ESTIMATION FOR STOCHASTIC DIFFERENTIAL EQUATIONS WITH A SMALL DIFFUSION COEFFICIENT

ARNAUD GLOTER AND MICHAEL SØRENSEN

Université Paris-Est and University of Copenhagen

Abstract. We consider a multidimensional diffusion X with drift coefficient $b(X_t, \alpha)$ and diffusion coefficient $\varepsilon a(X_t, \beta)$ where α and β are two unknown parameters, while ε is known. For a high-frequency sample of observations of the diffusion at the time points k/n, $k = 1, \ldots, n$, we propose a class of contrast functions and thus obtain estimators of (α, β). The estimators are shown to be consistent and asymptotically normal when $n \to \infty$ and $\varepsilon \to 0$ in such a way that $\varepsilon^{-1} n^{-\rho}$ remains bounded for some $\rho > 0$. The main focus is on the construction of explicit contrast functions, but it is noted that the theory covers quadratic martingale estimating functions as a special case. In a simulation study we consider the finite sample behaviour and the applicability to a financial model of an estimator obtained from a simple explicit contrast function.

1. Introduction

In this paper we consider a family of d-dimensional processes defined as the solution of

$$\text{(1)} \qquad \mathrm{d}X_t = b(X_t, \alpha)\mathrm{d}t + \varepsilon\sigma(X_t, \beta)\mathrm{d}W_t, \quad t \in [0, 1],$$
$$X_0 = x_0,$$

where $(\alpha, \beta) \in \overline{\Theta}_\alpha \times \overline{\Theta}_\beta$ with Θ_α and Θ_β being two open convex bounded subsets of respectively $\mathbb{R}^p$ and $\mathbb{R}^q$. The process (W_t) is an r-dimensional Wiener process; the function b is $\mathbb{R}^d$-valued and defined on $\mathbb{R}^d \times \overline{\Theta}_\alpha$; the function σ is defined on $\mathbb{R}^d \times \overline{\Theta}_\beta$

2000 *Mathematics Subject Classification.* 62F12, 62M05.

Key words and phrases. asymptotics, CIR-model, diffusion process with small noise, discrete-time observation, high-frequency data, minimum contrast estimation.

and takes values on the space of matrices $\mathbb{R}^d \otimes \mathbb{R}^r$; the initial value of the diffusion, $x_0 \in \mathbb{R}^d$, and $\varepsilon > 0$ are known. The only unknown quantity in (1) is the parameter $\theta = (\alpha, \beta)$. We denote the true value of the parameter by $\theta_0 = (\alpha_0, \beta_0)$ and assume that $\theta_0 \in \Theta = \Theta_\alpha \times \Theta_\beta$.

The small diffusion asymptotic $\varepsilon \to 0$ has been widely studied and has proved fruitful in applied problems. For illustrations of applications to contingent claim pricing, see Uchida & Yoshida (2004b) and references therein; for filtering problems, see e.g. Picard (1986, 1991). Several papers have been devoted to small diffusion asymptotics for parameter estimators in diffusion models. If the diffusion X is continuously observed on some finite interval, then the problem of estimation of the parameter α was treated by Kutoyants (1994), while semi-parametric problems were studied later (Kutoyants 1998; Iacus & Kutoyants 2001). Information criteria were studied by Uchida & Yoshida (2004a).

For a discretely observed process $(X_{t_k})_{k=0,\dots,n}$, Sørensen (2000) showed that by using martingale estimating functions parameters in the drift and diffusion coefficient may be estimated at rate the ε^{-1} as $\varepsilon \to 0$ even when the number of observations is fixed. For high-frequency data, where the process is observed at times $t_k = k/n$ with $n \to \infty$ and $\varepsilon = O(n^{-1/2})$, Genon–Catalot (1990) obtained an estimator of a drift coefficient parameter that is asymptotically equivalent to the maximum likelihood estimator based of the continuous time observation $(X_t)_{t\in[0,1]}$ and thus is efficient. Uchida (2004) considered a similar situation and obtained an efficient estimator from an approximate martingale estimating function. In Sørensen & Uchida (2003) estimation of both drift and diffusion coefficients parameters from the discrete time sampling $(X_{t_k})_{k=0,\dots,n}$ was treated. They obtained estimators that in a high-frequency and small-diffusion asymptotics are consistent, asymptotically Gaussian, and efficient for the estimation of the drift component parameter. However, they needed the restrictive condition that $\lim(\varepsilon\sqrt{n})^{-1} = M < \infty$.

