Available online at www.sciencedirect.com

SciVerse ScienceDirect

stochastic processes and their applications

Stochastic Processes and their Applications 122 (2012) 2521–2552

www.elsevier.com/locate/spa

A contrast estimator for completely or partially observed hypoelliptic diffusion

Adeline Samson[a,*], Michèle Thieullen[b]

[a] PRES Sorbonne Paris Cité, Université Paris Descartes, Laboratoire MAP5 UMR CNRS 8145, 45 rue des St Pères, 75006 Paris, France

[b] Université Pierre et Marie Curie - Paris 6, Laboratoire de Probabilités et Modèles Aléatoires (LPMA), UMR CNRS 7599, Boîte 188, 4, Place Jussieu, 75252 Paris cedex 05, France

Received 6 June 2011; received in revised form 11 April 2012; accepted 12 April 2012
Available online 25 April 2012

Abstract

Parametric estimation of two-dimensional hypoelliptic diffusions is considered when complete observations – both coordinates discretely observed – or partial observations – only one coordinate observed – are available. Since the volatility matrix is degenerate, Euler contrast estimators cannot be used directly. For complete observations, we introduce an Euler contrast based on the second coordinate only. For partial observations, we define a contrast based on an integrated diffusion resulting from a transformation of the original one. A theoretical study proves that the estimators are consistent and asymptotically Gaussian. A numerical application to Langevin systems illustrates the nice properties of both complete and partial observations' estimators.

Keywords: Hypoelliptic diffusion; Langevin system; Stochastic differential equations; Partial observations; Contrast estimator

1. Introduction

In this paper, we consider parametric estimation for hypoelliptic diffusions. We focus on two dimensional diffusions, which are generalizations of systems called *Langevin* or *hypoelliptic* by different communities. They appear in many domains such as random mechanics, finance

* Corresponding author.
E-mail addresses: adeline.samson@parisdescartes.fr (A. Samson), michele.thieullen@upmc.fr (M. Thieullen).

0304-4149/$ - see front matter
doi:10.1016/j.spa.2012.04.006

modeling and biology. Their common form is as follows:

$$\begin{cases} dY_t = g(Y_t, Z_t)dt \\ dZ_t = \beta(Y_t, Z_t)dt + \alpha(Y_t, Z_t)dB_t \end{cases} \tag{1}$$

where g, β and α are real functions depending on unknown parameters θ. In these systems, noise acts directly on the "speed" Z_t and on the "position" Y_t only through Z_t. We refer to [15] for examples of such systems arising in applications.

In some applications, it is not possible to measure the two coordinates. Therefore, we consider two observations cases. The complete observations case assumes that both (Y_t) and (Z_t) are discretely observed. The partial observations case assumes that only the first coordinate (Y_t) is observed.

Statistical inference for discretely observed diffusion processes is complex and has been widely investigated (see e.g. [16,18]). It is not possible in general to express the density of stochastic differential equation (SDE) explicitly. So different types of contrast estimators have been introduced for elliptic SDEs estimation, such as the multidimensional Euler contrast [6,10]. However for the hypoelliptic system (1), Euler contrast methods are not directly applicable as the volatility matrix is noninvertible. References on hypoelliptic estimations are few, even in the case of complete observations. The main paper is [15]. They propose an empirical approximation of the likelihood based on Itô–Taylor expansion so that the variance matrix becomes invertible. They construct a Bayesian estimator of θ based on a Gibbs sampler. They consider both complete and partial observation cases. Their method is limited to $g(Y_t, Z_t) = Z_t$, a drift function $\beta(Y_t, Z_t)$ which is linear with respect to the parameter and a constant volatility function $\alpha(Y_t, Z_t)$. In this paper, we consider more general models. We assume that g belongs to a family of functions such that it is possible to reduce to the case of integrated diffusions with a non-autonomous diffusion for (Z_t). Then, we propose to reduce to an Euler contrast based only on the second equation. This allows to consider general drift and volatility functions. We prove the consistency and the asymptotic normality of this contrast estimator when the number of observations $n \to \infty$ and the time step between two observations $\Delta_n \to 0$.

The case of partial observations introduces more difficulties because (Y_t) is not Markovian while (Y_t, Z_t) is Markovian. A maximum-likelihood estimation from discrete and partial observations of a two-dimensional linear system with a non-degenerate volatility function has been proposed [3]. However, their approach cannot be extended to a degenerate volatility function. Main references for partial observations of hypoelliptic diffusions are when the function $g(Y_t, Z_t)$ is equal to Z_t. In this case, model (1) can be viewed as an integrated diffusion process. Parametric estimation methods have been proposed in this context under the additional condition that Z_t satisfies an autonomous equation, meaning that the only coupling between Y_t and Z_t is through the identity $Y_t = \int_0^t Z_s ds$. Prediction-based estimating functions have been studied [2]. Gloter (2006) proposes an Euler contrast function and studies the properties of this estimator when the sampling interval Δ_n tends to zero [8]. However, their approaches are not adapted when Z_t does not satisfy an autonomous equation and when $g(Y_t, Z_t) \neq Z_t$. In this paper, we extend the approach of [8] to this case. We prove the consistency and the asymptotic normality of this contrast estimator when $\Delta_n \to 0$ when $n \to \infty$.

In order to establish asymptotic properties of our estimators we need existence and uniqueness of an invariant measure for system (1). This is a major difference with respect to Gloter's work since in his framework the second component Z_t satisfies an autonomous equation. Hence the invariant measure he introduces is that of a one dimensional diffusion. In our case, we need an invariant measure for the vector (Y_t, Z_t). Ergodicity of Langevin systems has been widely studied, relying on the hypoellipticity of the system as well as a Lyapunov condition involving a Lyapunov function [12,13]. We detail these conditions and propose examples where our assumptions are verified. A numerical study is performed on these examples, to which we compare results obtained by [15].

The paper is organized as follows. Section 2 presents the hypoelliptic system, general assumptions and more details for Langevin systems. Section 3 defines the two observations cases and the contrast estimators. The main results are presented, which consist in consistency and asymptotic normality of both estimators. Asymptotic properties of functionals of the processes are given in Section 4. Proofs of the estimator asymptotic properties are given in Section 5. Estimation methods are illustrated in Section 6 on simulated data. Section 7 presents some conclusions and discussions. Supplementary proofs are given in Appendix.

2. Hypoelliptic system and assumptions

2.1. The model

Let us consider system (1) and assume that the following condition holds

(C1) $\forall (y,z) \in \mathbb{R} \times \mathbb{R}, \partial_z g(y,z) \neq 0$.

Under assumption (C1), system (1) is hypoelliptic in the sense of stochastic calculus of variations [14]. Indeed, the Stratonovich form of (1) is

$$\begin{cases} dY_t = g(Y_t, Z_t)dt \\ dZ_t = \tilde{\beta}(Y_t, Z_t)dt + \alpha(Y_t, Z_t) \circ dB_t \end{cases} \tag{2}$$

with $\tilde{\beta}(y,z) := \beta(y,z) - \frac{1}{2}\alpha(y,z)\partial_z\alpha(y,z)$. Writing the coefficients of (2) as vector fields

$$A_0(y,z) = \begin{pmatrix} g(y,z) \\ \tilde{\beta}(y,z) \end{pmatrix} \quad \text{and} \quad A_1(y,z) = \begin{pmatrix} 0 \\ \alpha(y,z) \end{pmatrix}$$

and computing their Lie bracket leads to

$$[A_0, A_1] = \begin{pmatrix} \partial_z g(y,z) \\ \gamma(y,z) \end{pmatrix}.$$

The form of γ is explicit but not detailed here. Under condition (C1) the vectors A_1 and $[A_0, A_1]$ generate $\mathbb{R}^2$ and system (1) is hypoelliptic. We will discuss the consequence of this property in Section 2.2.

Condition (C1) plays also a crucial role to reduce model (1) to an integrated diffusion. Indeed, by the change of variable $X_t := g(Y_t, Z_t)$, the first equation of system (1) becomes $dY_t = X_t dt$ which suggests that the process (Y_t, X_t) should be an integrated diffusion. Condition (C1) enables us to apply the implicit function theorem which states that Z_t can be uniquely defined as a function of (Y_t, X_t). Consequently the vector (Y_t, X_t) satisfies

$$\begin{cases} dY_t = X_t dt \\ dX_t = b(Y_t, X_t)dt + a(Y_t, X_t)dB_t \end{cases} \tag{3}$$

where b and a result from the combination of the implicit function theorem and Itô formula. However the result of the implicit function theorem is only local and no explicit expression is available in general for Z_t as a function of (Y_t, X_t). Therefore in this paper, we assume that system (1) verifies the following condition

(C2) The process (Y_t, X_t) with $X_t := g(Y_t, Z_t)$ satisfies a system of the form (3) with explicit functions b and a.

This includes in particular functions g for which an explicit function f is available such that $Z_t = f(Y_t, X_t)$. Examples are $g(y, z) = \theta_1 y + \theta_2 z$ or $g(y, z) = \phi_\theta(y) + \theta_2 z$ for a function ϕ_θ which depends on parameter θ. This condition is also satisfied for more general systems. The following system

$$\begin{cases} dY_t = -(\theta_1 Y_t - \theta_2 Z_t^2)dt \\ dZ_t = -(\theta_3 Z_t + Z_t F_\theta(Y_t))dt + \alpha(Y_t, Z_t)dB_t \end{cases}$$

with $F \in C^\infty(\mathbf{R}, [0, +\infty[)$ a possibly non-linear function depending on parameter θ and $\alpha(Y_t, Z_t) = \sigma Z_t$ is an example where the change of variables $X_t := g(Y_t, Z_t)$ yields to explicit functions b and a, even if there exists no explicit function f such that $Z_t = f(Y_t, X_t)$. Variants with volatility functions $\alpha(Y_t, Z_t) = \sigma Z_t/(1+Z_t^2)$ or $\alpha(Y_t, Z_t) = \sigma Z_t F(Y_t)$ are other examples of systems that we consider in this paper.

In this paper, we focus on systems (1) for which conditions (C1) and (C2) hold. The first step of the estimation method consists in transforming system (1) into system (3). We denote μ and σ the parameters of functions b and a, respectively. These parameters include parameters of functions g, β and α of system (1). Our parameter is the vector $(\mu, \sigma^2) = \theta$. In the sequel, we denote $b_\mu(Y_t, X_t)$ and $a_\sigma(Y_t, X_t)$ the drift and volatility functions.

2.2. Assumptions

We assume that the vector θ belongs to $\Theta = \Theta_1 \times \Theta_2$ for $\Theta_1 \subset \mathbb{R}^{d_1}$ and $\Theta_2 \subset \mathbb{R}^{d_2}$ two compact subsets.

We now come to the assumptions regarding the drift and volatility functions. In this paper we work under conditions (C1)–(C2). In the present section we list our additional assumptions (A1)–(A4). Then we provide a set of sufficient conditions (S1)–(S3) ensuring that these assumptions are satisfied. We also examine the particular case of Langevin systems.

(A1) (a) There exists a constant c such that $\sup_{\sigma\in\Theta_2} |a_\sigma^{-1}(y, x)| \le c(1 + |y| + |x|)$

(b) for all $\theta \in \Theta$, b_μ and a_σ belong to the class $\mathcal{F}$ of functions $f \in \mathcal{C}^2(\mathbb{R}^2)$ for which there exists a constant c such that the function, its first and second partial derivatives with respect to y and x are bounded by $c(1 + |y| + |x|)$, for all $x, y \in \mathbb{R}$, uniformly in θ.

(A2) (a) $\forall k \in]0, \infty[\sup_{t\ge 0} \mathbb{E}(|X_t|^k + |Y_t|^k) < \infty$

(b) there exists a constant c such that $\forall t \ge 0$, $\forall \delta \ge 0$,

$$\mathbb{E}\left(\sup_{s\in[t,t+\delta[} |X_s|^k|\mathcal{G}_t\right) + \mathbb{E}\left(\sup_{s\in[t,t+\delta[} |Y_s|^k|\mathcal{G}_t\right) \le c(1 + |X_t|^k + |Y_t|^k)$$

where $\mathcal{G}_t = \sigma(B_s, s \le t)$.

(A3) (Y_t, X_t) admits a unique invariant probability measure ν_0 with finite moments of any order i.e. $\forall k > 0$, $\nu_0(|\cdot|^k) < \infty$.

(A4) (Y_t, X_t) satisfies a weak version of the ergodic theorem namely

$$\frac{1}{T}\int_0^T f(Y_s, X_s)ds_0(f) \quad \text{a.s}$$

for any continuous function f with polynomial growth at infinity.

Remark 1. 1. Actually we need assumption (A2) only for $k \leq 4$ to prove the properties of our estimators.

2. We need (A4) for all $f \in \{a^j, a^j \log a^2, b^k/a^j, (\partial b)^k/a^j, (\partial^2 b)^k/a^j, (\partial a^2)^k/a^j, (\partial^2 a^2)^k/a^j, j \in \{0, 1, 2, 4, 6\}, k \in \{0, 1, 2\}\}$. These have indeed polynomial growth at infinity thanks to (A1).

We now provide a set of sufficient conditions, sometimes called stability conditions, for (A2)–(A4) to hold. They are based on the existence of a function V, called Lyapunov function. Lyapunov functions are efficient tools in the asymptotic study of systems; their use is classical for the Langevin systems that we consider in Section 6.

(S1) For all $\theta \in \Theta$,
 (a) $V(y,x) \geq 1$, $\lim_{\|(y,x)\| \to +\infty} V(y,x) = +\infty$
 (b) there exist $c_1 > 0$ and $c_2 > 0$ such that $L_\theta V(y,x) \leq -c_1 V(y,x) + c_2, \forall (y,x) \in \mathbb{R}^2$ where L_θ denotes the infinitesimal generator corresponding to (3).

(S2) (Y_t, X_t) admits a unique invariant probability measure ν_0.

(S3) For all $\theta \in \Theta$, $\exists C > 0$ such that $(a_\sigma(y,x)\partial_x V(y,x))^2 \leq C V(y,x)$ for all $(y,x) \in \mathbb{R}^2$.

Assumption (S1) implies existence and uniqueness of a solution to system (3) as well as existence of an invariant probability measure. Moreover, under (S1) the process $S_t := \mathrm{e}^{c_1 t}(V(Y_t, X_t) - \frac{c_2}{c_1})$ is a local submartingale, hence (cf. [17]) for all $k \geq 1$ and all $t \geq 0$,

$$\mathbb{E}\left(\sup_{s \in [t, t+\delta[} |S_s|^k | \mathcal{G}_t\right) \leq \left(\frac{k}{k-1}\right)^k \mathbb{E}\left(|S_{t+\delta}|^k \,|\mathcal{G}_t\right) \leq \left(\frac{k}{k-1}\right)^k |S_t|^k. \tag{4}$$

Uniqueness of the invariant probability measure is not guaranteed by (S1) and is the purpose of assumption (S2). We show now that, if there exists a polynomial Lyapunov function (or a Lyapunov function with polynomial growth at infinity), then (S1)–(S3) imply (A2)–(A4). The examples of Section 6 admit quadratic Lyapunov functions. So, let us assume here that a polynomial V in y and x satisfies (S1). An analogous argument can be used when V is dominated at infinity by a polynomial. From (4) and the polynomial character of V we deduce (A2) and (A3). From (S1) to (S3), we know that for any $\lambda < \frac{2c_1}{C}$, the function $\tilde{V}(y,x) := \exp(\lambda V(y,x))$ satisfies (S1) or in other words is a Lyapunov function, and also that $\forall f \in \mathcal{C}$, $\frac{1}{T}\int_0^T f(Y_s, X_s)ds \xrightarrow[T\to\infty]{} \nu_0(f)$ a.s. where $\mathcal{C}$ denotes the class of measurable functions f such that $|f|$ is negligible w.r.t. $\exp(\frac{\lambda}{2}V(y,x))$. The class $\mathcal{C}$ contains all polynomials.

As already mentioned, (S3) is satisfied when a_σ is constant and V quadratic at infinity. The following assumption can also be used

(S3') For all $\theta \in \Theta$, $\exists C > 0$ and $\zeta \in [0, 1[$ such that $(a_\sigma(y,x)\partial_x V(y,x))^2 \leq C V(y,x)^{2-\zeta}$ for all $(y,x) \in \mathbb{R}^2$.

In this case it is still possible to generate Lyapunov functions from V which are polynomials in V of bounded degree and a weak version of (A4) holds on a class $\mathcal{C}$ which contains all polynomials of degree smaller than some value. The reader can find more details about (S1)–(S3) and their consequences for the long time behavior of (3) in [12].

We test our estimator numerically in Section 6 on particular Langevin systems. Such systems are defined by

$$\begin{cases} dY_t = X_t dt \\ dX_t = [-\gamma X_t - F'_D(Y_t)]dt + \sigma dW_t \end{cases} \tag{5}$$

with $\sigma > 0$, $F \in C^\infty(\mathbf{R}, [0, +\infty[)$ is a possibly non-linear function depending on parameter D and F' denotes the derivative of F w.r.t. y.

