformly in θ.

The convergence to zero of $\overline{\mathcal{I}}_n^{(3)}(\theta)$ and $\overline{\mathcal{I}}_n^{(5)}(\theta)$ is easily obtained since we have by (A1), (A2), CU_γ, proposition 1(2) and (16):

$$E\left(\sup_{\theta\in\Theta}\left|v_{2i,2n}^{(3)}(\theta)\right|+\sup_{\theta\in\Theta}\left|v_{2i,2n}^{(5)}(\theta)\right|\,|\,\mathcal{G}_{2i}^{2n}\right)$$
$$\le c\Delta_{2n}^{3/2}\left(\mathcal{B}_l^{(\gamma+\alpha_1)\vee(\gamma+\alpha_2/2)}(X_{2i\Delta_{2n}})+\mathcal{B}_r^{3+\gamma+\alpha_1+\alpha_2/2}(X_{2i\Delta_{2n}})\right).$$

Now, we treat $\overline{\mathcal{I}}_n^{(4)}(\theta)$. Using (14) and (15) we show that, if $(2\gamma+2\alpha_1)\vee(2\gamma+\alpha_2)\vee(\gamma+\beta_1+\beta_2)\leq M_l$ and $(2+\gamma+\beta_1+\beta_2)\vee(3+2\gamma+2\alpha_1+\alpha_2)\leq M_r$, then

$$(n\Delta_n)^{-1}\sum_{i=0}^{n-1}E\left(v_{2i,2n}^{(4)}(\theta)\,|\,\mathcal{G}_{2i}^{2n}\right)\xrightarrow[\mathbf{L^1}]{n\to\infty}0,\quad (n\Delta_n)^{-1}\sum_{i=0}^{n-1}E\left((v_{2i,2n}^{(4)}(\theta))^2\,|\,\mathcal{G}_{2i}^{2n}\right)\xrightarrow[\mathbf{L^1}]{n\to\infty}0.$$

We deduce that for all θ, $\overline{\mathcal{I}}_n^{(4)}(\theta)\xrightarrow[\mathbf{P}]{n\to\infty}0$. Furthermore, this convergence is uniform in θ by application of proposition A1 and since we can show $E(\sup_\theta \nabla_\theta|v_{2i,2n}^{(4)}(\theta)|)\leq c\Delta_{2n}$.

Details on the proof of theorem 3

A few computations based on theorem 1 gives, for $l=2,3,4$,

$$E\left(\left|w_{2i,2n}^{(l)}(\theta)\right|\,|\,\mathcal{G}_{2i}^{2n}\right)\leq\Delta_{2n}^{3/2}\left(\mathcal{B}_l^{(\gamma+2\alpha_1)\vee(\gamma+\alpha_2)}(X_{2i\Delta_{2n}})+\mathcal{B}_r^{4+\gamma+2\alpha_1+\alpha_2}(X_{2i\Delta_{2n}})\right).$$

Now, (A5) implies the $\mathbf{L}^1$ convergence to 0 for $\overline{\mathcal{Q}}_n^{(l)}(\theta)$, with $l=2,3,4$.