In this paper, we extend the result of Sørensen & Uchida (2003) by proposing estimators for which the weaker condition that $\overline{\lim}\,(\varepsilon n^\rho)^{-1} < \infty$ for some $\rho > 0$ is sufficient. More precisely we obtain the following result. Let X^0 be the solution of the underlying deterministic system under the true value of the drift parameter:

$$\mathrm{d}X_t^0 = b(X_t^0, \alpha_0)\mathrm{d}t, \quad X_0^0 = x_0,$$

and introduce the matrix:

$$I(\theta_0) = \begin{pmatrix} \left(I_b^{i,j}(\theta_0)\right)_{1\le i,j\le p} & 0 \\ 0 & \left(I_\sigma^{i,j}(\theta_0)\right)_{1\le i,j\le q} \end{pmatrix}, \tag{2}$$

with

$$I_b^{i,j}(\theta_0) = \int_0^1 \left(\frac{\partial}{\partial\alpha_i} b(X_s^0, \alpha_0)\right)^* [\sigma\sigma^*]^{-1}(X_s^0, \beta_0) \left(\frac{\partial}{\partial\alpha_j} b(X_s^0, \alpha_0)\right) \mathrm{d}s,$$

$$I_\sigma^{i,j}(\theta_0) = \frac{1}{2}\int_0^1 \mathrm{tr}\left[\left(\frac{\partial}{\partial\beta_i}[\sigma\sigma^*]\right)[\sigma\sigma^*]^{-1}\left(\frac{\partial}{\partial\beta_j}[\sigma\sigma^*]\right)[\sigma\sigma^*]^{-1}(X_s^0, \beta_0)\right]\mathrm{d}s.$$

By M^* we denote the transpose of a matrix M. We will present an estimator $(\hat\alpha_{\varepsilon,n}, \hat\beta_{\varepsilon,n})$ obtained by minimizing an explicit contrast function based on the observations $(X_{t_k})_{k=0,\dots,n}$ (with $t_k = k/n$) for which

$$\begin{pmatrix} \varepsilon^{-1}(\hat\alpha_{\varepsilon,n} - \alpha_0) \\ \sqrt{n}(\hat\beta_{\varepsilon,n} - \beta_0) \end{pmatrix} \to \mathcal{N}(0, I(\theta_0)^{-1}).$$

The estimator of the drift parameter is efficient. The asymptotic variance of the diffusion parameter equals that of the estimator in Sørensen & Uchida (2003).

The structure of the paper is as follow. In Section 2 we present the construction of the contrast based estimator. Then we state the main result precisely and show that martingale estimating functions appear as a special case of this work. Let us emphasize that the conditions needed on the diffusion X are less restrictive than those needed in Sørensen & Uchida (2003). In particular, the coefficient σ need not be a Lipschitz function. As an example, we consider in detail the case of a two factor model with Cox–Ingersoll–Ross component and explore the applicability of the estimator to financial data in a simulation study in Section 3. The Section 4 is devoted to the proofs of the results.

2. Main results

We start this section by presenting the necessary conditions and the construction of the estimator.

2.1. **Basic assumptions.** Let us first introduce the following set of assumptions.

- **[A1]** For all $\varepsilon > 0$, the equation (1), with the true value of the parameter, admits a unique strong solution $X = X^\varepsilon$ on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$.
- **[A2]** The function b is smooth and Lipschitz on $\mathbb{R}^d \times \overline{\Theta}_\alpha$ (by smooth we mean that b is the restriction of some $\mathcal{C}^\infty$ function defined on a larger open set).
- **[A3]** σ is continuous, and there exists some open convex subset $\mathcal{U}$ of $\mathbb{R}^d$ such that $X_t^0 \in \mathcal{U}$ for all $t \in [0,1]$, and σ is smooth on $\mathcal{U} \times \overline{\Theta}_\beta$. Moreover $\sigma\sigma^*(x,\beta)$ is invertible on $\mathcal{U} \times \overline{\Theta}_\beta$.
- **[A4]** If $\alpha \neq \alpha_0$ then the two functions $t \mapsto b(X_t^0, \alpha_0)$ and $t \mapsto b(X_t^0, \alpha)$ are not equal.
  If $\beta \neq \beta_0$ then the two functions $t \mapsto \sigma\sigma^*(X_t^0, \beta_0)$ and $t \mapsto \sigma\sigma^*(X_t^0, \beta)$ are not equal.
- **[B]** $\varepsilon = \varepsilon_n \to 0$ and there exists a $\rho > 0$ such that $\overline{\lim}_{n\to\infty} (\varepsilon_n n^\rho)^{-1} < \infty$.

In Section 1 we introduced (X_t^0), the solution of the ordinary differential equation corresponding to $\varepsilon = 0$. Now more generally let us consider the flow $(\xi_t(x,\alpha))_t$ defined by

$$\frac{\partial}{\partial t}\xi_t(x,\alpha) = b(\xi_t(x,\alpha),\alpha), \quad \xi_0(x,\alpha) = x. \tag{3}$$

The condition [A2] ensures that the flow $(\xi_t(x,\alpha))_t$ exists and is smooth; [A3] means that the coefficients are smooth on a convex neighborhood of the deterministic limiting path, (X_t^0).