For these systems the invariant probability ν_0 is unique and admits the density

$$\rho(y,x) = C \exp -\frac{\gamma}{\sigma^2}(x^2 - 2F_D(y))$$

where C is a multiplicative constant. Hence (A3) is fulfilled. Stability conditions for these systems are presented in [13]. If $F_D(y) \geq 0, \forall y \in \mathbf{R}$ and satisfies

$$\beta F_D(y) - \frac{1}{2}F'_D(y)y + \frac{\gamma^2}{8}\frac{\beta(2-\beta)}{(1-\beta)}y^2 \leq \alpha \tag{6}$$

for some $\beta \in]0, 1[$ and $\alpha > 0$, a Lyapunov function (which satisfies (S1)) is provided by

$$V(y,x) = \frac{1}{2}x^2 + F_D(y) + \frac{\gamma}{2}\langle y, x\rangle + \frac{\gamma^2}{4}y^2 + 1 \tag{7}$$

and condition (S3) is fulfilled. Note that the hypoelliptic property of these Langevin systems is exploited in [13] in order to establish their geometric ergodicity. In our numerical Section 6 we study respectively $\gamma = 0$, $F_D \equiv 0$ which corresponds to our Model I, $\gamma > 0$, $F_D(y) \equiv \frac{D}{2}y^2$ in Model II and $\gamma > 0$, $F_D(y) \equiv -\Sigma_{j=1}^n j^{-1}D_j(\cos y)^j$ in Model III. In these three examples (A1) is satisfied as well as (A3) (cf. [13]). Moreover a_σ is constant and V quadratic so (S3) holds which, as noticed previously, implies (A2) and (A4). Moreover, Models II and III satisfy (6).

3. Estimators and their properties

In this section, we first present the two observations frameworks. Then, for both frameworks, we introduce a discretized scheme of the system. The properties of these schemes are studied. They yield to the definition of the two contrast functions. Finally, we present the main results of the two contrast estimators, namely their consistency and the asymptotic normality.

3.1. Observations

We assume that (Y_t, X_t) is the unique solution of the system

$$\begin{cases} dY_t = X_t dt \\ dX_t = b_{\mu_0}(Y_t, X_t)dt + a_{\sigma_0}(Y_t, X_t)dB_t \end{cases} \tag{8}$$

where $\theta_0 = (\mu_0, \sigma_0)$ is the true value of the parameter and functions b_μ, a_σ are such that assumptions (A1)–(A4) are fulfilled. We assume that $\theta_0 \in \Theta$. From now on, we set $b(y,x) = b_{\mu_0}(y,x)$ and $a(y,x) = a_{\sigma_0}(y,x)$.

Now, we describe the two observations frameworks. The first case assumes that both components (Y_t) and (X_t) are observed at discrete times $0 = t_0 < t_1 < \cdots < t_n$. The second case assumes that the process $(X_t)_{t\geq 0}$ is hidden or not observed and that we only observe at discrete times t_i the process $(Y_t)_{t\geq 0}$. In both cases, we assume that discrete times are equally spaced and denote $\Delta_n = t_i - t_{i-1}$ the step size, so $t_i = i\Delta_n$. We denote $(Y_{i\Delta_n}, X_{i\Delta_n})$ the observation of the bidimensional process $(Y_t, X_t)_{t\geq 0}$ at time t_i for the first case, and $(Y_{i\Delta_n})$ the observation of the process $(Y_t)_{t\geq 0}$ for the second case. Our purpose is to estimate θ from the complete and partial observations. As for notation, in the sequel we use an upper index C (resp. P) for the case of complete (resp. partial) observations. The asymptotic behavior of the two estimators is studied for a step size Δ_n such that $\Delta_n \to 0$, as $n \to \infty$, $n\Delta_n \to \infty$ and $n\Delta_n^2 \to 0$.

3.2. A contrast estimator for complete observations

When (Y_t) and (X_t) are both observed at discrete times $(i\Delta_n)$, we can consider the classical two-dimensional Euler–Maruyama approximation $(\widetilde{Y}_{(i+1)\Delta_n}, \widetilde{X}_{(i+1)\Delta_n})$ of $(Y_{(i+1)\Delta_n}, X_{(i+1)\Delta_n})$ which is

$$\begin{pmatrix}\widetilde{Y}_{(i+1)\Delta_n}\\ \widetilde{X}_{(i+1)\Delta_n}\end{pmatrix} = \begin{pmatrix}\widetilde{Y}_{i\Delta_n}\\ \widetilde{X}_{i\Delta_n}\end{pmatrix} + \Delta_n \begin{pmatrix}\widetilde{X}_{i\Delta_n}\\ b(\widetilde{Y}_{i\Delta_n}, \widetilde{X}_{i\Delta_n})\end{pmatrix} + \sqrt{\Delta_n}\Sigma \begin{pmatrix}\eta_i^1\\ \eta_i^2\end{pmatrix}, \tag{9}$$
$$\Sigma = \begin{pmatrix}0 & 0\\ 0 & a(\widetilde{Y}_{i\Delta_n}, \widetilde{X}_{i\Delta_n})\end{pmatrix}$$

with (η_i^1, η_i^2) independent identically distributed centered Gaussian vector.

The two-dimensional Euler contrast cannot be used directly to estimate parameters θ because Σ is not invertible. To circle this problem, [15] considers an Itô–Taylor expansion of higher order, by adding the first non-zero noise term arising in the first coordinate. This yields to an invertible covariance matrix for some hypoelliptic models, which may be complex to calculate.

On the contrary, our estimation approach remains based on the Euler scheme. As said previously, it cannot be used directly. However, as we focus on parameter estimation of drift and volatility functions of the second coordinate which is observed in this subsection, we propose to consider a contrast based on the Euler–Maruyama approximation of this second equation. Dependence between successive terms $(X_{i\Delta_n})$ are described in the following Proposition:

Proposition 1. *Set $\mathcal{G}_i^n = \mathcal{G}_{i\Delta_n}$. We have*

$$X_{(i+1)\Delta_n} - X_{i\Delta_n} - \Delta_n b(Y_{i\Delta_n}, X_{i\Delta_n}) = a(Y_{i\Delta_n}, X_{i\Delta_n})\eta_{i,n} + \varepsilon_{i,n}^C$$

where $\eta_{i,n}$ is such that $\mathbb{E}(\eta_{i,n}^{2k+1}|\mathcal{G}_i^n) = 0$ and $\mathbb{E}(\eta_{i,n}^{2k}|\mathcal{G}_i^n) = (2k)!/(2^k k!)\Delta_n^k$ for $k \geq 0$; $\varepsilon_{i,n}^C$ is such that $\mathbb{E}(|\varepsilon_{i,n}^C|\,|\mathcal{G}_i^n) \leq c\Delta_n^{3/2}(1+|Y_{i\Delta_n}|+|X_{i\Delta_n}|)$ and $\mathbb{E}(|\varepsilon_{i,n}^C|^k|\mathcal{G}_i^n) \leq c\Delta_n^{k/2+1}(1+|Y_{i\Delta_n}|^k+|X_{i\Delta_n}|^k)$ for $k \geq 2$.

This leads to the definition of the following estimation contrast

$$\mathcal{L}_n^C(\theta) = \sum_{i=0}^{n-1}\left(\frac{(X_{(i+1)\Delta_n} - X_{i\Delta_n} - \Delta_n b_\mu(Y_{i\Delta_n}, X_{i\Delta_n}))^2}{\Delta_n a_\sigma^2(Y_{i\Delta_n}, X_{i\Delta_n})} + \log(a_\sigma^2(Y_{i\Delta_n}, X_{i\Delta_n}))\right) \tag{10}$$

which is an extension of the classical Euler contrast for unidimensional SDE (see [10]) when drift b_μ and volatility a_σ depend on both Y and X. We define the minimum contrast estimator $\hat{\theta}_n^C$ for complete observations as

$$\hat{\theta}_n^C = \arg\min_{\theta\in\Theta} \mathcal{L}_n^C(\theta).$$

3.3. A contrast estimator for partial observations

Contrast (10) cannot be used in the second case of observations, as $(X_{i\Delta_n})$ is not observed. In the context of integrated diffusion, [8] proposes to approximate $X_{i\Delta_n}$ by increments of (Y_t). We study the behavior of the process of increments in Proposition 2. The basic idea which consists in replacing directly $X_{i\Delta_n}$ by $\overline{Y}_{i,n}$ in contrast (10) leads to a biased estimator (see [8], for the case where (X_t) satisfies an autonomous diffusion). This is due to the dependence between two successive terms of the rate process $\overline{Y}_{i,n}$ (Proposition 3). The estimation contrast for partial observation must be corrected to take into account this correlation.

Now, we present more precisely these ideas. First, we introduce the increment or rate process

$$\overline{Y}_{i,n} = \frac{Y_{(i+1)\Delta_n} - Y_{i\Delta_n}}{\Delta_n}. \tag{11}$$

Model (3) implies

$$\overline{Y}_{i,n} = \frac{1}{\Delta_n}\int_{i\Delta_n}^{(i+1)\Delta_n} X_s ds.$$

Thus, when Δ_n is small, $\overline{Y}_{i,n}$ is close to $X_{i\Delta_n}$. More precisely, we have:

Proposition 2. *Assume* (A1)–(A2). *We have*

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

where there exists a constant c *such that* $|\mathbb{E}(e_{i,n}|\mathcal{G}_i^n) \leq c\Delta_n(1 + |X_{i\Delta_n}| + |Y_{i\Delta_n}|)$ *and* $|\mathbb{E}(e_{i,n}^2|\mathcal{G}_i^n) \leq c\Delta_n^2(1 + |X_{i\Delta_n}|^4 + |Y_{i\Delta_n}|^4)$.

Furthermore, if k *is a real number* ≥ 1*, then for all* i, n*, we have*

$$\mathbb{E}\left(\left|\overline{Y}_{i,n} - X_{i\Delta_n}\right|^k |\mathcal{G}_i^n\right) \leq c\Delta_n^{k/2}(1 + |X_{i\Delta_n}|^k + |Y_{i\Delta_n}|^k).$$

The link between two successive terms of the non-Markovian rate process $\overline{Y}_{i,n}$ is studied in the following Proposition.

Proposition 3. *Assume* (A1)–(A2). *Then*

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

where $U_{i,n} = \xi_{i,n} + \xi'_{i+1,n}$ *with*

$$\xi_{i,n} = \frac{1}{\Delta_n^{3/2}}\int_{i\Delta_n}^{(i+1)\Delta_n} (s - i\Delta_n)dB_s \quad \text{for } i, n \geq 0$$

$$\xi'_{i+1,n} = \frac{1}{\Delta_n^{3/2}}\int_{(i+1)\Delta_n}^{(i+2)\Delta_n} ((i+1)\Delta_n - s)dB_s \quad \text{for } i \geq -1, n \geq 0.$$

If k *is a real number* ≥ 1*, then for all* i, n

$$\mathbb{E}\left(\left|\overline{Y}_{i+1,n} - \overline{Y}_{i,n}\right|^k |\mathcal{G}_i^n\right) \leq c\Delta_n^{k/2}(1 + |X_{(i+1)\Delta_n}|^k + |Y_{(i+1)\Delta_n}|^k).$$

Moreover there exist constants c *such that*

$$\mathbb{E}(\varepsilon_{i,n}^P|\mathcal{G}_i^n) \leq c\Delta_n^2(1 + |X_{(i+1)\Delta_n}|^3 + |Y_{(i+1)\Delta_n}|^3)$$
$$\mathbb{E}((\varepsilon_{i,n}^P)^2|\mathcal{G}_i^n) \leq c\Delta_n^2(1 + |X_{(i+1)\Delta_n}|^4 + |Y_{(i+1)\Delta_n}|^4)$$
$$\mathbb{E}((\varepsilon_{i,n}^P)^4|\mathcal{G}_i^n) \leq c\Delta_n^4(1 + |X_{(i+1)\Delta_n}|^8 + |Y_{(i+1)\Delta_n}|^8)$$
$$\mathbb{E}(\varepsilon_{i,n}^P U_{i,n}|\mathcal{G}_i^n) \leq c\Delta_n^{3/2}(1 + |X_{(i+1)\Delta_n}|^2 + |Y_{(i+1)\Delta_n}|^2).$$

Remark that Proposition 3 implies that for any function f of the two variables Y_t and X_t, $f(Y_{i\Delta_n}, X_{i\Delta_n})$ and $\overline{Y}_{i+1,n} - \overline{Y}_{i,n} - \Delta_n b(Y_{i\Delta_n}, \overline{Y}_{i,n})$ have a correlation of order $\Delta_n^{1/2}$. Moreover the variance of $U_{i,n}$ is $2/3\Delta_n$, while the variance of $\eta_{i,n}$ in Proposition 1 is 1. Gloter [8] proposes a correction of the contrast by weighting the first sum in (10) by a factor $3/2$. We extend this contrast to the case of drift and volatility functions depending on both processes (X_t) and (Y_t). Thus we consider the following contrast

$$\mathcal{L}_n^P(\theta) = \sum_{i=1}^{n-2} \left(\frac{3}{2} \frac{\left(\overline{Y}_{i+1,n} - \overline{Y}_{i,n} - \Delta_n b_\mu(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n})\right)^2}{\Delta_n a_\sigma^2(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n})} + \log(a_\sigma^2(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n})) \right). \tag{12}$$

Remark that as $(\overline{Y}_{i,n})$ is not Markovian, we introduce a shift in the index of the drift and the diffusion functions to avoid a correlation term of order $\Delta_n^{1/2}$ between $(\overline{Y}_{i+1,n} - \overline{Y}_{i,n})$ and functionals $f(Y_{i\Delta_n}, \overline{Y}_{i,n})$.

We define the minimum contrast estimator for partial observations $\hat{\theta}_n^P$ as

$$\hat{\theta}_n^P = \arg\min_{\theta \in \Theta} \mathcal{L}_n^P(\theta).$$

3.4. Main results

To simplify notations and proofs, we restrict to one-dimensional parameters μ and σ. This could easily be extended to multidimensional parameters (see Remark 5 of [8]). Simulations (Section 6) illustrate this extension.

In this paper, we prove the consistency and asymptotic normality of both estimators under the following identifiability assumption

$$\begin{aligned} &a_\sigma(y,x) = a_{\sigma_0}(y,x)\, d\nu_0(y,x) \quad \text{almost everywhere implies } \sigma = \sigma_0 \\ &b_\mu(y,x) = b_{\mu_0}(y,x)\, d\nu_0(y,x) \quad \text{almost everywhere implies } \mu = \mu_0. \end{aligned}$$

Classically, the consistency of the estimator $\hat{\theta}_n$ requires $\Delta_n \to 0$.

Theorem 1. *Under assumptions* (A1)–(A4)*, the estimators $\hat{\theta}_n^C$ and $\hat{\theta}_n^P$ are consistent:*

$$\hat{\theta}_n^C \xrightarrow[n\to\infty]{\mathbb{P}} \theta_0, \quad \text{and} \quad \hat{\theta}_n^P \xrightarrow[n\to\infty]{\mathbb{P}} \theta_0.$$

The asymptotic distribution requires the additional condition $n\Delta_n^2 \to 0$. The rate of convergence is different for $\hat{\mu}_n$ and $\widehat{\sigma_n^2}$. The drift term is estimated with rate $(n\Delta_n)^{1/2}$ and the diffusion term is estimated with rate $n^{1/2}$.

Theorem 2. *Set assumptions* (A1)–(A4)*, $n\Delta_n^2 \xrightarrow[n\to\infty]{} 0$. In the complete observations case, $\left(\sqrt{n\Delta_n}\left(\hat{\mu}_n^C - \mu_0\right), \sqrt{n}\left(\widehat{\sigma_n^2}^C - \sigma_0^2\right)\right)$ converges in distribution to*

$$\mathcal{N}\left(0, \left\{\nu_0\left(\frac{(\partial_\mu b)^2(\cdot,\cdot)}{a^2(\cdot,\cdot)}\right)\right\}^{-1}\right) \otimes \mathcal{N}\left(0, 2\left\{\nu_0\left(\frac{(\partial_{\sigma^2} a^2)^2(\cdot,\cdot)}{a^4(\cdot,\cdot)}\right)\right\}^{-1}\right)$$

and in the partial observations case, $\left(\sqrt{n\Delta_n}\left(\hat{\mu}_n^P - \mu_0\right), \sqrt{n}\left(\widehat{\sigma_n^2}^P - \sigma_0^2\right)\right)$ *converges in distribution to*

$$\mathcal{N}\left(0, \left\{\nu_0\left(\frac{(\partial_\mu b)^2(\cdot,\cdot)}{a^2(\cdot,\cdot)}\right)\right\}^{-1}\right) \otimes \mathcal{N}\left(0, \frac{9}{4}\left\{\nu_0\left(\frac{(\partial_{\sigma^2} a^2)^2(\cdot,\cdot)}{a^4(\cdot,\cdot)}\right)\right\}^{-1}\right).$$

Theorem 2 is an extension of several results. We first comment the complete observations case. When (X_t) is an autonomous diffusion, [10] proves that the asymptotic distribution is

$$\mathcal{N}\left(0, \left\{\nu_{X,0}\left(\frac{(\partial_\mu b)^2(\cdot)}{a^2(\cdot)}\right)\right\}^{-1}\right) \otimes \mathcal{N}\left(0, 2\left\{\nu_{X,0}\left(\frac{(\partial_{\sigma^2} a^2)^2(\cdot)}{a^4(\cdot)}\right)\right\}^{-1}\right)$$

where the limit distribution $\nu_{X,0}$ is the stationary distribution of the diffusion (X_t) itself. In that case, observations of (Y_t) are not used. When (X_t) is not autonomous and the diffusion is bidimensional, this result can be generalized if the volatility matrix Σ is non-degenerate. The asymptotic variance is then based on ν_0, the stationary distribution of the vector (Y_t, X_t). When the volatility matrix is degenerate as in model (3), the first assertion of Theorem 2 shows that reducing the contrast to the Euler approximation of the second coordinate yields to asymptotic normality of $\hat{\theta}_n$, the asymptotic variance involving the stationary distribution of the vector (Y_t, X_t). This is a major difference with respect to the case of an autonomous diffusion for (X_t).