2.2. **The estimator and its properties.** In the following the function

$$\tilde{\delta}_n(x,\alpha) = \xi_{1/n}(x,\alpha) - x \tag{4}$$

plays a crucial role. The quantity $\tilde{\delta}_n(X_{t_{k-1}}, \alpha_0)$ is an approximation of $X_{t_k} - X_{t_{k-1}}$ as $\varepsilon \to 0$ and $n \to \infty$ (recall that $t_k = k/n$). Basic properties of $\tilde{\delta}_n(x,\alpha)$ are given in Section 4.3. We introduce a contrast function approximating the law of the observations in a way analogous to the approach in Kessler (1997) or Sørensen & Uchida (2003):

$$\tilde{U}_{\varepsilon,n}(\theta) = \left(\sum_{k=1}^{n} \{ \log \det \Xi_{k-1}(\beta) + \varepsilon^{-2} n \tilde{P}_k^*(\alpha) \Xi_{k-1}(\beta)^{-1} \tilde{P}_k(\alpha) \} \right) \mathbb{1}_{\{Z>0\}},$$

where

$$\tilde{P}_k(\alpha) = X_{t_k} - X_{t_{k-1}} - \tilde{\delta}_n(X_{t_{k-1}}, \alpha), \tag{5}$$

$$\Xi_k(\beta) = [\sigma\sigma^*](X_{t_k}, \beta) \tag{6}$$

and the random variable $Z = \inf_{k=0,\dots,n-1;\beta\in\overline{\Theta}_\beta} \det \Xi_k(\beta)$ is introduced to insure that $\tilde{U}_{\varepsilon,n}$ is well defined.

This contrast function is only explicit (due to $\tilde{\delta}_n$) if the flow (3) admits an explicit expression, which is not generally the case. However, useful explicit approximations are often available (see Section 2.3 for details). Therefore we denote by $\delta_n(x, \alpha)$ an approximation of the quantity $\tilde{\delta}_n(x, \alpha)$ and make the following assumptions on the quality of this approximation.

- **[C1]** The function δ_n is smooth on $\mathbb{R}^d \times \overline{\Theta}_\alpha$, and for any compact subset K of $\mathbb{R}^d$, there exists $c(K)$ such that

$$\sup_{x\in K, \alpha\in\overline{\Theta}_\alpha} \left| \delta_n(x, \alpha) - \tilde{\delta}_n(x, \alpha) \right| \le c(K)\varepsilon n^{-3/2}.$$

  Similar bounds hold for the first two derivatives of δ_n and $\tilde{\delta}_n$ with respect to the parameter α.
- **[C2]** The functions $n\delta_n$ are Lipschitz in the variable α, with a constant independent of n, on any compact subset of $\mathbb{R}^d \times \overline{\Theta}_\alpha$. The same holds for derivatives of any order with respect to α.

By Proposition 2 in Section 4.3, the choice $\delta_n = \tilde{\delta}_n$ satisfies these conditions under [A2] (of course, only [C2] needs verification in this case).

We can define a more general contrast function using the approximation δ_n instead of $\tilde{\delta}_n$:

$$U_{\varepsilon,n}(\theta) = \left(\sum_{k=1}^{n} \{ \log \det \Xi_{k-1}(\beta) + \varepsilon^{-2} n P_k^*(\alpha) \Xi_{k-1}(\beta)^{-1} P_k(\alpha) \} \right) \mathbb{1}_{\{Z>0\}}, \tag{7}$$

where now

$$P_k(\alpha) = X_{t_k} - X_{t_{k-1}} - \delta_n(X_{t_{k-1}}, \alpha). \tag{8}$$

Let $\hat{\theta}_{\varepsilon,n} = (\hat{\alpha}_{\varepsilon,n}, \hat{\beta}_{\varepsilon,n})$ be a minimum contrast estimator; i.e. a family of random variables satisfying

$$\hat{\theta}_{\varepsilon,n} = \underset{\theta \in \overline{\Theta}}{\operatorname{argmin}}\, U_{\varepsilon,n}(\theta). \tag{9}$$

The main result of the paper is the following.

Theorem 1. *Assume [A1]–[A4], [B], [C1]–[C2] and that $\theta_0 \in \Theta$ with the matrix $I(\theta_0)$ (given in (2)) being positive definite. Then*

$$\hat{\theta}_{\varepsilon,n} \to \theta_0$$

in $\mathbb{P}$-probability as $\varepsilon \to 0$ and $n \to \infty$. Further, we have the convergence

$$\begin{pmatrix} \varepsilon^{-1}(\hat{\alpha}_{\varepsilon,n} - \alpha_0) \\ \sqrt{n}(\hat{\beta}_{\varepsilon,n} - \beta_0) \end{pmatrix} \to \mathcal{N}(0, I(\theta_0)^{-1})$$

in distribution under $\mathbb{P}$ as $\varepsilon \to 0$ and $n \to \infty$.

Thus the estimator of the diffusion parameter, $\hat{\alpha}_{\varepsilon,n}$, and the estimator of the diffusion parameter, $\hat{\beta}_{\varepsilon,n}$, are asymptotically independent. The matrix $I_b(\theta_0)$ is equal to the Fisher information matrix for estimation of α_0 from the continuous time observation $(X_t)_{t\in[0,1]}$; see Kutoyants (1994). Hence the estimator of the drift parameter α is efficient. The asymptotic information matrix $I_\sigma(\theta_0)$, is related to the expression for the Fisher information matrix for estimation of the the diffusion parameter β from high frequency data with fixed ε found by Gobet (2002) in the same way that $I_b(\theta_0)$ is related to Gobet's expression for the Fisher information matrix for estimation of the drift parameter. This leads us to conjecture that our estimator for the diffusion parameter is efficient too.

The proof of Theorem 1 is given in Section 4.

2.3. **A possible choice of δ_n.** As stated previously the choice $\delta_n(x,\alpha) = \tilde{\delta}_n(x,\alpha) = \xi_{1/n}(x,\alpha) - x$ is possible under [A2], but typically this choice does not provide an explicit contrast function. Hence it can be useful to let δ_n equal the following approximation of $\tilde{\delta}_n$. Define the operator $\mathcal{L}^0_\alpha(f)(x) = \sum_{i=1}^d b^i(x,\alpha)\frac{\partial}{\partial x_i} f(x)$ for any differentiable function f, and set for any integer $v \geq 1$

$$\delta_n^v = \sum_{u=1}^{v} \left(\mathcal{L}^0_\alpha\right)^{u-1} (b(.,\alpha)) \frac{n^{-u}}{u!}.$$

For instance, we have the approximations $\delta_n^1(x,\alpha) = n^{-1}b(x,\alpha)$ and $\delta_n^2(x,\alpha) = n^{-1}b(x,\alpha) + (1/2)n^{-2}\sum_{i=1}^d b^i(x,\alpha)\frac{\partial}{\partial x_i} b(x,\alpha)$. The same approach was used by Uchida (2004) to approximate martingale estimating functions.

By the Assumption [A2] and (3), we easily prove that for any compact subset K of $\mathbb{R}^d$:

$$\sup_{x \in K, \alpha \in \overline{\Theta}_\alpha} \left| \delta_n^v(x,\alpha) - \tilde{\delta}_n(x,\alpha) \right| \leq c n^{-(v+1)}. \tag{10}$$

Moreover both δ_n^v and $\tilde{\delta}_n$ are smooth (see Proposition 2 for details) and one may show that (10) hold too for any derivatives with respect to α.

Hence by Assumption [B], δ_n^v satisfies [C1] provided that v is large enough, or ρ is small enough. Since [C2] is immediate, the choice $\delta_n = \delta_n^v$ is valid when $n^{-(v+1)} = O(\varepsilon n^{-3/2})$, i.e. if $(\varepsilon n^{v-1/2})^{-1}$ is bounded, or if $v - 1/2 \geq \rho$.

Remark 1. The choice $\delta_n(x,\alpha) = b(x,\alpha)n^{-1}$, which was considered in the paper by Sørensen & Uchida (2003), is sufficient by (10) when $(\varepsilon n^{1/2})^{-1}$ is bounded. Hence we find the set-up of Sørensen & Uchida (2003) as a particular case of the general framework considered here.

Remark 2. By inspecting the step 1 in the proof of consistency in Section 4.4.1 bellow, it can be seen that if the parameter β_0 is known and the contrast function (7) is used to estimate the parameter α, then the condition [B] is not necessary for consistency and might be replaced by $\varepsilon = \varepsilon_n \xrightarrow{n\to\infty} 0$.

2.4. **Martingale estimating functions.** A useful tool for estimating parameters in diffusion models is provided by quadratic martingale estimating functions, see Bibby, Jacobsen & Sørensen (2004). These estimators work well for low frequency data too. Here we briefly consider how our theory covers quadratic martingale estimating functions. Such an estimating function can be obtained by differentiation as the pseudo-score corresponding to a contrast function like (7), but with exact conditional moments instead of approximations. In particular, we must choose $P_k(\alpha) = X_{t_k} - m_n(X_{t_{k-1}}, \alpha)$ where $m_n(x,\alpha)$ is the conditional expectation of $X_{1/n}$ given $X_0 = x$ (under the assumption that this conditional expectation depends only on α). The corresponding choice $\delta_n(x,\alpha) = m_n(x,\alpha) - x$ is not always explicit, but it is interesting to note that this choice automatically satisfies the approximation condition [C1]. Indeed we have, by using (1) and then (3):

$$\begin{aligned}(11) \qquad \delta_n(x,\alpha) =& \mathbb{E}_\theta \left[\int_0^{1/n} b(X_s,\alpha)\mathrm{d}s \mid X_0 = x \right] \\ &= \tilde{\delta}_n(x,\alpha) + \mathbb{E}_\theta \left[\int_0^{1/n} \{ b(X_s,\alpha) - b(\xi_s(x,\alpha),\alpha)\} \mathrm{d}s \mid X_0 = x \right],\end{aligned}$$

and under smoothness conditions on b and σ we can prove (for details see condition [A3'] and Lemma 3 in Section 4):

$$(12) \qquad \sup_{0 \le s \le 1/n} \mathbb{E}_\theta \left[|X_s - \xi_s(x,\alpha)| \mid X_0 = x \right] \le c\varepsilon n^{-1/2}(1+|x|^c).$$

It follows that $\delta_n - \tilde{\delta}_n$ is of the order of magnitude required in [C1]. To see that the same order of approximation holds for the two first derivatives with respect to α is more delicate. Under smoothness assumptions on the coefficients, one can differentiate (11) with respect to α and then use that we have bounds analogous to (12) for $\left| \frac{\partial^i}{\partial \alpha^i} X_s - \frac{\partial^i}{\partial \alpha^i} \xi_s(x,\alpha) \right|$ with $i = 1, 2$. We omit the details here.