For the partial observations case, when (X_t) is autonomous, [8] proves that replacing $X_{i\Delta_n}$ by $\overline{Y}_{i,n}$ underestimates the asymptotic variance, as a consequence of Proposition 3. As in the complete observation case, when (X_t) is autonomous, the asymptotic variance is based on the stationary distribution $\nu_{X,0}$. In model (3) where the diffusion is not autonomous, second assertion of Theorem 2 shows that the invariant measure of (Y_t, X_t) is required in the asymptotic variance.

The estimation of μ is asymptotically efficient since $\nu_0\left(\frac{(\partial_\mu b)^2(\cdot,\cdot)}{a^2(\cdot,\cdot)}\right)$ is the Fisher information of the continuous time model. This is not the case for σ^2 as its asymptotic variance is increased with a factor $9/16$ instead of $1/2$ for directly observed diffusion [10].

Proofs of Theorems 1 and 2 are given in Section 5. They are based on properties of functionals of $(Y_{i\Delta_n}, X_{i\Delta_n})$ and $(Y_{i\Delta_n}, \overline{Y}_{i,n})$, which are studied in Section 4.

4. Functionals of $(Y_{i\Delta_n}, X_{i\Delta_n})$ and $(Y_{i\Delta_n}, \overline{Y}_{i,n})$

Contrast properties rely on convergence results for functionals appearing in the contrast functions. These functionals are of different types: functional mean, variation and quadratic variation of $X_{i\Delta_n}$ and $\overline{Y}_{i,n}$. We consider for the complete observations case, for a measurable function f, the three functionals:

$$\overline{\nu}_n^C(f) = \frac{1}{n}\sum_{i=0}^{n-1} f(Y_{i\Delta_n}, X_{i\Delta_n}, \theta),$$

$$\overline{I}_n^C(f) = \frac{1}{n\Delta_n}\sum_{i=0}^{n-1} f(Y_{i\Delta_n}, X_{i\Delta_n}, \theta)(X_{(i+1)\Delta_n} - X_{i\Delta_n} - \Delta_n b(Y_{i\Delta_n}, X_{i\Delta_n}))$$

$$\overline{Q}_n^C(f) = \frac{1}{n\Delta_n}\sum_{i=0}^{n-1} f(Y_{i\Delta_n}, X_{i\Delta_n}, \theta)(X_{(i+1)\Delta_n} - X_{i\Delta_n})^2$$

and for the partial observations case, the three functionals

$$\overline{\nu}_n^P(f) = \frac{1}{n}\sum_{i=0}^{n-1} f(Y_{i\Delta_n}, \overline{Y}_{i,n}, \theta)$$

$$\overline{I}_n^P(f) = \frac{1}{n\Delta_n}\sum_{i=1}^{n-2} f(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n}, \theta)(\overline{Y}_{i+1,n} - \overline{Y}_i - \Delta_n b(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n}))$$

$$\overline{Q}_n^P(f) = \frac{1}{n\Delta_n}\sum_{i=1}^{n-2} f(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n}, \theta)\left(\overline{Y}_{i+1,n} - \overline{Y}_i\right)^2.$$

Note that in $\overline{I}_n^P(f)$ and $\overline{Q}_n^P(f)$, we introduce shifted processes $Y_{(i-1)\Delta_n}$ and $\overline{Y}_{i-1,n}$ in the function f as a consequence of the remark following Proposition 3. Consequently, the drift term $b(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n})$ in $\overline{I}_n^P(f)$ and in the contrast $\mathcal{L}_n^P$ are also shifted so that, when the square quantity in $\mathcal{L}_n^P$ is developed, functionals to be studied have the proper index. Asymptotic study of these functionals is difficult because it involves $(Y_{i\Delta_n}, \overline{Y}_{i,n})$ instead of the original Markovian process $(Y_{i\Delta_n}, X_{i\Delta_n})$.

We first study the uniform convergence of these functionals, then their convergence in distribution. In the following, we assume that f belongs to the class $\mathcal{F}$ introduced in Assumption (A1).

4.1. Uniform convergence

The first result concerns the empirical mean of the discretized process $(X_{i\Delta_n})_{i\geq 0}$ and the rate process $(\overline{Y}_{i,n})_{i\geq 1}$. The limits involve the stationary distribution ν_0 of the vector (Y_t, X_t). Proofs are given in the Appendix. They are essentially based on Propositions 2 and 3 and generalize the proofs of [8] to a non-autonomous diffusion (X_t).

Proposition 4. *Under assumptions* (A1)–(A4)*, we have uniformly in* θ

$$\overline{\nu}_n^C(f) \xrightarrow[n\to\infty]{\mathbb{P}} \nu_0(f), \qquad \overline{\nu}_n^P(f) \xrightarrow[n\to\infty]{\mathbb{P}} \nu_0(f).$$

We see that replacing $X_{i\Delta_n}$ by $\overline{Y}_{i,n}$ in the partial observations case does not change the limit. The next result concerns the functionals $\overline{I}_n^C$ and $\overline{I}_n^P$ which involve the variations of the processes $(X_{i\Delta_n})_{i\leq 0}$ and $(\overline{Y}_{i,n})_{i\leq 0}$, respectively.

Theorem 3. *Under assumptions* (A1)–(A4)*, we have uniformly in* θ

$$\overline{I}_n^C(f) \xrightarrow[n\to\infty]{\mathbb{P}} 0, \qquad \overline{I}_n^P(f) \xrightarrow[n\to\infty]{\mathbb{P}} 0. \tag{13}$$

The limit is the same for the complete and partial functionals. This is due to the introduction of the lag in the definition of $\overline{I}_n^P(f)$: $f(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n}, \theta)$ and $b(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n})$ instead of $f(Y_{i\Delta_n}, \overline{Y}_{i,n}, \theta)$ and $b(Y_{i\Delta_n}, \overline{Y}_{i,n})$. This enables us to avoid correlation terms of order $\Delta_n^{1/2}$. When no lag is introduced, the limit is not 0, see for instance [8].

The last result deals with the quadratic variations of $(X_{i\Delta_n})_{i\geq 0}$ and $(\overline{Y}_{i,n})_{i\geq 1}$.

Theorem 4. *Under assumptions* (A1)–(A4)*, we have uniformly in* θ

$$\overline{Q}_n^C(f(\cdot,\cdot,\theta)) \xrightarrow[n\to\infty]{\mathbb{P}} \nu_0(f(\cdot,\cdot,\theta)a^2(\cdot,\cdot))$$

$$\overline{Q}_n^P(f(\cdot,\cdot,\theta)) \xrightarrow[n\to\infty]{\mathbb{P}} \frac{2}{3}\nu_0(f(\cdot,\cdot,\theta)a^2(\cdot,\cdot)).$$

Theorem 4 is an extension of various results. It implies several comments which have already been partially addressed in Section 3.4. We first comment the complete observations case. When (X_t) is an autonomous diffusion, [10] proves that for a function $f:\mathbb{R}\times\Theta\to\mathbb{R}$

$$\frac{1}{n\Delta_n}\sum_{i=0}^{n-1} f(X_{i\Delta_n},\theta)\left(X_{(i+1)\Delta_n}-X_{i\Delta_n}\right)^2 \xrightarrow[n\to\infty]{\mathbb{P}} \nu_{X,0}(f(\cdot,\theta)a^2(\cdot))$$

where the limit distribution $\nu_{X,0}$ is the stationary distribution of the diffusion (X_t) itself. When the diffusion is two-dimensional and the volatility matrix Σ is non-degenerate, the limit is then $\nu_0(f(\cdot,\cdot,\theta)\Sigma\Sigma'(\cdot,\cdot))$ where ν_0 is the stationary distribution of the vector (Y_t,X_t). When the volatility matrix is degenerate as in model (3), the first assertion of Theorem 4 shows that the problem is reduced to the Euler approximation of the second equation of the system with the limit involving the stationary distribution of the vector (Y_t,X_t).

For the partial observations case, when (X_t) is autonomous, [8] proves that replacing $X_{i\Delta_n}$ by $\overline{Y}_{i,n}$ modifies the result by underestimating $\nu_{X,0}(f(\cdot,\theta)a^2(\cdot))$. In the case of model (3) where the diffusion is not autonomous, the second assertion of Theorem 4 shows that the invariant measure of (Y_t,X_t) is required.

4.2. Convergence in distribution of functionals of the process

In this section, we study some central limit theorems for the functionals $\overline{I}_n^C$, $\overline{I}_n^P$ and $\overline{Q}_n^C$, $\overline{Q}_n^P$. As $\overline{I}_n^C$ and $\overline{I}_n^P$ converge in probability to 0 (Theorem 3), they also satisfy a central limit theorems as follows:

Theorem 5. *Under assumptions* (A1)–(A4) *and* $n\Delta_n^2 \xrightarrow[n\to\infty]{} 0$*, we have*

$$\sqrt{n\Delta_n}\,\overline{I}_n^C(f) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0,\nu_0(f^2a^2))$$

$$\sqrt{n\Delta_n}\,\overline{I}_n^P(f) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0,\nu_0(f^2a^2)).$$

The condition $n\Delta_n^2 \xrightarrow[n\to\infty]{} 0$ is classical (see [4]). This condition imposes that the discretization step decreases to zero fast enough to ensure that the contribution of the error terms tends to 0 as $n\to\infty$. The lag introduced in the definition of $\overline{I}_n^P$ makes the result very similar for both complete and partial observations case.

We now present a central limit theorem for $\overline{Q}_n^P$ and $\overline{Q}_n^C$. Theorem 4 shows that $\overline{Q}_n^P$ underestimates $\nu_0(f(\cdot,\cdot,\theta)a^2(\cdot,\cdot))$. The correction factor $2/3$ is thus required in the associated central limit theorem:

Theorem 6. *Under assumptions* (A1)–(A4) *and* $n\Delta_n^2 \xrightarrow[n\to\infty]{} 0$, *we have*

$$\sqrt{n}\left(\overline{Q}_n^C(f) - \overline{\nu}_n^C(fa^2)\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, \nu_0(f^2a^4))$$
$$\sqrt{n}\left(\overline{Q}_n^P(f) - \frac{2}{3}\overline{\nu}_n^P(fa^2)\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, \nu_0(f^2a^4)).$$

In the partial observations case, when we replace $X_{i\Delta_n}$ by $\overline{Y}_{i,n}$, the asymptotic variance increases due to the factor $3/2$. This can also be compared to the results of [8] when the diffusion (X_t) is autonomous.

5. Proofs of main results

In this section, asymptotic properties of estimators $\hat{\theta}_n^C$ and $\hat{\theta}_n^P$ are proved.

5.1. *Proof of Theorem 1*

We follow the proof of [11]. We have to show that, uniformly in θ,

$$\frac{1}{n}\mathcal{L}_n^C(\theta) \xrightarrow[n\to\infty]{\mathbb{P}} \nu_0\left(\frac{a_{\sigma_0}^2(y,x)}{a_\sigma^2(y,x)} + \log a_\sigma^2(y,x)\right) \tag{14}$$

$$\frac{1}{n}\mathcal{L}_n^P(\theta) \xrightarrow[n\to\infty]{\mathbb{P}} \nu_0\left(\frac{a_{\sigma_0}^2(y,x)}{a_\sigma^2(y,x)} + \log a_\sigma^2(y,x)\right). \tag{15}$$

This ensures the convergence of $\hat{\sigma}_n^2$ to σ_0^2 for both cases. Then, if we prove that

$$\frac{1}{n\Delta_n}(\mathcal{L}_n^C(\mu,\sigma) - \mathcal{L}_n^C(\mu_0,\sigma)) \xrightarrow[n\to\infty]{\mathbb{P}} \nu_0\left(\frac{(b_\mu(y,x) - b_{\mu_0}(y,x))^2}{a_\sigma^2(y,x)}\right) \tag{16}$$

$$\frac{1}{n\Delta_n}(\mathcal{L}_n^P(\mu,\sigma) - \mathcal{L}_n^P(\mu_0,\sigma)) \xrightarrow[n\to\infty]{\mathbb{P}} \frac{3}{2}\nu_0\left(\frac{(b_\mu(y,x) - b_{\mu_0}(y,x))^2}{a_\sigma^2(y,x)}\right) \tag{17}$$

this ensures the convergence of $\hat{\mu}_n$ to μ_0 for both cases. We start by proving (14)–(15). In the complete observations case, we have

$$\begin{aligned}\frac{1}{n}\mathcal{L}_n^C(\theta) = {} & \overline{Q}_n^C(a_\sigma^{-2}(\cdot,\cdot)) + \overline{\nu}_n^C(\log a_\sigma^2(\cdot,\cdot)) - 2\Delta_n\overline{I}_n^C(a_\sigma^{-2}(\cdot,\cdot)b_\mu(\cdot,\cdot)) \\ & + \Delta_n\overline{\nu}_n^C(a_\sigma^{-2}(\cdot,\cdot)(b_\mu^2(\cdot,\cdot) - 2b_\mu(\cdot,\cdot)b_{\mu_0}(\cdot,\cdot))).\end{aligned}$$

Using Proposition 4, Theorems 3 and 4, we easily prove (14). In the partial observations case, we have

$$\begin{aligned}\frac{1}{n}\mathcal{L}_n^P(\theta) = {} & \frac{3}{2}\overline{Q}_n^P(a_\sigma^{-2}(\cdot,\cdot)) + \overline{\nu}_n^P(\log a_\sigma^2(\cdot,\cdot)) - 3\Delta_n\overline{I}_n^P(a_\sigma^{-2}(\cdot,\cdot)b_\mu(\cdot,\cdot)) \\ & + \frac{3}{2}\Delta_n\overline{\nu}_n^P(a_\sigma^{-2}(\cdot,\cdot)(b_\mu^2(\cdot,\cdot) - 2b_\mu(\cdot,\cdot)b_{\mu_0}(\cdot,\cdot))).\end{aligned}$$

Using Proposition 4, Theorems 3 and 4, we easily prove (15). For the proof of (16), we have

$$\frac{1}{n\Delta_n}\left(\mathcal{L}_n^C(\mu,\sigma)-\mathcal{L}_n^C(\mu_0,\sigma)\right)=2\overline{I}_n^C\left(\frac{b_{\mu_0}}{a_\sigma^2}(\cdot,\cdot)-\frac{b_\mu}{a_\sigma^2}(\cdot,\cdot)\right)+\overline{\nu}_n^C\left(\frac{\left(b_\mu(\cdot,\cdot)-b_{\mu_0}(\cdot,\cdot)\right)^2}{a_\sigma^2(\cdot,\cdot)}\right).$$

We conclude with Proposition 4 and Theorem 3. For the proof of (17), we write

$$\frac{1}{n\Delta_n}\left(\mathcal{L}_n^P(\mu,\sigma)-\mathcal{L}_n^P(\mu_0,\sigma)\right)=3\overline{I}_n^P\left(\frac{b_{\mu_0}}{a_\sigma^2}(\cdot,\cdot)-\frac{b_\mu}{a_\sigma^2}(\cdot,\cdot)\right)+\frac{3}{2}\overline{\nu}_n^P\left(\frac{\left(b_\mu(\cdot,\cdot)-b_{\mu_0}(\cdot,\cdot)\right)^2}{a_\sigma^2(\cdot,\cdot)}\right).$$

We conclude with Proposition 4 and Theorem 3.