Remark 3. To obtain exactly a quadratic martingale estimating function, the contrast function must be defined by means of the exact second conditional moment $v(x,\beta) = cov_\theta(X_{1/n} \mid X_0 = x)$ instead of its short time approximation $\varepsilon^2 n^{-1}[\sigma\sigma^*](x,\beta)$ (assuming here that v only depends on β). The modification of the contrast (7) obtained by replacing the approximation $\Xi_{k-1}(\beta)$ by the exact moment $v(X_{t_{k-1}}, \beta)$ is not considered here because it is not important for high frequency data; however one could prove directly that Theorem 1 holds too for the modified contrast.

3. An example: a two factor model

As illustration we consider a two factor model $X = (Y, R)$ given by:

$$\mathrm{d}Y_t = (R_t + \mu_1)\mathrm{d}t + \varepsilon\kappa_1 \mathrm{d}W_t^1, \quad Y_0 = y_0 \in \mathbb{R}, \tag{13}$$

$$\mathrm{d}R_t = \mu_2(m - R_t)\mathrm{d}t + \varepsilon\kappa_2\sqrt{R_t}(\rho \mathrm{d}W_t^1 + (1-\rho^2)^{\frac{1}{2}}\mathrm{d}W_t^2), \quad R_0 = r_0 > 0, \tag{14}$$

with parameter $\theta = (\mu_1, \mu_2, m, \kappa_1^2, \kappa_2^2, \rho) \in \mathbb{R} \times (0, \infty)^4 \times (-1, 1)$.

The second component represents the short term rate while Y is the log price of some asset (see e.g. Longstaff & Schwartz 1995, Overhaus *et al* 2006). The parameter ρ allows correlation between innovation terms of the two factors. This diffusion satisfies [A1]–[A3], and [A4] holds if $r_0 \neq m_0$. The bi-dimension equation (3) is linear and has the solution:

$$\xi_t(y, r, \mu_1, \mu_2, m) = \begin{pmatrix} y + \mu_1 t + y + \frac{r-m}{\mu_2}(1 - e^{-\mu_2 t}) \\ m + (r - m)e^{-\mu_2 t} \end{pmatrix},$$

yielding to an explicit expression for the contrast function with the choice $\delta_n = \tilde{\delta}_n$. The information matrix $I(\theta)$ is explicit too, and one can check that $I(\theta_0)$ is invertible if $r_0 \neq m_0$ with:

$$I_b(\theta)^{-1} = (1-\rho^2)\begin{bmatrix} \frac{1}{\kappa_1^2} & 0 & 0 \\ 0 & \frac{-m\mu_2 - m\ln(q) + (m-r_0)(e^{-\mu_2}-1)}{\kappa_2^2\mu_2} & \frac{-\mu+\ln(q)}{\kappa_2^2} \\ 0 & \frac{-\mu+\ln(q)}{\kappa_2^2} & -\frac{\mu_2\ln(q)}{m\kappa_2^2} \end{bmatrix}^{-1}$$

where $q = r_0/[r_0 + m(e^{\mu_2} - 1)]$, and

$$I_\sigma(\theta)^{-1} = \begin{bmatrix} 2\kappa_1^4 & 2\rho^2\kappa_1^2\kappa_2^2 & \rho(1-\rho^2)\kappa_1^2 \\ 2\rho^2\kappa_1^2\kappa_2^2 & 2\kappa_2^4 & \rho(1-\rho^2)\kappa_2^2 \\ \rho(1-\rho^2)\kappa_1^2 & \rho(1-\rho^2)\kappa_2^2 & (1-\rho^2)^2 \end{bmatrix}$$

Let us remark that the asymptotic variance for the estimation of drift parameters decreases to zero as the correlation parameter ρ^2 increases to 1.

We explore the behaviour of the estimator for finite samples using Monte Carlo simulations. For each of the following situations and for each sample size a number of independent realizations of the process were simulated by means of the Euler scheme, and the estimators were calculated for each realization. Means and standard deviations of the simulated estimator values are reported in the Tables 1, 2 and 3. The Tables 1 and 2 are based on 400 replications, while 1000 replications were used for Table 3.