5.2. *Proof of Theorem 2*

The scheme of the proof is the same for both complete and partial observations cases. Let $\hat{\theta}_n$ and $\mathcal{L}_n(\theta)$ denote the estimator and contrast either for complete or partial observations. A Taylor's formula around $\hat{\theta}_n$ yields: $\mathcal{D}_n=\int_0^1\mathcal{C}_n(\theta_0+u(\hat{\theta}_n-\theta_0)du\,\mathcal{E}_n)$ where

$$\mathcal{D}_n=\begin{pmatrix}-(\sqrt{n\Delta_n})^{-1}\partial_\mu\mathcal{L}_n(\theta_0)\\ -(\sqrt{n})^{-1}\partial_\sigma\mathcal{L}_n(\theta_0)\end{pmatrix},\qquad \mathcal{E}_n=\begin{pmatrix}\sqrt{n\Delta_n}\left(\hat{\mu}_n-\mu_0\right)\\ \sqrt{n}\left(\widehat{\sigma^2}_n-\sigma_0^2\right)\end{pmatrix},$$

$$\mathcal{C}_n(\theta)=\begin{pmatrix}\dfrac{1}{n\Delta_n}\,\partial^2_{\mu^2}\,\mathcal{L}_n(\theta) & \dfrac{1}{n\sqrt{\Delta_n}}\,\partial^2_{\sigma\mu}\,\mathcal{L}_n(\theta)\\ \dfrac{1}{n\sqrt{\Delta_n}}\,\partial^2_{\mu\sigma}\,\mathcal{L}_n(\theta) & \dfrac{1}{n}\,\partial^2_{\sigma^2}\,\mathcal{L}_n(\theta)\end{pmatrix}.$$

Let now detail the two cases. In the complete observations case, we have

$$\frac{1}{\sqrt{n\Delta_n}}\partial_\mu\mathcal{L}_n^C(\theta_0)=-2\sqrt{n\Delta_n}\overline{I}_n^C\left(\frac{\partial_\mu b_{\mu_0}(\cdot,\cdot)}{a_{\sigma_0}^2(\cdot,\cdot)}\right)$$

$$\frac{1}{\sqrt{n}}\partial_{\sigma^2}\mathcal{L}_n^C(\theta_0)=-\sqrt{n}\left(\overline{Q}_n^C\left(\frac{\partial_{\sigma^2}(a_{\sigma_0}^2(\cdot,\cdot))}{a_{\sigma_0}^4(\cdot,\cdot)}\right)-\overline{\nu}_n^C\left(\frac{\partial_{\sigma^2}(a_{\sigma_0}^2(\cdot,\cdot))}{a_{\sigma_0}^2(\cdot,\cdot)}\right)\right)+o_{\mathbb{P}}(1).$$

By Theorems 5 and 6, this yields

$$\mathcal{D}_n^C\xrightarrow[n\to\infty]{\mathcal{D}}\mathcal{N}\left(0,\begin{pmatrix}4\nu_0\left(\dfrac{(\partial_\mu b_{\mu_0})^2(\cdot,\cdot)}{a_{\sigma_0}^2(\cdot,\cdot)}\right) & 0\\ 0 & 2\nu_0\left(\dfrac{(\partial_{\sigma^2}a_{\sigma_0}^2)^2(\cdot,\cdot)}{a_{\sigma_0}^4(\cdot,\cdot)}\right)\end{pmatrix}\right).$$

The proof of the convergence in distribution of $\mathcal{E}_n$ follows from the consistency of $\hat{\theta}_n^C$ and if we prove the uniform (with respect to θ) convergence in probability of $\mathcal{C}_n^C(\theta)$. To prove the uniform convergence, we differentiate twice $\mathcal{L}_n^C$. Proposition 4 and Theorem 3 show that $\mathcal{C}_n^C(\theta)$ converges uniformly in θ in probability to $\mathcal{C}^C(\theta)$ where

$$\mathcal{C}^C(\theta) = \begin{pmatrix} \mathcal{C}_{11}^C(\theta) & 0 \\ 0 & \mathcal{C}_{22}^C(\theta) \end{pmatrix}$$

$$\mathcal{C}_{11}^C(\theta) = 2\nu_0 \left(\frac{(\partial_\mu b_\mu)^2(\cdot,\cdot)}{a_\sigma^2(\cdot,\cdot)} + \frac{\partial^2_{\mu^2} b_\mu}{a_\sigma^2}(\cdot,\cdot)(b_\mu(\cdot,\cdot) - b_{\mu_0}(\cdot,\cdot)) \right)$$

$$\begin{aligned} \mathcal{C}_{22}^C(\theta) \;&=\; \nu_0 \left((\partial_{\sigma^2} a_\sigma^2)^2(\cdot,\cdot) \left(\frac{2a_{\sigma_0}^2(\cdot,\cdot)}{a_\sigma^6(\cdot,\cdot)} - \frac{1}{a_{\sigma_0}^4(\cdot,\cdot)} \right) \right) \\ &\quad + \nu_0 \left(\partial^2_{\sigma^2} a_\sigma^2(\cdot,\cdot) \left(\frac{1}{a_\sigma^2(\cdot,\cdot)} - \frac{a_{\sigma_0}^2(\cdot,\cdot)}{a_\sigma^4(\cdot,\cdot)} \right) \right). \end{aligned}$$

Hence the result for the complete observations case. In the partial observations case, we have

$$\frac{1}{\sqrt{n\Delta_n}} \partial_\mu \mathcal{L}_n^P(\theta_0) = -3\sqrt{n\Delta_n}\,\overline{I}_n^P \left(\frac{\partial_\mu b_{\mu_0}(\cdot,\cdot)}{a_{\sigma_0}^2(\cdot,\cdot)} \right)$$

$$\frac{1}{\sqrt{n}} \partial_{\sigma^2} \mathcal{L}_n^P(\theta_0) = -\frac{3}{2}\sqrt{n} \left(\overline{Q}_n^P \left(\frac{\partial_{\sigma^2} a_{\sigma_0}^2(\cdot,\cdot)}{a_{\sigma_0}^4(\cdot,\cdot)} \right) - \frac{2}{3}\overline{\nu}_n^P \left(\frac{\partial_{\sigma^2} a_{\sigma_0}^2(\cdot,\cdot)}{a_{\sigma_0}^2(\cdot,\cdot)} \right) \right) + o_{\mathbb{P}}(1).$$

By Theorems 5 and 6, this yields

$$\mathcal{D}_n^P \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\left(0, \begin{pmatrix} 9\nu_0 \left(\frac{(\partial_\mu b_{\mu_0})^2(\cdot,\cdot)}{a_{\sigma_0}^2(\cdot,\cdot)} \right) & 0 \\ 0 & \frac{9}{4}\nu_0 \left(\frac{(\partial_{\sigma^2} a_{\sigma_0}^2)^2(\cdot,\cdot)}{a_{\sigma_0}^4(\cdot,\cdot)} \right) \end{pmatrix} \right).$$

Proof of the convergence in distribution of $\mathcal{E}_n$ follows from the consistency of $\hat{\theta}_n^P$ and the uniform (with respect to θ) convergence in probability of $\mathcal{C}_n^P(\theta)$. To prove the uniform convergence, we differentiate twice $\mathcal{L}_n^P$. Proposition 4 and Theorem 3 show that $\mathcal{C}_n^P(\theta)$ converges uniformly in θ in probability to $\mathcal{C}^P(\theta)$ where

$$\mathcal{C}^P(\theta) = \begin{pmatrix} \mathcal{C}_{11}^P(\theta) & 0 \\ 0 & \mathcal{C}_{22}^P(\theta) \end{pmatrix}$$

$$\mathcal{C}_{11}^P(\theta) = 3\nu_0 \left(\frac{(\partial_\mu b_\mu)^2(\cdot,\cdot)}{a_\sigma^2(\cdot,\cdot)} + \frac{\partial^2_{\mu^2} b_\mu}{a_\sigma^2}(\cdot,\cdot)(b_\mu(\cdot,\cdot) - b_{\mu_0}(\cdot,\cdot)) \right)$$

$$\begin{aligned} \mathcal{C}_{22}^P(\theta) \;&=\; \nu_0 \left((\partial_{\sigma^2} a_\sigma^2)^2(\cdot,\cdot) \left(\frac{2a_{\sigma_0}^2(\cdot,\cdot)}{a_\sigma^6(\cdot,\cdot)} - \frac{1}{a_{\sigma_0}^4(\cdot,\cdot)} \right) \right) \\ &\quad + \nu_0 \left(\partial^2_{\sigma^2} a_\sigma^2(\cdot,\cdot) \left(\frac{1}{a_\sigma^2(\cdot,\cdot)} - \frac{a_{\sigma_0}^2(\cdot,\cdot)}{a_\sigma^4(\cdot,\cdot)} \right) \right). \end{aligned}$$

Hence the result. □

6. Simulation study

We consider three models of simulation, which are those proposed by [15]. Their general form is given as the Langevin system (5) where F_D is some (possibly non-linear) force function parameterized by D. Model I corresponds to a simple linear stochastic growth with $\gamma = 0$, $F_D \equiv 0$. Model II corresponds to a linear oscillator subject to noise and damping with $\gamma > 0$, $F_D(y) \equiv \frac{D}{2}y^2$. Model III is a non-linear oscillator subject to noise and damping with $\gamma > 0$, $F_D(y) \equiv -\Sigma_{j=1}^{n} j^{-1} D_j (\cos y)^j$. Stability conditions for these models have been detailed in Section 2.

6.1. Model I: stochastic growth

We consider the following simple model

$$\begin{cases} dY_t = X_t dt \\ dX_t = \sigma_0 dB_t. \end{cases} \tag{18}$$

The process has one unknown parameter, σ_0, that describes the size of the fluctuations. Model (18) has a matricial form $dU_t = AU_t dt + \Gamma dB_t$ where $U_t = (Y_t, X_t)^t$ and

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \Gamma = \begin{pmatrix} 0 & 0 \\ 0 & \sigma_0 \end{pmatrix}.$$

This model has an explicit solution $U_t = e^{A(t-t_0)}U_0 + \int_{t_0}^{t} e^{A(t-s)} \Gamma dB_s$. Given the fact that $e^{At} = I + At$ for this simple model, the process (U_t) is Gaussian with expectation vector and covariance matrix

$$\mathbb{E}(U_t|U_0) = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} U_0, \qquad \mathrm{Var}(U_t|U_0) = \sigma_0^2 \Sigma_t = \sigma_0^2 \begin{pmatrix} t^3/3 & t^2/2 \\ t^2/2 & t \end{pmatrix}.$$

The covariance matrix Σ_t is invertible. Note that the process (U_t) has no stationary probability distribution. It is usual to consider the Lebesgue measure, which is not a probability measure, as its invariant measure. Although the theory developed in this paper has to be extended to the existence of an invariant measure which is not a probability measure, this is beyond the scope of this paper. Nevertheless, this example illustrates that estimators have good properties in that case. An exact discrete sampling scheme can be deduced from the exact distribution of (U_t)

$$\begin{pmatrix} Y_{(i+1)\Delta_n} \\ X_{(i+1)\Delta_n} \end{pmatrix} = \begin{pmatrix} Y_{i\Delta_n} + \Delta_n X_{i\Delta_n} \\ X_{i\Delta_n} \end{pmatrix} + \sigma_0 \Sigma_{\Delta_n}^{1/2} \begin{pmatrix} \varepsilon_{i\Delta_n}^{(1)} \\ \varepsilon_{i\Delta_n}^{(2)} \end{pmatrix}. \tag{19}$$

As the exact distribution is available and easily computable, the estimation of σ can be obtained from the exact likelihood when complete observations are available. The exact maximum likelihood estimator (MLE) is thus

$$\widehat{\sigma}^{\mathrm{MLE}} = \frac{1}{2n} \sum_{i=0}^{n-1} (U_{(i+1)\Delta_n} - e^{A\Delta_n} U_{i\Delta_n})' (\Sigma_{\Delta_n})^{-1} (U_{(i+1)\Delta_n} - e^{A\Delta_n} U_{i\Delta_n}).$$

The fact that the MLE is explicit is very specific to this simple model. It is also interesting to study the two contrast estimators, which are defined for more general models. The estimator for the complete observations case is equal to

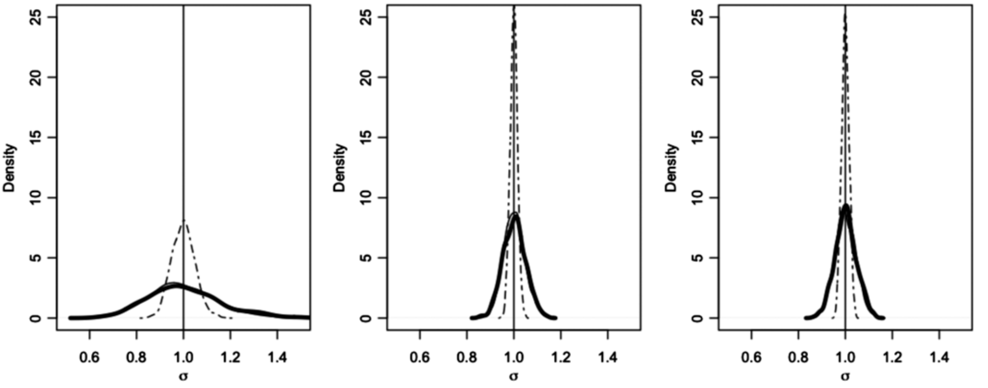

Fig. 1. Model I: stochastic growth. Estimator densities of parameter σ computed on 1000 simulated datasets for three designs $\Delta_n = 0.1, n = 100$ (a), $\Delta_n = 0.1, n = 1000$ (b) and $\Delta_n = 0.01, n = 1000$ (c). True value of σ is 1 (vertical line). Three estimators are compared: MLE (dotted line), complete observations contrast estimator $\widehat{\sigma}^C$ (plain line), partial observations contrast estimator $\widehat{\sigma}^P$ (bold line).

$$\widehat{\sigma}^C = \frac{1}{\Delta_n n} \sum_{i=0}^{n-1} (X_{(i+1)\Delta_n} - X_{i\Delta_n})^2.$$

When partial observations are available, the estimator is

$$\widehat{\sigma}^P = \frac{3}{2} \frac{1}{\Delta_n (n-2)} \sum_{i=1}^{n-2} (\overline{Y}_{i+1,n} - \overline{Y}_{i,n})^2.$$

The behavior of these three estimators are compared on simulated data. Three designs (Δ_n, n) of simulations are considered: $\Delta_n = 0.1, n = 100$; $\Delta_n = 0.1, n = 1000$ and $\Delta_n = 0.01, n = 1000$. A thousand of datasets are simulated for each design with the exact discrete scheme (19), the true parameter value $\sigma_0 = 1$ and $U_0 = (1, 1)'$. The three estimators are computed on each dataset. Kernel estimations of the density of these estimators are represented in Fig. 1. The three estimators are unbiased for the three designs. Their variances are small and decrease when n increases, whatever the value of Δ_n. The maximum likelihood estimator $\widehat{\sigma}^{\text{MLE}}$ has a smaller variance than the two contrast estimators $\widehat{\sigma}^C$ and $\widehat{\sigma}^P$, whatever the values of n and Δ_n. This is expected as the MLE is based on the exact distribution of the diffusion, while $\widehat{\sigma}^C$ and $\widehat{\sigma}^P$ are based on Euler approximation. The two contrast estimators behave very similarly. Empirical means and standard deviations of the three estimators for the three designs are presented in Table 1. Means and standard deviations obtained by Pokern et al. [15] on the same example are also reported. With complete observations, the MLE and the contrast estimator have similar estimate means and are unbiased. The standard deviations of $\widehat{\sigma}^C$ are three times larger than for $\widehat{\sigma}^{\text{MLE}}$. With partial observations, the contrast estimator $\widehat{\sigma}^P$ has similar mean than the one of [15], but twice greater standard deviations.

6.2. *Model II: harmonic oscillator*

We consider a harmonic oscillator that is driven by a white noise forcing:

$$\begin{cases} dY_t = X_t dt \\ dX_t = (-D_0 Y_t - \gamma_0 X_t)dt + \sigma_0 dB_t \end{cases} \tag{20}$$

Table 1
Model I: stochastic growth. True value is $\sigma = 1$. Mean and standard error of estimators of parameter σ computed on 1000 simulated datasets for three designs $\Delta_n = 0.1, n = 100$, $\Delta_n = 0.1, n = 1000$ and $\Delta_n = 0.01, n = 1000$. Four estimators are compared: MLE with complete observations, complete observations contrast estimator $\widehat{\sigma}^C$, partial observations contrast estimator $\widehat{\sigma}^P$ and Gibbs estimates obtained by Pokern et al. [15] with partial observations.

Observations	Estimator	Design		
		$\Delta_n = 0.1$ $n = 100$	$\Delta_n = 0.1$ $n = 1000$	$\Delta_n = 0.01$ $n = 1000$
Complete	$\widehat{\sigma}^{\mathrm{MLE}}$	0.999 (0.050)	1.000 (0.015)	1.000 (0.016)
Complete	$\widehat{\sigma}^C$	0.998 (0.142)	1.001 (0.044)	1.000 (0.044)
Partial	$\widehat{\sigma}^P$	1.005 (0.158)	1.002 (0.048)	1.001 (0.046)
Partial	Pokern et al.	0.993 (0.077)	0.999 (0.024)	1.000 (0.024)

with $\gamma_0 > 0$ and $D_0 > 0$. The process has three unknown parameters $(D_0, \gamma_0, \sigma_0)$. Model (20) has a matricial form $dU_t = AU_t dt + \Gamma dB_t$ where $U_t = (Y_t, X_t)^t$,

$$A = \begin{pmatrix} 0 & 1 \\ -D_0 & -\gamma_0 \end{pmatrix}, \qquad \Gamma = \begin{pmatrix} 0 & 0 \\ 0 & \sigma_0 \end{pmatrix}.$$

The stationary distribution of (U_t) is Gaussian with zero mean and an explicit variance matrix [5]

$$\mathrm{Var}(U_t) = \Sigma_U = \frac{1}{-2tr(A)\det(A)} \det(A)\Gamma\Gamma' + (A - tr(A)I_2)\Gamma\Gamma'(A - tr(A)I_2)'$$

where $tr(A)$ and $\det(A)$ are the trace and the determinant of A.