First the parameters are set to $\mu_1 = \mu_2 = m = \kappa_1^2 = \kappa_2^2 = 1$, $\rho = 0.3$ and $(y_0, r_0) = (0, 1.5)$. In Table 1 with $\varepsilon = 0.01$ the estimator gives good results, and it is very noticeable that the estimation of the drift remains good even if n is small. This is not surprising since for the model (14) our estimators for μ_2, m are the same one as considered in Sørensen (2000), and it is proved in this paper that for n fixed and $\varepsilon \to 0$ these estimators are consistent and asymptotically normal. That the estimating function obtained from our contrast function is a martingale is due to the fact that when the drift is linear, the conditional expectation of X_{t_k} given $X_{t_{k-1}}$ equals $\xi_{1/n}(X_{t_{k-1}}, \alpha)$. We made additional simulations that showed that, for the different choice $\delta_n(x, \alpha) = b(x, \alpha)n^{-1}$, the estimation is biased when n is too small. In Table 2 we give results for $\varepsilon = 0.1$, and it clearly appears that the behaviour of the estimators worsen.

To investigate how the estimator could perform on real financial data, we set the parameters to $\mu_1 = 5.7$, $m = 2$, $\kappa_2^2 = 450$, $\varepsilon = 0.1$, $x_0 = 3$ and $n = 300$. This set of parameter values corresponds to the estimates obtained in Chan *et al.* (1992) for 300 monthly observations over 25 years, if $r(t)/25$ is the annualized short time interest rate. We set arbitrarily $\rho = 0$, $\kappa_1^2 = 25$ and $\mu_1 = -0.125$ so that the risk premium of the asset is null. The estimator of μ_2 appears biased and μ_1 is clearly too small to be sharply estimated, but the other parameters are well estimated (see Table 3).

4. Proof of the main result

The details of the proof of the main result are split into 4 subsections. First, we introduce a set of more restrictive assumptions under which the proof will be easier and show that it is enough to prove the result under these assumptions. Then we present some crucial lemmas on the random variables $P_k(\alpha_0)$ (section 4.2). Third, we study the asymptotic behaviour of the functions $\tilde{\delta}_n(x, \alpha)$, and finally we study the contrast function $U_{\varepsilon,n}$ and prove Theorem 1 in section 4.4.

4.1. **A stronger set of assumptions.** Before proving Theorem 1, let us introduce a set of more restrictive assumptions under which the proof will be easier.

[A3'] For all $(x,\beta) \in \mathbb{R}^d \times \overline{\Theta}_\beta$, the matrix $\sigma\sigma^*(x,\beta)$ is positive definite. Moreover, the functions σ and $[\sigma\sigma^*]^{-1}$ (respectively b) are bounded and smooth with bounded derivatives of any order on $\mathbb{R}^p \times \overline{\Theta}_\beta$ (respectively $\mathbb{R}^p \times \overline{\Theta}_\alpha$).

[A5] $\sup_{0\le t\le 1} \left|X_t - X_t^0\right|$ tends to zero in $\mathbb{P}$-probability as $\varepsilon \to 0$.

[C1'] The function δ_n is smooth and there exists a constant c, such that

$$\sup_{x\in\mathbb{R}^d,\alpha\in\overline{\Theta}_\alpha} \left|\delta_n(x,\alpha) - \tilde{\delta}_n(x,\alpha)\right| \le c\varepsilon n^{-3/2},$$

and similar approximations hold for the first two derivatives of δ_n and $\tilde{\delta}_n$ with respect to the parameter α.

[C2'] The functions $n\delta_n$ are bounded and Lipschitz in the variable α, on $\mathbb{R}^d \times \overline{\Theta}_\alpha$, with constants independent of n. The same holds for derivatives of any order with respect to α.

By the following proposition, it is enough to prove Theorem 1 under [A1], [A2], [A3'], [A4], [A5], [B], [C1']–[C2'].

Proposition 1. *To prove that the conclusions of Theorem 1 hold under [A1]–[A4], [B] and [C1]–[C2], it is enough to prove that they hold under the stronger conditions [A1], [A2], [A3'], [A4], [A5], [B], [C1']–[C2'].*

Proof. Assume [A1]–[A4], [B] and [C1]–[C2]. By [A3] we can find two compact sets K, K' such that $K' \subset \overset{\circ}{K} \subset K \subset \mathcal{U}$ with $X_t^0 \in \overset{\circ}{K'}$, $\forall t \in [0,1]$. Now, by Lemma 6 below, smooth modifications, b' and σ', of b and σ exist such that:

1. $\forall x \in K, \forall \alpha \in \overline{\Theta}_\alpha$, $b'(x,\alpha) = b(x,\alpha)$ and b' has compact support
2. $\forall x \in K, \forall \beta \in \overline{\Theta}_\beta$, $\sigma'(x,\beta) = \sigma(x,\beta)$, σ' is constant except on some compact set, and $\inf_{x\in\mathbb{R}^d,\beta\in\overline{\Theta}_\beta} \det \sigma'\sigma'^*(x,\beta) > 0$.