The estimator for the complete observations case is defined as

$$\widehat{\theta}^C = \arg\min_\theta \left[\sum_{i=0}^{n-1} \frac{\left(X_{(i+1)\Delta_n} - X_{i\Delta_n} + \Delta_n(DY_{i\Delta_n} + \gamma X_{i\Delta_n})\right)^2}{\Delta_n \sigma^2} + n \log \sigma^2 \right].$$

When only partial observations $(Y_{i\Delta_n})$ are available, the contrast is

$$\widehat{\theta}^P = \arg\min_\theta \left[\frac{3}{2} \sum_{i=1}^{n-2} \frac{\left(\overline{Y}_{i+1,n} - \overline{Y}_{i,n} + \Delta_n(DY_{(i-1)\Delta_n} + \gamma \overline{Y}_{i-1,n})\right)^2}{\Delta_n \sigma^2} \right.$$
$$\left. + \ (n-2) \log \sigma^2 \right].$$

The behavior of these two estimators is compared on simulated data. Three designs (Δ_n, n) of simulations are considered: $\Delta_n = 0.1, n = 1000$; $\Delta_n = 0.1, n = 100$ and $\Delta_n = 0.01, n = 1000$. A thousand of datasets are simulated for each design with the exact stationary distribution, the true parameter values $D_0 = 4$, $\gamma_0 = 0.5$ and $\sigma_0 = 1$ and $U_0 = (1, 0)'$. The two estimators $\widehat{\theta}^C$ and $\widehat{\theta}^P$ are computed on each dataset. Empirical mean and standard deviations of the estimators are reported on Table 2 (simultaneous estimation of the three parameters). Results obtained by Pokern et al. [15] when estimating only $\widehat{\sigma}^P$ are reported in Table 2. In their paper, the authors do not detail their results for drift parameters when using the Gibbs loop, which corresponds to a simultaneous estimation. So we only report results for σ. Parameter σ is estimated with very small bias whatever the design and the kind of observations. The bias of σ is slightly less than the one of [15]. Its standard deviation decreases with n. The drift parameters D and γ are estimated with bias for the two first designs. The bias decrease when $n = 1000$, $\Delta_n = 0.01$. This

Table 2
Model II: harmonic growth, estimation of the three parameters D, γ, σ. Mean and standard error of parameter estimators D, γ and σ computed on 1000 simulated datasets for three designs $\Delta_n = 0.1, n = 100$ (a), $\Delta_n = 0.1, n = 1000$ (b) and $\Delta_n = 0.01, n = 1000$ (c). Three estimators are compared: complete observations contrast estimator $\widehat{\theta}^C$, partial observations contrast estimator $\widehat{\theta}^P$ and Gibbs estimator obtained by Pokern et al. [15] with partial observations.

Estimator	True value	Design		
		$\Delta_n = 0.1$ $n = 100$	$\Delta_n = 0.1$ $n = 1000$	$\Delta_n = 0.01$ $n = 1000$
$\widehat{\sigma}^C$	1	0.980 (0.069)	0.974 (0.021)	0.996 (0.021)
$\widehat{\sigma}^P$		0.946 (0.074)	0.956 (0.021)	0.994 (0.023)
Pokern et al.		1.154 (0.074)	1.114 (0.025)	1.016 (0.013)
$\widehat{D}^C$	4	3.567 (0.489)	3.488 (0.187)	4.034 (0.642)
$\widehat{D}^P$		3.588 (0.494)	3.501 (0.188)	4.032 (0.644)
$\widehat{\gamma}^C$	0.5	1.022 (0.098)	1.086 (0.271)	0.678 (0.326)
$\widehat{\gamma}^P$		1.285 (0.275)	1.215 (0.096)	0.699 (0.330)

is corroborated by the theoretical results, as the asymptotic conditions are not the same for drift and volatility parameter estimation. The bias is very small for D but still remains for γ when $n = 1000$, $\Delta_n = 0.01$. The bias for drift parameters obtained with partial observations are larger than with complete observations. For example, with $n = 100$, $\Delta_n = 0.1$, the mean estimated value for parameter D is 3.588 with partial observations and 3.567 for complete observations, to be compared to the true value 4. When n increases and Δ_n decreases, this difference decreases and the bias is small.

6.3. Model III: trigonometric oscillator

We consider the dynamics of a particle moving in a trigonometric potential (see [15]). The model is

$$\begin{cases} dY_t = X_t dt \\ dX_t = \left(-\gamma_0 X_t - \sum_{j=1}^{c} D_{0,j} \sin(Y_t) \cos^{j-1}(Y_t)\right) dt + \sigma_0 dB_t \end{cases} \tag{21}$$

with parameters $\theta = (\gamma_0, D_{0j}, j = 1, \ldots, c, \sigma_0)$. This system is non-linear. No explicit closed form expression for the solution is known.
The estimator for the complete observations case is defined as

$$\widehat{\theta}^C = \arg\min_{\theta} \left[n \log \sigma^2 + \sum_{i=0}^{n-1} \frac{\left(X_{(i+1)\Delta_n} - X_{i\Delta_n} + \Delta_n \left(\gamma X_{i\Delta_n} + \sum_{j=1}^{c} D_j \sin(Y_{i\Delta_n}) \cos^{j-1}(Y_{i\Delta_n})\right)\right)^2}{\Delta_n \sigma^2} \right].$$

When only partial observations $(Y_{i\Delta_n})$ are available, the estimator is

$$\widehat{\theta}^P = \arg\min_{\theta} \left[(n-2) \log \sigma^2 + \frac{3}{2} \sum_{i=1}^{n-2} \frac{\left(\overline{Y}_{i+1,n} - \overline{Y}_{i,n} + \Delta_n \left(\gamma \overline{Y}_{i-1,n} + \sum_{j=1}^{c} D_j \sin(Y_{(i-1)\Delta_n}) \cos^{j-1}(Y_{(i-1)\Delta_n})\right)\right)^2}{\Delta_n \sigma^2} \right]$$

Table 3
Model III: trigonometric growth, estimation of the five parameters $D_1, D_2, D_3, \gamma, \sigma$. Mean and standard error of parameter estimators D, γ and σ computed on 1000 simulated datasets for four designs $\Delta_n = 0.1, n = 100$, $\Delta_n = 0.1, n = 1000$, $\Delta_n = 0.01, n = 1000$ and $\Delta_n = 0.01, n = 10,000$. Two estimators are compared: complete observations contrast estimator $\widehat{\theta}^C$ and partial observations contrast estimator $\widehat{\theta}^P$.

Estimator	True value	Design			
		$\Delta_n = 0.1$ $n = 100$	$\Delta_n = 0.1$ $n = 1000$	$\Delta_n = 0.01$ $n = 1000$	$\Delta_n = 0.01$ $n = 10,000$
$\widehat{\sigma}^C$	0.7	0.886 (0.110)	0.861 (0.032)	0.713 (0.019)	0.714 (0.006)
$\widehat{\sigma}^P$		1.012 (0.118)	1.021 (0.034)	0.873 (0.024)	0.784 (0.008)
$\widehat{D_1}^C$	1	0.987 (0.414)	1.003 (0.125)	1.043 (0.381)	1.010 (0.111)
$\widehat{D_1}^P$		1.002 (0.378)	1.002 (0.116)	1.036 (0.378)	1.005 (0.110)
$\widehat{D_2}^C$	−8	−8.221 (1.451)	−8.020 (0.367)	−8.082 (1.878)	−8.042 (0.498)
$\widehat{D_2}^P$		−7.340 (1.382)	−7.251 (0.339)	−8.019 (1.859)	−7.998 (0.495)
$\widehat{D_3}^C$	8	8.271 (2.424)	8.001 (0.597)	7.722 (3.589)	8.010 (0.764)
$\widehat{D_3}^P$		7.068 (2.235)	7.007 (0.565)	7.641 (3.559)	7.964 (0.758)
$\widehat{\gamma}^C$	0.5	0.638 (0.290)	0.524 (0.074)	0.671 (0.384)	0.522 (0.099)
$\widehat{\gamma}^P$		0.889 (0.304)	0.763 (0.074)	0.701 (0.384)	0.548 (0.100)

The behavior of these estimators is compared on simulated data. Four designs (Δ_n, n) of simulations are considered: $\Delta_n = 0.1, n = 100$; $\Delta_n = 0.1, n = 1000$; $\Delta_n = 0.01, n = 1000$ and $\Delta_n = 0.01, n = 10,000$. A thousand of datasets are simulated for each design with the exact stationary distribution and the true parameter values proposed by Pokern et al. [15] $D_{01} = 1$, $D_{02} = -8$, $D_{03} = 8$, $\gamma_0 = 0.5$ and $\sigma_0 = 0.7$ and $U_0 = (1, 1)'$. The two estimators $\widehat{\theta}^C$ and $\widehat{\theta}^P$ are computed on each dataset. Simultaneous estimation of the five parameters is performed. Empirical mean and standard deviations of the estimators are reported on Table 3. Pokern et al. [15]'s results are presented as figures and are not reported here. Bias and standard deviations of drift and volatility parameters decrease when n increases and Δ_n decreases. For example, for σ with complete observations, the mean estimated value is 0.886 when $n = 100$, $\Delta_n = 0.1$ and 0.714 when $n = 1000$, $\Delta_n = 0.01$, to be compared to the true value 0.7. For γ with complete observations, the mean estimated values is 0.638 when $n = 100$, $\Delta_n = 0.1$ and 0.522 when $n = 1000$, $\Delta_n = 0.01$, to be compared to the true value 0.5. Estimators obtained from partial observations have greater bias than those obtained from complete observations. For example, when $n = 100$, $\Delta_n = 0.1$, for σ with complete observations, the mean estimated value for σ is 0.886 with complete observations and 1.012 with partial observations, to be compared to the true value 0.7.

7. Discussion

We consider two cases of observations (partial and complete) of a hypoelliptic two-dimensional diffusion, with non-autonomous equations. The contrast estimators are based on Euler approximations of the second coordinate. We prove their consistency and give their asymptotic distribution. The case of complete observations leads to efficient estimator. On the contrary, in the case of partial observations, our estimator is not efficient. This extends the results of [8] to non-autonomous diffusion.

We compare our estimators to [15]'s estimator. Pokern et al. [15] limit their study to linear drift and constant diffusion coefficient. Their estimator is based on a hybrid Gibbs sampler in a

Bayesian framework. Their algorithm may be time consuming. Our estimator has the advantage to be simple to compute. For example, on the three examples considered in the simulation study, which are the same than those handled by Pokern et al. [15], our estimators are explicit and thus computed in less than one second.

Only second-order hypoelliptic systems have been considered in this paper. The estimation method proposed by Pokern et al. [15] works for larger order. The extension of our approach to these higher order hypoelliptic systems would require higher order approximation schemes, as Runge–Kutta schemes.

Although Model (1) involves a function g in the first coordinate, we reduce to the case $dY_t = X_t$ for the definition of the contrast functions as explained in Section 2. Our estimation procedure could be used to estimate parameters of function g. Numerical study of such models would be explored in future works. This could have great usefulness to consider more complex models and real data.

Acknowledgment

This work was supported by Agence Nationale de la Recherche through the project MANDy ANR-09-BLAN-0008-01.

Appendix. Proofs

Proposition A of [7] can be extended to drift and volatility depending both on y and x:

Proposition 5. *Let $f \in \mathcal{C}^1$. If $\exists c, \forall y, x$ such that $|f'_y(y,x)| + |f'_x(y,x)| \leq c(1+|y|+|x|)$ then, for all integer $k \geq 1$, we have*

$$\mathbb{E}\left(\sup_{t\in[i\Delta_n,(i+1)\Delta_n[} \left|f(Y_t,X_t) - f(Y_{i\Delta_n},X_{i\Delta_n})\right|^k |\mathcal{G}_i^n\right) \leq c\Delta_n^{k/2}(1+|Y_{i\Delta_n}|^k + |X_{i\Delta_n}|^k).$$

Proof. With start with $f(y,x) = x$. Let $\delta_{i,n} = \sup_{t\in[i\Delta_n,(i+1)\Delta_n[} |X_t - X_{i\Delta_n}|$. Using the Burkholder inequality, we get

$$\mathbb{E}(\delta_{i,n}^k|\mathcal{G}_i^n) \leq c\,\mathbb{E}\left[\left(\int_{i\Delta_n}^{(i+1)\Delta_n} |b(Y_t,X_t)|dt\right)^k \middle|\mathcal{G}_i^n\right] + c\,\mathbb{E}\left[\left(\int_{i\Delta_n}^{(i+1)\Delta_n} |a^2(Y_t,X_t)|dt\right)^{k/2} \middle|\mathcal{G}_i^n\right].$$

Using Assumptions (A2), we get

$$\begin{aligned}\mathbb{E}(\delta_{i,n}^k|\mathcal{G}_i^n) &\leq c\,\Delta_n^k\mathbb{E}\left[\sup_{t\in[i\Delta_n,(i+1)\Delta_n[} |b^k(Y_t,X_t)\|\mathcal{G}_i^n\right] \\ &\quad + c\,\Delta_n^{k/2}\mathbb{E}\left[\sup_{t\in[i\Delta_n,(i+1)\Delta_n[} |a^k(Y_t,X_t)\|\mathcal{G}_i^n\right] \\ &\quad + c\,\Delta_n^{k/2}(1+|Y_{i\Delta_n}|^k + |X_{i\Delta_n}|^k).\end{aligned}$$

Now for a general f, we study $f(Y_t, X_t) - f(Y_{i\Delta_n}, X_{i\Delta_n}) = f(Y_t, X_t) - f(Y_{i\Delta_n}, X_t) + f(Y_{i\Delta_n}, X_t) - f(Y_{i\Delta_n}, X_{i\Delta_n})$. We have

$$\begin{aligned}
&\mathbb{E}\left(\sup_{t\in[i\Delta_n,(i+1)\Delta_n[} \left|f(Y_t, X_t) - f(Y_{i\Delta_n}, X_{i\Delta_n})\right|^k |\mathcal{G}_i^n\right) \\
&\quad \le 2^{k-1}\mathbb{E}\left(\sup_{t\in[i\Delta_n,(i+1)\Delta_n[} \left|f(Y_t, X_t) - f(Y_{i\Delta_n}, X_t)\right|^k |\mathcal{G}_i^n\right) \\
&\qquad + 2^{k-1}\mathbb{E}\left(\sup_{t\in[i\Delta_n,(i+1)\Delta_n[} \left|f(Y_{i\Delta_n}, X_t) - f(Y_{i\Delta_n}, X_{i\Delta_n})\right|^k |\mathcal{G}_i^n\right).
\end{aligned}$$

We first study $f(Y_t, X_t) - f(Y_{i\Delta_n}, X_t)$. Burkholder inequality yields

$$\mathbb{E}\left(\sup_{t\in[i\Delta_n,(i+1)\Delta_n[} \left|f(Y_t, X_t) - f(Y_{i\Delta_n}, X_t)\right|^k |\mathcal{G}_i^n\right) \le c\, \Delta_n^{k/2}(1 + |Y_{i\Delta_n}|^k + |X_{i\Delta_n}|^k).$$

We then study $f(Y_{i\Delta_n}, X_t) - f(Y_{i\Delta_n}, X_{i\Delta_n})$. Burkholder inequality yields the result. $\square$

Proof of Proposition 1. We have

$$X_{(i+1)\Delta_n} - X_{i\Delta_n} - \Delta_n b(Y_{i\Delta_n}, X_{i\Delta_n}) = a(Y_{i\Delta_n}, X_{i\Delta_n})\eta_{i,n} + \alpha_{in} + \beta_{in}$$

where $\eta_{i,n} = \int_{i\Delta_n}^{(i+1)\Delta_n} dB_s$, $\alpha_{i,n} = \int_{i\Delta_n}^{(i+1)\Delta_n}(a(Y_s, X_s) - a(Y_{i\Delta_n}, X_{i\Delta_n}))dB_s$ and $\beta_{i,n} = \int_{i\Delta_n}^{(i+1)\Delta_n}(b(Y_s, X_s) - b(Y_{i\Delta_n}, X_{i\Delta_n}))ds$. Properties of $\eta_{i,n}$ are directly deduced from properties of the Brownian motion. Let $\varepsilon_{i,n}^C = \alpha_{i,n} + \beta_{i,n}$. Assumptions (A1)–(A2) lead to $|\mathbb{E}(\beta_{i,n}|\mathcal{G}_i^n)| \le c\Delta_n^{3/2}(1+|Y_{i\Delta_n}|+|X_{i\Delta_n}|)$. Proposition 5 provides $\mathbb{E}(|\beta_{i,n}|^k|\mathcal{G}_i^n) \le c\Delta_n^{k/2}(1+|Y_{i\Delta_n}|^k+|X_{i\Delta_n}|^k)$ for $k \ge 2$. Burkholder inequality gives $\mathbb{E}(|\alpha_{i,n}|^k |\mathcal{G}_i^n) \le c\Delta_n^k(1 + |Y_{i\Delta_n}|^k + |X_{i\Delta_n}|^k)$ for $k \ge 2$. $\square$

Proof of Proposition 2. We have $\overline{Y}_{i,n} - X_{i\Delta_n} = \frac{1}{\Delta_n}\int_{i\Delta_n}^{(i+1)\Delta_n}(X_v - X_{i\Delta_n})dv$ and $X_v - X_{i\Delta_n} = \int_{i\Delta_n}^{v} b(Y_s, X_s)ds + \int_{i\Delta_n}^{v} a(Y_s, X_s)dB_s$. By the Fubini theorem, we get $\overline{Y}_{i,n} - X_{i\Delta_n} = \Delta_n^{1/2}a(Y_{i\Delta_n}, X_{i\Delta_n})\xi'_{i,n} + e_{i,n}$ where $e_{i,n} = \alpha_{i,n} + \beta_{i,n}$ and

$$\begin{aligned}
\alpha_{i,n} &= \frac{1}{\Delta_n}\int_{i\Delta_n}^{(i+1)\Delta_n} \left(a(Y_v, X_v) - a(Y_{i\Delta_n}, X_{i\Delta_n})\right)((i+1)\Delta_n - v)dB_v \\
\beta_{i,n} &= \frac{1}{\Delta_n}\int_{i\Delta_n}^{(i+1)\Delta_n} \int_{i\Delta_n}^{v} b(Y_s, X_s)dsdv.
\end{aligned}$$

By Assumption (A1), we get $|\beta_{i,n}| \le c\Delta_n(1+\sup_{s\in[i\Delta_n,(i+1)\Delta_n[}(|Y_s|+|X_s|))$. As $\mathbb{E}(\alpha_{i,n}|\mathcal{G}_i^n) = 0$, we get $|\mathbb{E}(e_{i,n}|\mathcal{G}_i^n)| \le c\Delta_n(1 + |X_{i\Delta_n}| + |Y_{i\Delta_n}|)$. By Assumption (A2), for all $k \ge 0$, we get $\mathbb{E}\left(|\beta_{i,n}|^k|\mathcal{G}_i^n\right) \le c\Delta_n^k(1+|Y_{i\Delta_n}|^k+|X_{i\Delta_n}|^k)$. For $k \ge 2$, applying the Burkholder–Davis–Gundy and the Jensen inequalities yields:

$$\mathbb{E}\left(\left|\alpha_{i,n}^k\right| |\mathcal{G}_i^n\right) \le c\int_{i\Delta_n}^{(i+1)\Delta_n} \mathbb{E}\left(|a(Y_s, X_s) - a(Y_{i\Delta_n}, X_{i\Delta_n})|^k|\mathcal{G}_i^n\right) ds.$$

By Proposition 5 and Assumption (A1), we get $\mathbb{E}\left(\left|\alpha_{i,n}^k\right|\right) \leq c\Delta_n^{k/2+1}(1+|Y_{i\Delta_n}|^k+|X_{i\Delta_n}|^k)$. Finally, we get $|\mathbb{E}(e_{i,n}^2|\mathcal{G}_i^n)| \leq c\Delta_n^2(1+|X_{i\Delta_n}|^2+|Y_{i\Delta_n}|^2)$. Using Proposition 5, we have

$$\mathbb{E}\left(\sup_{s\in[i\Delta_n,(i+1)\Delta_n[}|X_s-X_{i\Delta_n}|^k|\mathcal{G}_i^n\right) \leq \Delta_n^{k/2}(1+|Y_{i\Delta_n}|^k+|X_{i\Delta_n}|^k)$$

thus we directly deduce

$$\begin{aligned}\mathbb{E}\left(\left|\overline{Y}_{i,n}-X_{i\Delta_n}\right|^k|\mathcal{G}_i^n\right) &= \mathbb{E}\left(\left|\frac{1}{\Delta_n}\int_{i\Delta_n}^{(i+1)\Delta_n}(X_s-X_{i\Delta_n})ds\right|^k\middle|\mathcal{G}_i^n\right)\\ &\leq \Delta_n^{k/2}(1+|Y_{i\Delta_n}|^k+|X_{i\Delta_n}|^k). \quad \square\end{aligned}$$

Proof of Proposition 3. We have

$$\begin{aligned}\overline{Y}_{i+1,n}-\overline{Y}_{i,n} &= \frac{1}{\Delta_n}\underbrace{\int_{i\Delta_n}^{(i+1)\Delta_n}\int_s^{s+\Delta_n}a(Y_v,X_v)dB_vds}_{A_i}\\ &\quad+\frac{1}{\Delta_n}\underbrace{\int_{i\Delta_n}^{(i+1)\Delta_n}\int_s^{s+\Delta_n}b(Y_v,X_v)dvds}_{B_i}.\end{aligned}$$

By Fubini theorem, we have

$$\begin{aligned}A_i &= \int_{i\Delta_n}^{(i+1)\Delta_n}a(Y_v,X_v)(v-i\Delta_n)dB_v+\int_{(i+1)\Delta_n}^{(i+2)\Delta_n}a(Y_v,X_v)((i+2)\Delta_n-v)dB_v\\ B_i &= \int_{i\Delta_n}^{(i+1)\Delta_n}b(Y_v,X_v)(v-i\Delta_n)dv+\int_{(i+1)\Delta_n}^{(i+2)\Delta_n}b(Y_v,X_v)((i+2)\Delta_n-v)dv.\end{aligned}$$

We can rewrite A_i as $A_i = a(Y_{i\Delta_n},X_{i\Delta_n})\Delta_n^{3/2}(\xi_{i,n}+\xi'_{i+1,n})+a_{i,n}+a'_{i+1,n}$ where $a_{i,n}=\int_{i\Delta_n}^{(i+1)\Delta_n}\left(a(Y_v,X_v)-a(Y_{i\Delta_n},X_{i\Delta_n})\right)(v-i\Delta_n)dB_v$ and $a'_{i+1,n}=\int_{(i+1)\Delta_n}^{(i+2)\Delta_n}\big(a(Y_v,X_v) - a(Y_{i\Delta_n},X_{i\Delta_n})\big)((i+2)\Delta_n-v)dB_v$. Similarly,

$$B_i = b(Y_{i\Delta_n},X_{i\Delta_n})\Delta_n^2+b_{i,n}+b'_{i+1,n}$$

where $b_{i,n} = \int_{i\Delta_n}^{(i+1)\Delta_n}(b(Y_v,X_v)-b(Y_{i\Delta_n},\overline{Y}_{i,n}))(v-i\Delta_n)dB_v$ and $b'_{i+1,n} = \int_{(i+1)\Delta_n}^{(i+2)\Delta_n}(b(Y_v,X_v)-b(Y_{i\Delta_n},\overline{Y}_{i,n}))((i+2)\Delta_n-v)dB_v$. Therefore, this yields

$$\overline{Y}_{i+1,n}-\overline{Y}_{i,n}-\Delta_nb(Y_{i\Delta_n},\overline{Y}_{i,n}) = a(Y_{i\Delta_n},X_{i\Delta_n})\Delta_n^{1/2}(\xi_{i,n}+\xi'_{i+1,n})+\varepsilon_{i,n}^P$$

with $\varepsilon_{i,n}^P=\frac{a_{i,n}}{\Delta_n}+\frac{a'_{i+1,n}}{\Delta_n}+\frac{b_{i,n}}{\Delta_n}+\frac{b'_{i+1,n}}{\Delta_n}$.

∘ Let us prove $|\mathbb{E}(\varepsilon^P_{i,n}|\mathcal{G}^n_i)| \le c\Delta_n^2(1+|X_{(i+1)\Delta_n}|^3+|Y_{(i+1)\Delta_n}|^3)$. We have $\mathbb{E}(a_{i,n}|\mathcal{G}^n_i) = \mathbb{E}(a'_{i+1,n}|\mathcal{G}^n_i)=0$ and

$$\mathbb{E}\left(\frac{b_{i,n}}{\Delta_n}|\mathcal{G}^n_i\right) = \frac{1}{\Delta_n}\int_{i\Delta_n}^{(i+1)\Delta_n}(v-i\Delta_n)\mathbb{E}(b(Y_v,X_v)-b(Y_{i\Delta_n},X_{i\Delta_n})|\mathcal{G}^n_i)dv + \frac{1}{\Delta_n}\int_{i\Delta_n}^{(i+1)\Delta_n}(v-i\Delta_n)\mathbb{E}(b(Y_{i\Delta_n},X_{i\Delta_n})-b(Y_{i\Delta_n},\overline{Y}_{i,n})|\mathcal{G}^n_i)dv.$$

By Itô's formula, Assumptions (A1)–(A2) and Proposition 5, we get

$$\sup_{v\in[i\Delta_n,(i+1)\Delta_n[}\left|\mathbb{E}\left(b(Y_v,X_v)-b(Y_{i\Delta_n},X_{i\Delta_n})|\mathcal{G}^n_i\right)\right| \le \Delta_n c(1+|Y_{i\Delta_n}|^2+|X_{i\Delta_n}|^2).$$

By Taylor's formula of order two, there exists $Z\in(\overline{Y}_{i,n},X_{i\Delta_n})$ such that

$$b(Y_{i\Delta_n},\overline{Y}_{i,n})-b(Y_{i\Delta_n},X_{i\Delta_n}) = b'_x(Y_{i\Delta_n},X_{i\Delta_n})(\overline{Y}_{i,n}-X_{i\Delta_n}) + \frac{1}{2}b''_{x^2}(Y_{i\Delta_n},Z)(\overline{Y}_{i,n}-X_{i\Delta_n})^2.$$

Using the Cauchy–Schwartz inequality, we get

$$\left|\mathbb{E}\left(b(Y_{i\Delta_n},\overline{Y}_{i,n})-b(Y_{i\Delta_n},X_{i\Delta_n})|\mathcal{G}^n_i\right)\right| \le c\Delta_n(1+|Y_{i\Delta_n}|^2+|X_{i\Delta_n}|^2).$$

Hence

$$\sup_{v\in[i\Delta_n,(i+1)\Delta_n[}\left|\mathbb{E}\left(b(Y_v,X_v)-b(Y_{i\Delta_n},\overline{Y}_{i,n})|\mathcal{G}^n_i\right)\right| \le \Delta_n c(1+|Y_{i\Delta_n}|^2+|X_{i\Delta_n}|^2)$$

and $\left|\mathbb{E}\left(\frac{b_{i,n}}{\Delta_n}|\mathcal{G}^n_i\right)\right| \le \Delta_n^2 c(1+|Y_{i\Delta_n}|^2+|X_{i\Delta_n}|^2)$. Similarly, $\left|\mathbb{E}\left(\frac{b'_{i+1,n}}{\Delta_n}|\mathcal{G}^n_i\right)\right| \le \Delta_n^2 c(1+|Y_{(i+1)\Delta_n}|^2+|X_{(i+1)\Delta_n}|^2)$ and the bound on $|\mathbb{E}(\varepsilon^P_{i,n}|\mathcal{G}^n_i)|$ is proved.

∘ We now bound $|\mathbb{E}((\varepsilon^P_{i,n})^2|\mathcal{G}^n_i)|$ and $|\mathbb{E}((\varepsilon^P_{i,n})^4|\mathcal{G}^n_i)|$. Using the Cauchy–Schwarz inequality, it is sufficient to bound $|\mathbb{E}((\varepsilon^P_{i,n})^4|\mathcal{G}^n_i)|$. By Assumption (A2) and Proposition 5, we obtain $\mathbb{E}\left(\left|\frac{b_{i,n}}{\Delta_n}\right|^4|\mathcal{G}^n_i\right) \le \Delta_n^4 c(1+|Y_{i\Delta_n}|^4+|X_{i\Delta_n}|^4)$ and similarly $\mathbb{E}\left(\left|\frac{b'_{i+1,n}}{\Delta_n}\right|^4|\mathcal{G}^n_i\right) \le \Delta_n^4 c(1+|Y_{(i+1)\Delta_n}|^4+|X_{(i+1)\Delta_n}|^4)$. We have to bound $\mathbb{E}\left(\left|\frac{a_{i,n}}{\Delta_n}\right|^4|\mathcal{G}^n_i\right)$. Using the Burkholder–Davis–Gundy inequality, Proposition 5 and Assumption (A2), we get

$$\begin{aligned}\mathbb{E}\left(\left|\frac{a_{i,n}}{\Delta_n}\right|^4\middle|\mathcal{G}^n_i\right) &\le \frac{c}{\Delta_n^4}\mathbb{E}\left(\int_{i\Delta_n}^{(i+1)\Delta_n}\left(a(Y_v,X_v)-a(Y_{i\Delta_n},X_{i\Delta_n})\right)^4dv \times \int_{i\Delta_n}^{(i+1)\Delta_n}\left((v-i\Delta_n)^4dv\right)|\mathcal{G}^n_i\right)\\ &\le c\Delta_n^4(1+|Y_{i\Delta_n}|^4+|X_{i\Delta_n}|^4).\end{aligned}$$

Similarly, we obtain $\mathbb{E}\left(\left|\frac{a'_{i+1,n}}{\Delta_n}\right|^4|\mathcal{G}^n_i\right) \le c\Delta_n^4(1+|Y^4_{(i+1)\Delta_n}|+|X_{(i+1)\Delta_n}|^4)$. This completes the proof for the bound of $|\mathbb{E}(\varepsilon^4_{i,n}|\mathcal{G}^n_i)|$.

∘ We now proof $|\mathbb{E}(\varepsilon_{i,n}U_{i,n}|\mathcal{G}_i^n)| \le c\Delta_n^{3/2}(1+|X_{i\Delta_n}|^2+|Y_{i\Delta_n}|^2)$. From the definitions of $(a_{i,n}, a'_{i+1,n}, b_{i,n}, b'_{i+1,n})$, we can prove the following inequalities

$$|\mathbb{E}(a_{i,n}\xi_{i,n}|\mathcal{G}_i^n)| \le \Delta_n^{5/2}c(1+|Y_{i\Delta_n}|+|X_{i\Delta_n}|), \qquad |\mathbb{E}(a'_{i+1,n}\xi_{i,n}|\mathcal{G}_i^n)| = 0$$
$$|\mathbb{E}(b_{i,n}\xi_{i,n}|\mathcal{G}_i^n)| \le \Delta_n^{5/2}c(1+|Y_{i\Delta_n}|+|X_{i\Delta_n}|),$$
$$|\mathbb{E}(b'_{i+1,n}\xi_{i,n}|\mathcal{G}_i^n)| \le \Delta_n^{5/2}c(1+|Y_{(i+1)\Delta_n}|+|X_{(i+1)\Delta_n}|).$$

Hence the results for $|\mathbb{E}(\varepsilon_{i,n}U_{i,n}|\mathcal{G}_i^n)|$.

∘ Proposition 5 yields

$$\mathbb{E}\left(\sup_{s\in[i\Delta_n,(i+2)\Delta_n[}|X_s - X_{i\Delta_n}|^k\right) \le c\Delta_n^{k/2}(1+|Y_{(i+1)\Delta_n}|^k+|X_{(i+1)\Delta_n}|^k)$$

which provides

$$\mathbb{E}\left(\left|\overline{Y}_{i+1,n}-\overline{Y}_{i,n}\right|^k|\mathcal{G}_i^n\right) \le c\Delta_n^{k/2}(1+|Y_{(i+1)\Delta_n}|^k+|X_{(i+1)\Delta_n}|^k). \quad \square$$

Proof of Proposition 4. The first assertion in the complete observations case is based on the convergence of the Euler scheme [1]. For partial observations case, Taylor's expansion ensures that there exists $s \in (\overline{Y}_{i,n}, X_{i\Delta_n})$ such that

$$f(Y_{i\Delta_n}, \overline{Y}_{i,n}, \theta) = f(Y_{i\Delta_n}, X_{i\Delta_n}, \theta) + f'_x(Y_{i\Delta_n}, X_s, \theta)(\overline{Y}_{i,n} - X_{i\Delta_n}).$$

Thus we deduce that $\mathbb{E}(\sup|f(Y_{i\Delta_n}, \overline{Y}_{i,n}, \theta) - f(Y_{i\Delta_n}, X_{i\Delta_n}, \theta)||\mathcal{G}_i^n) \le c\Delta_n^{1/2}(1+|X_{i\Delta_n}| + |Y_{i\Delta_n}|)$. Hence, the L^1 convergence of $\sup\frac{1}{n}\sum_{i=0}^n|f(Y_{i\Delta_n}, \overline{Y}_{i,n}, \theta) - f(Y_{i\Delta_n}, X_{i\Delta_n}, \theta)|$ is proved. The results yields by applying Proposition 2. $\square$

Proof of Theorem 3. The scheme of the proof is the same for both complete and partial observations cases, but the arguments are simpler for the complete observations case. We only detail the second case. Set $\tilde{I}_n^P(f) = \frac{1}{n\Delta_n}\sum_{i=0}^{n-2} f(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n}, \theta)\left(\overline{Y}_{i+1,n} - \overline{Y}_i - \Delta_n b(Y_{i\Delta_n}, \overline{Y}_{i,n})\right)$. We can write

$$\overline{I}_n^P(f) = \tilde{I}_n^P(f) + \frac{1}{n\Delta_n}\sum_{i=1}^{n-1}\Delta_n f(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n}, \theta)(b(Y_{i\Delta_n}, \overline{Y}_i) - b(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n}))$$

we first study the convergence of $\tilde{I}_n^P(f)$ and then we deduce the result for $\overline{I}_n^P(f)$.