Clearly these new coefficients satisfy the condition [A3']. Define $X'^\varepsilon = X'$ as a solution of (1) with the coefficients σ and b replaced by $\sigma'(.,\beta_0)$ and $b'(.,\alpha_0)$. Then by the uniqueness of solutions of stochastic differential equations (see for instance Gihman & Skorohod (1972), p.44) we have, $\mathbb{P}(X = X'; X'_t \in K',\ \forall t \in [0,1]) = P(X'_t \in K',\ \forall t \in [0,1])$. Using that by theorem 1.2 at p.45 of Freidlin

& Wentzell (1998), $\sup_{0\leq s\leq 1}\left|X'^{\varepsilon}_s - X'^0_s\right| \xrightarrow[\mathbb{P}]{\varepsilon\to 0} 0$, it follows that for any $r > 0$, $\mathbb{P}(\sup_{0\leq s\leq 1}\left|X^{\varepsilon}_s - X^0_s\right| \geq r) \leq \mathbb{P}(\sup_{0\leq s\leq 1}\left|X'^{\varepsilon}_s - X'^0_s\right| \geq r) + \mathbb{P}(\exists t\in[0,1],\ X'_t \notin K')$ converges to zero as $\varepsilon \to 0$. Hence condition [A5] holds for the diffusion X'^{ε}, and in turn we deduce that $\mathbb{P}(X = X'; X'_t \in K',\ \forall t\in[0,1])$ tends to one as $\varepsilon \to 0$.

Consider now the flow defined by (3) with the coefficient b replaced by b', and the associated quantity defined by (4), which we denote by $\tilde{\delta}'_n(x,\alpha)$. It is easy to see that for n sufficiently large, $\tilde{\delta}'_n(x,\alpha) = \tilde{\delta}_n(x,\alpha)$ for all $\alpha\in\overline{\Theta}_\alpha$ and all x in some compact set K'' with $K' \subset \overset{\circ}{K''} \subset K'' \subset K$. Next, we modify the approximation δ_n accordingly by defining $\delta'_n(x,\alpha) = \delta_n(x,\alpha)\psi(x) + \tilde{\delta}'_n(x,\alpha)(1-\psi(x))$, where ψ is a non-negative smooth function that is equal to 1 on K' and vanishes outside K''. Then for all $(x,\alpha)\in\mathbb{R}^d\times\overline{\Theta}_\alpha$:

$$\delta'_n(x,\alpha) - \tilde{\delta}'_n(x,\alpha) = (\delta_n(x,\alpha) - \tilde{\delta}'_n(x,\alpha))\psi(x) = (\delta_n(x,\alpha) - \tilde{\delta}_n(x,\alpha))\psi(x).$$

Thus condition [C1] for $(\delta_n, \tilde{\delta}_n)$ implies [C1'] for $(\delta'_n, \tilde{\delta}'_n)$ with $c = c(K')\|\psi\|_\infty$. Moreover, $n\delta'_n$ vanishes outside some compact set, so [C2'] follows from [C2].

By construction, we now have two statistical problems on the same probability space. The one indicated by a "prime" satisfies [A1], [A2], [A3'], [A4], [A5], [B], [C1']–[C2']. We assume that theorem 1 has been proved under these conditions, so the conclusions of theorem 1 hold for the "prime"-model. On the event $\{X = X'; X'_t \in K',\ \forall t\in[0,1]\}$, the contrast functions of the two statistical problems coincide, and asymptotically this event has probability equal to one. Hence the conclusions of theorem 1 hold for the initial statistical problem too. $\square$

4.2. **Preliminary lemmas.** We introduce the σ-field $\mathcal{G}^n_k = \sigma(W_s, s\leq t_k)$. Let us denote by $\mathrm{R}(a,x)$ any function defined on $\mathbb{R}^d$ such that there exists $c\geq 0$, with $|\mathrm{R}(a,x)| \leq ac(1+|x|^c)$ for all x. Moreover, denote by $\mathcal{C}^\infty_\uparrow(\mathbb{R}^d\times\overline{\Theta},\mathbb{R})$ the set of smooth functions f on $\mathbb{R}^d\times\overline{\Theta}$ for which the derivatives of any order have at most polynomial growth: $\sup_{\theta\in\overline{\Theta}}\left|\frac{\partial^r}{\partial\theta^r}\frac{\partial^s}{\partial x^s}f(x,\theta)\right| \leq c(1+|x|^c)$ while $\mathcal{C}^\infty_\uparrow(\mathbb{R}^d,\mathbb{R})$ denotes the subset of $\mathcal{C}^\infty_\uparrow(\mathbb{R}^d\times\overline{\Theta},\mathbb{R})$ consisting of the functions only dependent on x.