We have $\tilde{I}_n^P(f) = \frac{1}{n\Delta_n}\sum_{i=0}^{n-1} Z_{i,n}(\theta)$ with $Z_{i,n}(\theta) = f(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n}, \theta)\ (\overline{Y}_{i+1,n} - \overline{Y}_i - \Delta_n b(Y_{i\Delta_n}, \overline{Y}_{i,n}))$. The random variable $\overline{Y}_{i,n}$ is $\mathcal{G}_{i+1}^n$-measurable and $Z_{i,n}(\theta)$ is $\mathcal{G}_{i+2}^n$-measurable. We split $\tilde{I}_n^P(f)$ into the sum of three terms

$$\tilde{I}_n^P(f) = \frac{1}{n\Delta_n}\left(\sum_{i=0}^{n-1} Z_{3i,n}(\theta) + \sum_{i=0}^{n-1} Z_{3i+1,n}(\theta) + \sum_{i=0}^{n-1} Z_{3i+2,n}(\theta)\right).$$

To prove (13), it is enough to show that $\frac{1}{n\Delta_n}\sum_{i=0}^{n-1} Z_{3i,n}(\theta) \xrightarrow[n\to\infty]{\mathbb{P}} 0$ uniformly in θ, in probability (the proof for the convergence of $\frac{1}{n\Delta_n}\sum_{i=0}^{n-1} Z_{3i+1,n}(\theta)$ and $\frac{1}{n\Delta_n}\sum_{i=0}^{n-1} Z_{3i+2,n}(\theta)$ is analogous). Using Proposition 3, we set $Z_{i,n}(\theta) = z_{i,n}^{(2)}(\theta) + z_{i,n}^{(2)}(\theta)$ with

$$z_{i,n}^{(1)}(\theta) = f(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n}, \theta)\Delta_n^{1/2} a(Y_{i\Delta_n}, X_{i\Delta_n})U_{i,n}$$
$$z_{i,n}^{(2)}(\theta) = f(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n}, \theta)\varepsilon_{i,n}^P.$$

To prove $\frac{1}{n\Delta_n}\sum_{i=0}^{n-1} z_{3i,n}^{(j)}(\theta) \xrightarrow[n\to\infty]{\mathbb{P}} 0$ for $j = 1, 2$, we use Lemma A2 of [8]. It is thus enough to prove

$$\frac{1}{n\Delta_n}\sum_{i=0}^{n-1}\mathbb{E}(z_{3i,n}^{(j)}(\theta)|\mathcal{G}_{3i}^n) \xrightarrow[n\to\infty]{\mathbb{P}} 0$$
$$\frac{1}{n^2\Delta_n^2}\sum_{i=0}^{n-1}\mathbb{E}((z_{3i,n}^{(j)}(\theta))^2|\mathcal{G}_{3i}^n) \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

As $\overline{Y}_{3i-1,n}$ is $\mathcal{G}_{3i}^n$ measurable and $\mathbb{E}(U_{3i,n}|\mathcal{G}_{3i}^n) = 0$, we have $\mathbb{E}(z_{i,n}^{(1)}(\theta)|\mathcal{G}_{3i}^n) = 0$. Using $\mathbb{E}(U_{3i,n}^2|\mathcal{G}_{3i}^n) = 2/6$, we get

$$\mathbb{E}((z_{3i,n}^{(1)}(\theta))^2|\mathcal{G}_{3i}^n) = \frac{2}{6}\Delta_n f^2(Y_{(3i-1)\Delta_n}, \overline{Y}_{3i-1,n}, \theta)a^2(Y_{3i\Delta_n}, X_{3i\Delta_n}).$$

Assumptions (A1)–(A2) yields $\frac{1}{n^2\Delta_n^2}\sum_{i=0}^{n-1}\mathbb{E}((z_{3i,n}^{(1)}(\theta))^2|\mathcal{G}_{3i}^n) \xrightarrow[n\to\infty]{\mathbb{P}} 0$. For $z_{3i,n}^{(2)}(\theta)$, using Proposition 3, we get $\mathbb{E}(z_{3i,n}^{(1)}(\theta)|\mathcal{G}_{3i}^n) \le cf(Y_{(3i-1)\Delta_n}, \overline{Y}_{3i-1,n}, \theta)\Delta_n^2(1 + |Y_{(3i-1)\Delta_n}|^3 + |X_{(3i-1)\Delta_n}|^3)$ and thus $\frac{1}{n\Delta_n}\sum_{i=0}^{n-1}\mathbb{E}(z_{3i,n}^{(2)}(\theta)|\mathcal{G}_{3i}^n) \xrightarrow[n\to\infty]{\mathbb{P}} 0$.

Similarly, we have $\mathbb{E}((z_{3i,n}^{(2)}(\theta))^2|\mathcal{G}_{3i}^n) \le cf^2(Y_{3i\Delta_n}, \overline{Y}_{3i,n}, \theta)\Delta_n^2(1 + |Y_{3i\Delta_n}|^4 + |X_{3i\Delta_n}|^4)$ and thus $\frac{1}{n^2\Delta_n^2}\sum_{i=0}^{n-1}\mathbb{E}((z_{3i,n}^{(2)}(\theta))^2|\mathcal{G}_{3i}^n) \xrightarrow[n\to\infty]{\mathbb{P}} 0$. This gives the convergence in probability of $\tilde{I}_n^P(f)$ for all θ.

To obtain uniformity with respect to θ, we use the Proposition 1 of [8]. It is enough to show $\sup_{n\in\mathbb{N}}\mathbb{E}\left(\sup_\theta|\partial_\theta\tilde{I}_n^P(f)|\right) < \infty$. We have

$$\partial_\theta\tilde{I}_n^P(f) = \frac{1}{n\Delta_n}\sum_{i=1}^{n-1}\partial_\theta f(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n}, \theta)\left(\Delta_n^{1/2}a(Y_{i\Delta_n}, X_{i\Delta_n})U_{i,n} + \varepsilon_{i,n}\right).$$

As $\mathbb{E}(U_i, n|\mathcal{G}_i^n) = 0$ and $\mathbb{E}(\varepsilon_{i,n}|\mathcal{G}_i^n) \le c\Delta_n^2(1 + |X_{i\Delta_n}|^3 + |Y_{i\Delta_n}|^3)$, we have $\mathbb{E}\left(\partial_\theta f_\theta(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n})\left(\Delta_n^{1/2}a(Y_{i\Delta_n}, X_{i\Delta_n})U_{i,n} + \varepsilon_{i,n}\right)|\mathcal{G}_i^n\right) \le c\Delta_n(1 + |X_{i\Delta_n}|^3 + |Y_{i\Delta_n}|^3)$. With Assumption (A2), it implies

$$\mathbb{E}\left(\partial_\theta f(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n}, \theta)\left(\Delta_n^{1/2}a(Y_{i\Delta_n}, X_{i\Delta_n})U_{i,n} + \varepsilon_{i,n}\right)|\mathcal{G}_i^n\right) \le c\Delta_n.$$

Hence, $\sup_{n\in\mathbb{N}}\mathbb{E}\left(\sup_\theta|\partial\tilde{I}_n^P(f)|\right) < \infty$ and uniformity in θ follows. We can now deduce the result for $\overline{I}_n^P(f)$. Taylor's formula gives the existence of s_1 and s_2 such that

$$\begin{aligned} b(Y_{i\Delta_n}, \overline{Y}_i) - b(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n}) &= b'_y(Y_{s_1}, \overline{Y}_i)(Y_{i\Delta_n} - Y_{(i-1)\Delta_n}) \\ &\quad + b'_x(Y_{(i-1)\Delta_n}, X_{s_2})(\overline{Y}_i - \overline{Y}_{i-1,n}). \end{aligned}$$

Assumptions (A1)–(A2), Cauchy–Schwarz inequality and $Y_{i\Delta_n} - Y_{(i-1)\Delta_n} = \Delta_n \overline{Y}_{i-1,n}$ imply

$$\begin{aligned}
&\mathbb{E}\left(\left|b(Y_{i\Delta_n}, \overline{Y}_i) - b(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n})\right| |\mathcal{G}_i^n\right) \\
&\le c\left(\mathbb{E}\left(\left|\Delta_n \overline{Y}_{i-1,n}\right|^2 |\mathcal{G}_i^n\right)^{1/2} + \mathbb{E}\left(\left|\overline{Y}_i - \overline{Y}_{i-1,n}\right|^2 |\mathcal{G}_i^n\right)^{1/2}\right)(1 + |X_{i\Delta_n}| + |Y_{i\Delta_n}|) \\
&\le c\Delta_n^{1/2}(1 + |X_{i\Delta_n}| + |Y_{i\Delta_n}|).
\end{aligned}$$

This implies

$$\frac{1}{n\Delta_n}\sum_{i=1}^{n-1} \Delta_n f(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n}, \theta)(b(Y_{i\Delta_n}, \overline{Y}_i) - b(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n})) \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Hence the result. □

Proof of Theorem 4. We only detail the partial observations case. We set $W_{i,n}(\theta) = f(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n}, \theta)(\overline{Y}_{i+1,n} - \overline{Y}_{i,n})^2$ such that $\overline{Q}_n^P(f) = \frac{1}{n\Delta_n}\sum W_{i,n}(\theta)$. We split the sum into the sum of three terms $W_{3i,n}$, $W_{3i+1,n}$ and $W_{3i+2,n}$. Given this partition, it is enough to show that

$$(n\Delta_n)^{-1}\sum_{i=1}^{n-1} W_{3i,n}(\theta) \xrightarrow[n\to\infty]{\mathbb{P}} \frac{2}{3}\nu_0(f(\cdot,\cdot,\theta)a^2(\cdot,\cdot)).$$

As the expression of $\overline{Q}_n^P$ is slightly different from [8], we are able to write $W_{3i,n}(\theta)$ as the sum of only three terms (instead four). Using Taylor's formula, there exists $X_s \in (\overline{Y}_{3i,n}, X_{3i\Delta_n})$ such that we can write $W_{3i,n}(\theta) = w_{3i,n}^{(1)}(\theta) + w_{3i,n}^{(2)}(\theta) + w_{3i,n}^{(3)}(\theta)$ with

$$\begin{aligned}
w_{3i,n}^{(1)}(\theta) &= \Delta_n a^2(Y_{3i\Delta_n}, X_{3i\Delta_n})U_{3i,n}^2 f(Y_{(3i-1)\Delta_n}, \overline{Y}_{3i-1,n}, \theta) \\
w_{3i,n}^{(2)}(\theta) &= 2\Delta_n^{1/2}U_{3i,n}a(Y_{3i\Delta_n}, X_{3i\Delta_n})f(Y_{(3i-1)\Delta_n}, \overline{Y}_{3i-1,n}, \theta)(\varepsilon_{3i,n} \\
&\quad + \Delta_n b(Y_{3i\Delta_n}, X_{3i\Delta_n}) + \Delta_n b_x'(Y_{3i\Delta_n}, X_s)(\overline{Y}_{3i,n} - X_{3i\Delta_n})) \\
w_{3i,n}^{(3)}(\theta) &= (\varepsilon_{3i,n} + \Delta_n b(Y_{3i\Delta_n}, X_{3i\Delta_n}) + \Delta_n b_x'(Y_{3i\Delta_n}, X_s)(\overline{Y}_{3i,n} - X_{3i\Delta_n}))^2 \\
&\quad \times f(Y_{(3i-1)\Delta_n}, \overline{Y}_{3i-1,n}, \theta).
\end{aligned}$$

We set $\overline{Q}_n^{(Pj)}(\theta) = (n\Delta_n)^{-1}\sum_{i=1}^{n-1} w_{3i,n}^{(j)}(\theta)$, for $j = 1, 2, 3$. We start by studying $\overline{Q}_n^{(P1)}(\theta)$. Using $\mathbb{E}(U_{3i,n}^2|\mathcal{G}_{3i}^n) = 2/3$ and $\mathbb{E}(U_{3i,n}^4|\mathcal{G}_{3i}^n) = 4/3$ and the fact that $\overline{Y}_{3i-1,n}$ is $\mathcal{G}_{3i}^n$-measurable, we obtain:

$$\begin{aligned}
\mathbb{E}(w_{3i}^{(1)}(\theta)|\mathcal{G}_{3i}^n) &= \frac{2\Delta_n}{3}a^2(Y_{3i\Delta_n}, X_{3i\Delta_n})f(Y_{(3i-1)\Delta_n}, \overline{Y}_{3i-1,n}, \theta) \\
\mathbb{E}((w_{3i}^{(1)}(\theta))^2|\mathcal{G}_{3i}^n) &= \frac{4\Delta_n^2}{3}a^4(Y_{3i\Delta_n}, X_{3i\Delta_n})f({}^2Y_{(3i-1)\Delta_n}, \overline{Y}_{3i-1,n}, \theta).
\end{aligned}$$

Thus, applying Lemma A1 of [8], we get

$$(n\Delta_n)^{-1}\sum_{i=0}^{n-1}\mathbb{E}(w_{3i}^{(1)}(\theta)|\mathcal{G}_{3i}^n) \xrightarrow[n\to\infty]{\mathbb{P}} \frac{2}{3}\nu_0(f(\cdot,\cdot,\theta)a^2(\cdot,\cdot))$$

and by Assumption (A2), we get $\mathbb{E}\left|\mathbb{E}((w_{3i}^{(1)}(\theta))^2|\mathcal{G}_{3i}^n)\right| \le c\Delta_n^2$ and therefore, $(n\Delta_n)^{-1}\sum_{i=0}^{n-1}$ $\mathbb{E}((w_{3i}^{(1)}(\theta))^2|\mathcal{G}_{3i}^n) \xrightarrow[n\to\infty]{\mathbb{P}} 0$. By Lemma A2 of [8], we deduce $\overline{Q}_n^{(P1)}(\theta) \xrightarrow[n\to\infty]{\mathbb{P}}$

$\frac{2}{3}\nu_0(f(\cdot,\cdot,\theta)a^2(\cdot,\cdot))$ in probability. Using Proposition 3 and Lemma A2 of [8], we easily prove that $\overline{Q}_n^{(P2)}(\theta) \xrightarrow[n\to\infty]{\mathbb{P}} 0$ and $\overline{Q}_n^{(P3)}(\theta) \xrightarrow[n\to\infty]{\mathbb{P}} 0$. The uniformity is obtained by bounding

$$\sup_{n\in\mathbb{N}}(n\Delta_n)^{-1}\sum_{i=0}^{n-1}\mathbb{E}\left((\overline{Y}_{i+1,n}-\overline{Y}_{i,n})^2\sup_\theta\left|\partial_\theta f(Y_{i\Delta_n},\overline{Y}_{i,n},\theta)\right|\right)<\infty$$

due to Proposition 1 of [8]. This is easily obtained using Proposition 2 and Assumption (A2). □

Proof of Theorem 5. We only detail the partial observations case. We have

$$\sqrt{n\Delta_n}\overline{I}_n^P(f) = \sqrt{n\Delta_n}\tilde{I}_n^P(f) + \frac{1}{\sqrt{n\Delta_n}} \times \sum_{i=2}^{n-1} f(Y_{(i-1)\Delta_n},\overline{Y}_{i-1,n},\theta)\Delta_n(b(Y_{i\Delta_n},\overline{Y}_i)-b(Y_{(i-1)\Delta_n},\overline{Y}_{i-1,n})).$$

We first study the distribution convergence of $\sqrt{n\Delta_n}\tilde{I}_n^P(f)$ then we deduce the result for $\sqrt{n\Delta_n}\overline{I}_n^P(f)$. Using the same notations as in Theorem 4, we set $\sqrt{n\Delta_n}\tilde{I}_n^P(f) = N_n^{(1)} + N_n^{(2)}$ with $N_n^{(1)} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n-1} f(Y_{(i-1)\Delta_n},\overline{Y}_{(i-1)\Delta_n},\theta)a(Y_{i\Delta_n},X_{i\Delta_n})(\xi_{i,n}+\xi'_{i+1,n})$ and $N_n^{(2)} = \frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^{n-1} f(Y_{(i-1)\Delta_n},\overline{Y}_{(i-1)\Delta_n},\theta)\varepsilon_{in}$. First, we study $N_n^{(1)}$. In order to use a martingale central limit theorem, we reorder the terms

$$N_n^{(1)} = \frac{1}{\sqrt{n}}f(Y_0,\overline{Y}_0,\theta)a(Y_{\Delta_n},X_{\Delta_n})\xi_{0,n} + \frac{1}{\sqrt{n}}\sum_{i=2}^{n-1}s_{in}^{(1)} + \frac{1}{\sqrt{n}}f(Y_{(n-2)\Delta_n},\overline{Y}_{(n-2)\Delta_n},\theta)a(Y_{(n-1)\Delta_n},X_{(n-1)\Delta_n})\xi'_{n,n} \tag{A.1}$$

with

$$s_{in}^{(1)} = f(Y_{(i-1)\Delta_n},\overline{Y}_{(i-1)\Delta_n},\theta)a(Y_{i\Delta_n},X_{i\Delta_n})\xi_{i,n} + f(Y_{(i-2)\Delta_n},\overline{Y}_{(i-2)\Delta_n},\theta)a(Y_{(i-1)\Delta_n},X_{(i-1)\Delta_n})\xi'_{i,n}.$$

We have $\mathbb{E}(s_{in}^{(1)}|\mathcal{G}_i^n)=0$ and we compute the conditional variance $\mathbb{E}[(s_{in}^{(1)})^2|\mathcal{G}_i^n]$:

$$\begin{aligned}\mathbb{E}[(s_{in}^{(1)})^2|\mathcal{G}_i^n] = \frac{1}{3}\{&f^2(Y_{(i-1)\Delta_n},\overline{Y}_{(i-1)\Delta_n},\theta)a^2(Y_{i\Delta_n},X_{i\Delta_n})\\ &+ f(Y_{(i-1)\Delta_n},\overline{Y}_{(i-1)\Delta_n},\theta)f(Y_{(i-2)\Delta_n},\overline{Y}_{(i-2)\Delta_n},\theta)\\ &\times a(Y_{i\Delta_n},X_{i\Delta_n})a(Y_{(i-1)\Delta_n},X_{(i-1)\Delta_n})\\ &+ f^2(Y_{(i-2)\Delta_n},\overline{Y}_{(i-2)\Delta_n},\theta)a^2(Y_{(i-1)\Delta_n},X_{(i-1)\Delta_n})\}.\end{aligned}$$

We want to prove that $\frac{1}{n}\sum_{i=2}^{n-1}\mathbb{E}[(s_{in}^{(1)})^2|\mathcal{G}_i^n] \xrightarrow[n\to\infty]{\mathbb{P}} \nu_0(f^2a^2)$. We first start with the term $\frac{1}{n}\sum_{i=2}^{n-1} f^2(Y_{(i-1)\Delta_n},\overline{Y}_{(i-1)\Delta_n},\theta)a^2(Y_{i\Delta_n},X_{i\Delta_n})$. By the ergodic theorem, we have

$$\frac{1}{n}\sum_{i=2}^{n-1} f^2(Y_{(i-1)\Delta_n},X_{(i-1)\Delta_n},\theta)a^2(Y_{i\Delta_n},X_{i\Delta_n}) \xrightarrow[n\to\infty]{\mathbb{P}} \nu_0(f^2a^2).$$

A Taylor development and the Cauchy–Schwarz inequality provide the convergence in L^1 towards 0 of $\sup_\theta \frac{1}{n}\sum_{i=2}^{n-1} |f^2(Y_{(i-1)\Delta_n},\overline{Y}_{(i-1)\Delta_n},\theta)a^2(Y_{i\Delta_n},X_{i\Delta_n})-f^2(Y_{(i-1)\Delta_n},X_{(i-1)\Delta_n},\theta)$

$a^2(Y_{iDn}, X_{i\Delta_n})|$ using Assumptions (A1)–(A2). The terms $f(Y_{(i-1)\Delta_n}, \overline{Y}_{(i-1)\Delta_n}, \theta)$ $f(Y_{(i-2)\Delta_n}, \overline{Y}_{(i-2)\Delta_n}, \theta)a(Y_{i\Delta_n}, X_{i\Delta_n})$ $a(Y_{(i-1)\Delta_n}, X_{(i-1)\Delta_n})$ and $f^2(Y_{(i-2)\Delta_n}, \overline{Y}_{(i-2)\Delta_n}, \theta)$ $a^2(Y_{(i-1)\Delta_n}, X_{(i-1)\Delta_n})$ are similar.

We easily bound $\mathbb{E}[(s_{in}^{(1)})^4|\mathcal{G}_i^n]$ and show that $\frac{1}{n^2}\sum_{i=2}^{n-1}\mathbb{E}[(s_{in}^{(1)})^4|\mathcal{G}_i^n] \xrightarrow[n\to\infty]{L^1} 0$. By the martingale central limit theorem, we deduce that $\frac{1}{\sqrt{n}}\sum_{i=2}^{n-1} s_{in}^{(1)} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, \nu_0(f^2a^2))$. By (A.1), we deduce $N_n^{(1)} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, \nu_0(f^2a^2))$.

We now have to prove the convergence to 0 of $N_n^{(2)}$. Using Proposition 3, we easily have the convergence to 0 of $\frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^{n-1}\mathbb{E}[f(Y_{(i-1)\Delta_n}, \overline{Y}_{(i-1)\Delta_n}, \theta)\varepsilon_{in}|\mathcal{G}_i^n]$ in probability. Similarly, we obtain that $\frac{1}{n\Delta_n}\sum_{i=1}^{n-1}\mathbb{E}[f^2(Y_{(i-1)\Delta_n}, \overline{Y}_{(i-1)\Delta_n}, \theta)\varepsilon_{in}^2|\mathcal{G}_i^n]$ converges to 0 in probability. Thus, using Proposition 5, we get $N_n^{(2)} \xrightarrow[n\to\infty]{\mathbb{P}} 0$. This implies

$$\frac{1}{n\Delta_n}\sum_{i=1}^{n-1}\Delta_n f(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n}, \theta)(b(Y_{i\Delta_n}, \overline{Y}_i) - b(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n})) \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

This gives the convergence in distribution of $\sqrt{n\Delta_n}\tilde{I}_n^P(f)$. To deduce the results for $\sqrt{n\Delta_n}\overline{I}_n^P(f)$, we remark that

$$\begin{aligned} b(Y_{i\Delta_n}, \overline{Y}_i) - b(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n}) &= b(Y_{i\Delta_n}, \overline{Y}_i) - b(Y_{(i-1)\Delta_n}, \overline{Y}_i) \\ &\quad + b(Y_{(i-1)\Delta_n}, \overline{Y}_i) - b(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n}). \end{aligned}$$

Taylor's development gives

$$\mathbb{E}[|b(Y_{i\Delta_n}, \overline{Y}_{i,n}) - b(Y_{i\Delta_n}, \overline{Y}_{i-1,n})|\mathcal{G}_i^n] \le c\sqrt{\Delta_n}(1 + |X_{i\Delta_n}| + |Y_{i\Delta_n}|).$$

Using $b(Y_{i\Delta_n}, \overline{Y}_i) - b(Y_{(i-1)\Delta_n}, \overline{Y}_i) = \int_{(i-1)\Delta_n}^{i\Delta_n} b'_y(Y_s, \overline{Y}_{i-1,n})(X_{(i-1)\Delta_n} + \int_{(i-1)\Delta_n}^{s} b(V_u, X_u)du + \int_{(i-1)\Delta_n}^{s} a(V_u, X_u)dB_u)ds$ and the Burkholder inequality, this yields

$$\mathbb{E}[|b(Y_{i\Delta_n}, \overline{Y}_i) - b(Y_{(i-1)\Delta_n}, \overline{Y}_i)|\,|\mathcal{G}_i^n] \le c\sqrt{\Delta_n}(1 + |X_{i\Delta_n}| + |Y_{i\Delta_n}|).$$

Using Assumptions (A1)–(A4), we deduce the convergence to 0 in probability of $\frac{1}{\sqrt{n\Delta_n}}\sum_{i=2}^{n-1} f(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n}, \theta)\Delta_n(b(Y_{i\Delta_n}, \overline{Y}_i) - b(Y_{(i-1)\Delta_n}, \overline{Y}_{i-1,n}))$. We deduce that $\sqrt{n\Delta_n}\overline{I}_n^P(f) - \sqrt{n\Delta_n}\tilde{I}_n^P(f) = o_{\mathbb{P}}(1)$. □

Proof of Theorem 6. We only detail the partial observations case. We use the same notations as in Theorem 4. Set $\overline{M}_n(f) = \sqrt{n}\left(\overline{Q}_n^P(f) - \frac{2}{3}\overline{\nu}_n(fa^2)\right)$ and $\beta(y,x) = a^2(y,x)f(y,x,\theta)$. We have

$$\begin{aligned} \overline{M}_n(f) = \sqrt{n}\Bigg[\frac{1}{n\Delta_n}\sum_{i=2}^{n-1}\Big(&\Delta_n f(Y_{i\Delta_n}, \overline{Y}_i, \theta)a^2(Y_{i\Delta_n}, X_{i\Delta_n})U_i^2 \\ &+ f(Y_{i\Delta_n}, \overline{Y}_i, \theta)(\varepsilon_{i,n}^P + \Delta_n b(Y_{i\Delta_n}, \overline{Y}_i))^2 \\ &+ 2\Delta_n^{1/2} f(Y_{i\Delta_n}, \overline{Y}_i, \theta)a(Y_{i\Delta_n}, X_{i\Delta_n})U_i(\varepsilon_{i,n}^P \\ &+ \Delta_n b(Y_{i\Delta_n}, \overline{Y}_i))\Big) - \frac{2}{3n}\sum_{i=2}^{n-1} f(Y_{i\Delta_n}, \overline{Y}_i, \theta)a^2(Y_{i\Delta_n}, \overline{Y}_i)\Bigg]. \end{aligned}$$

By Taylor expansion, there exists $X_v \in (X_{i\Delta_n}, \overline{Y}_{i,n})$ such that

$$\overline{M}_n(f) = \sqrt{n}\left[\frac{1}{n\Delta_n}\sum_{i=2}^{n-1}\left(\Delta_n\beta(Y_{i\Delta_n}, X_{i\Delta_n})\left(U_i^2 - \frac{2}{3}\right)\right.\right.$$
$$+ f(Y_{i\Delta_n}, X_{i\Delta_n}, \theta)(\varepsilon_{i,n}^P + \Delta_n b(Y_{i\Delta_n}, \overline{Y}_i))^2$$
$$+ 2\Delta_n^{1/2} f(Y_{i\Delta_n}, X_{i\Delta_n}, \theta)a(Y_{i\Delta_n}, X_{i\Delta_n})U_i(\varepsilon_{i,n}^P$$
$$\left. + \ \Delta_n b(Y_{i\Delta_n}, \overline{Y}_i))(\overline{Y}_{i+1,n} - \overline{Y}_{i,n})^2(\overline{Y}_{i,n} - X_{i\Delta_n})f_x'(Y_{i\Delta_n}, X_v, \theta)\right)$$
$$\left. - \ \frac{2}{3n}\sum_{i=2}^{n-1}(\beta(Y_{i\Delta_n}, \overline{Y}_i) - \beta(Y_{i\Delta_n}, X_{i\Delta_n}))\right].$$

Thus $\overline{M}_n(f) = \sum_{l=1}^{5}\overline{M}_n^{(l)}$ with

$$\overline{M}_n^{(1)} = \frac{1}{\sqrt{n}}\sum_{i=2}^{n-1}\beta(Y_{i\Delta_n}, X_{i\Delta_n})\left(U_i^2 - \frac{2}{3}\right)$$
$$\overline{M}_n^{(2)} = \frac{1}{\sqrt{n\Delta_n}}\sum_{i=2}^{n-1}2f(Y_{i\Delta_n}, X_{i\Delta_n}, \theta)a(Y_{i\Delta_n}, X_{i\Delta_n})U_i(\varepsilon_{i,n}^P + \Delta_n b(Y_{i\Delta_n}, \overline{Y}_i))$$
$$\overline{M}_n^{(3)} = \frac{1}{\sqrt{n}\Delta_n}\sum_{i=2}^{n-1}f(Y_{i\Delta_n}, X_{i\Delta_n}, \theta)(\varepsilon_{i,n}^P + \Delta_n b(Y_{i\Delta_n}, \overline{Y}_i))^2$$
$$\overline{M}_n^{(4)} = \frac{1}{\sqrt{n}\Delta_n}\sum_{i=2}^{n-1}(\overline{Y}_{i+1,n} - \overline{Y}_{i,n})^2(\overline{Y}_{i,n} - X_{i\Delta_n})f_x'(Y_{i\Delta_n}, X_v, \theta)$$
$$\overline{M}_n^{(5)} = \frac{2}{3\sqrt{n}}\sum_{i=2}^{n-1}(\beta(Y_{i\Delta_n}, \overline{Y}_i) - \beta(Y_{i\Delta_n}, X_{i\Delta_n})).$$

We first study the convergence of $\overline{M}_n^{(1)}$. Reordering terms to obtain a triangular array of martingale increments, we get

$$\overline{M}_n^{(1)} = \frac{1}{\sqrt{n}}\left\{\sum_{i=2}^{n-1}s_{in} + \left(\xi_{0,n}^2 - \frac{1}{3}\right)\beta(Y_0, X_0) + \left(\xi_{n,n}'^2 - \frac{1}{3}\right)\beta(Y_{(n-1)\Delta_n}, X_{(n-1)\Delta_n})\right.$$
$$\left. + \ 2\xi_{n-1,n}\xi_{n,n}'\beta(Y_{(n-1)\Delta_n}, X_{(n-1)\Delta_n})\right\}$$

where $s_{in} = \left(\xi_{i,n}^2 - \frac{1}{3}\right)\beta(Y_{i\Delta_n}, X_{i\Delta_n}) + \left(\xi_{i,n}'^2 - \frac{1}{3}\right)\beta(Y_{(i-1)\Delta_n}, X_{(i-1)\Delta_n}) + 2\xi_{i-1,n}\xi_{i,n}'\beta(Y_{(i-1)\Delta_n}, X_{(i-1)\Delta_n})$. But, s_{in} is $\mathcal{G}_{i+1}^n$ measurable and centered conditionally to $\mathcal{G}_i^n$. Furthermore, using the properties of $(\xi_{i,n}, \xi_{i,n}')$, we deduce

$$\begin{aligned}\mathbb{E}(s_{in}^2|\mathcal{G}_i^n) &= \frac{2}{9}\beta^2(Y_{i\Delta_n}, X_{i\Delta_n}) + \frac{2}{9}\beta^2(Y_{(i-1)\Delta_n}, X_{(i-1)\Delta_n}) \\ &\quad + \frac{4}{3}\xi_{i-1,n}^2\beta^2(Y_{(i-1)\Delta_n}, X_{(i-1)\Delta_n}) \\ &\quad + \frac{1}{9}\beta(Y_{(i-1)\Delta_n}, X_{(i-1)\Delta_n})\beta(Y_{i\Delta_n}, X_{i\Delta_n}).\end{aligned}$$

To prove the convergence of $\overline{M}_n^{(1)}$, it is sufficient to prove that

$$\frac{1}{n}\sum_{i=2}^{n-1}\mathbb{E}\left(|s_{in}^2|\,|\mathcal{G}_i^n\right) \xrightarrow[n\to\infty]{\mathbb{P}} \nu_0(\beta^2) \quad \text{and} \quad \frac{1}{n^2}\sum_{i=2}^{n-1}\mathbb{E}\left(|s_{in}^4|\,|\mathcal{G}_i^n\right) \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Indeed, applying Theorem 3.2 in [9], we get $\frac{1}{\sqrt{n}}\sum_{i=2}^{n-1} s_{in} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, \nu_0(\beta^2))$ and so does $\overline{M}_n^{(1)}$. By Lemma A2 of [8], we have $\frac{1}{n}\sum_{i=2}^{n-1}\xi_{i-1,n}^2\beta^2(Y_{(i-1)\Delta_n}, X_{(i-1)\Delta_n}) \xrightarrow[n\to\infty]{\mathbb{P}} 1/3\nu_0(\beta^2)$.

Thus, we deduce $\frac{1}{n}\sum_{i=2}^{n-1}\mathbb{E}\left(|s_{in}^2|\,|\mathcal{G}_i^n\right) \xrightarrow[n\to\infty]{\mathbb{P}} \nu_0(\beta^2)$. The bound on β^4 yields the convergence of $\frac{1}{n^2}\sum_{i=2}^{n-1}\mathbb{E}\left(|s_{in}^4|\,|\mathcal{G}_i^n\right)$.

We have to prove $\overline{M}_n^{(l)} \xrightarrow[n\to\infty]{\mathbb{P}} 0$ for $l = 2, \ldots, 5$. This holds true using that $n\Delta_n^2 \to 0$ and the hypothesis (A1) for $\overline{M}_n^{(5)}$. $\square$

References

[1] V. Bally, D. Talay, The law of the Euler scheme for stochastic differential equations I, convergence rate of the distribution function, Probab. Theory Related Fields 104 (1996) 43–60.

[2] S. Ditlevsen, M. Sørensen, Inference for observations of integrated diffusion processes, Scand. J. Stat. 31 (2004) 417–429.

[3] B. Favetto, A. Samson, Parameter estimation for a bidimensional partially observed Ornstein–Uhlenbeck process with biological application, Scand. J. Stat. 37 (2010) 200–220.

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

[5] C. Gardiner, Handbook of Stochastic Methods, Springer Verlag, New York, 1985.

[6] V. Genon-Catalot, J. Jacod, On the estimation of the diffusion coefficient for multi-dimensional diffusion processes, Ann. Inst. Henri Poincaré Probab. Stat. 29 (1993) 119–151.

[7] A. Gloter, Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient, ESAIM Probab. Stat. 4 (2000) 205–227.

[8] A. Gloter, Parameter estimation for a discretely observed integrated diffusion process, Scand. J. Stat. 33 (2006) 83–104.

[9] P. Hall, C.C. Heyde, Martingale Limit Theory and its Application, Academic Press Inc., Harcourt Brace Jovanovich Publishers, New York, 1980.

[10] M. Kessler, Estimation of an ergodic diffusion from discrete observations, Scand. J. Stat. 24 (1997) 211–229.

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

[12] V. Lemaire, Estimation récursive de la mesure invariante d'un processus de diffusion, Thèse de Doctorat de l' Univ. Marne-La-Vallée, 2005.

[13] J. Mattingly, A.M. Stuart, D. Higham, Ergodicity for sdes and approximations: locally lipschitz vector fields and degenerate noise, Stochastic Process. Appl. 101 (2002) 185–232.

[14] D. Nualart, The Malliavin calculus and related topics, in: Probability and its Applications (New York), second ed., Springer-Verlag, Berlin, 2006.

[15] Y. Pokern, A. Stuart, P. Wiberg, Parameter estimation for partially observed hypoelliptic diffusions, J. Roy. Stat. Soc. B 71 (2009) 49–73.

[16] B.L.S. Prakasa Rao, Statistical inference from sampled data for stochastic processes, in: Statistical Inference From Stochastic Processes (Ithaca, NY, 1987), in: Contemp. Math., vol. 80, Amer. Math. Soc., Providence, RI, 1988, pp. 249–284.

[17] D. Revuz, M. Yor, Continuous Martingales and Brownian Motion, in: Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences), vol. 293, Springer-Verlag, Berlin, 1991.

[18] N. Yoshida, Estimation for diffusion processes from discrete observation, J. Multivariate Anal. 41 (1992) 220–242.