Finally, we denote by $\mathcal{L}^\varepsilon$ the generator of the diffusion X: if f is smooth,

$$\mathcal{L}^\varepsilon(f)(x) = \sum_{i=1}^d b^i(x,\alpha_0)\frac{\partial}{\partial x_i}f(x) + 1/2\varepsilon^2\sum_{i,j=1}^d[\sigma\sigma^*]^{i,j}(x,\beta_0)\frac{\partial^2}{\partial x_i\partial x_j}f(x),$$

and set

$$\mathcal{L}^0(f)(x) = \sum_{i=1}^{d} b^i(x,\alpha_0)\frac{\partial}{\partial x_i} f(x).$$

Lemma 1. *Assume [A1], [A2], [A3'], [B] and [C1'], then*

1. $$\left|\mathbb{E}\left[P_k^i(\alpha_0) \mid \mathcal{G}_{k-1}^n\right]\right| = R(\varepsilon^2 n^{-1}, X_{t_{k-1}}) + R(\varepsilon n^{-3/2}, X_{t_{k-1}}).$$
2. $$\mathbb{E}\left[P_k^{i_1}(\alpha_0) P_k^{i_2}(\alpha_0) \mid \mathcal{G}_{k-1}^n\right] = \frac{\varepsilon^2}{n}\Xi_{k-1}^{i_1,i_2}(\beta_0) + R(\varepsilon^2 n^{-2}, X_{t_{k-1}}).$$
3. $$\left|\mathbb{E}\left[P_k^{i_1}(\alpha_0) P_k^{i_2}(\alpha_0) P_k^{i_3}(\alpha_0) \mid \mathcal{G}_{k-1}^n\right]\right| = R(\varepsilon^3 n^{-2}, X_{t_{k-1}}).$$
4. $$\begin{aligned}\mathbb{E}\left[\Pi_{j=1}^4 P_k^{i_j}(\alpha_0) \mid \mathcal{G}_{k-1}^n\right] = {}& \tfrac{\varepsilon^4}{n^2}\left\{\Xi_{k-1}^{i_1,i_2}\Xi_{k-1}^{i_3,i_4} + \Xi_{k-1}^{i_1,i_3}\Xi_{k-1}^{i_2,i_4} + \Xi_{k-1}^{i_1,i_4}\Xi_{k-1}^{i_2,i_3}\right\} \\ &+ R(\varepsilon^4 n^{-5/2}, X_{t_{k-1}}).\end{aligned}$$
5. *For all* $M \geq 2$, $\mathbb{E}\left[\left|P_k^i(\alpha_0)\right|^M \mid \mathcal{G}_{k-1}^n\right] = R(\varepsilon^M n^{-M/2}, X_{t_{k-1}})$.

Proof. First, remark that we only need to establish these inequalities with $\tilde{P}_k(\alpha_0)$ instead of $P_k(\alpha_0)$. Indeed, we see that $P_k(\alpha_0) - \tilde{P}_k(\alpha_0) = \delta_n(X_{t_{k-1}},\alpha_0) - \tilde{\delta}_n(X_{t_{k-1}},\alpha_0)$ is properly bounded by Assumption [C1'] to enable the substitutions in 1)–5).

Set $\phi_k(y) = y - X_{t_{k-1}} - \tilde{\delta}_n(X_{t_{k-1}},\alpha_0)$. Using the Markov property of X and applying iteratively the Ito formula, we have for any integer $v \geq 1$

$$\begin{aligned}\mathbb{E}\left[\tilde{P}_k^i(\alpha_0) \mid \mathcal{G}_{k-1}^n\right] &= \mathbb{E}\left[\phi_k^i(X_{t_k}) \mid X_{t_{k-1}}\right] = \textstyle\sum_{u=0}^{v-1} \left(\mathcal{L}^\varepsilon\right)^u (\phi_k^i)(X_{t_{k-1}})\frac{n^{-u}}{u!} \\ &+ \textstyle\int_{t_{k-1}}^{t_k}\int_{t_{k-1}}^{s_1}\dots\int_{t_{k-1}}^{s_{v-1}} \mathbb{E}\left[\left(\mathcal{L}^\varepsilon\right)^v (\phi_k^i)(X_{s_v}) \mid X_{t_{k-1}}\right] \mathrm{d}s_v \dots \mathrm{d}s_1.\end{aligned} \tag{15}$$

Now by (4), $0 = \xi_{1/n}(X_{t_{k-1}},\alpha_0) - X_{t_{k-1}} - \tilde{\delta}_n(X_{t_{k-1}},\alpha_0) = \phi_k^i(\xi_{1/n}(X_{t_{k-1}},\alpha_0))$. Hence, using a Taylor expansion, we obtain

$$\begin{aligned}0 = \sum_{u=0}^{v-1} \left(\mathcal{L}^0\right)^u (\phi_k^i)(X_{t_{k-1}})\frac{n^{-u}}{u!} + \\ \int_0^{1/n}\int_0^{s_1}\dots\int_0^{s_{v-1}} \left(\mathcal{L}^0\right)^v (\phi_k^i)(\xi_{s_v}(X_{t_{k-1}},\alpha_0))\mathrm{d}s_v\dots\mathrm{d}s_1.\end{aligned} \tag{16}$$

Using that all derivatives of order ≥ 1 of ϕ_k are bounded and assumption $[A3']$, we see that the multiple integrals in (15)–(16) are bounded by cn^{-v}. Hence, by (15)–(16):

$$\mathbb{E}\left[\tilde{P}_k^i(\alpha_0) \mid \mathcal{G}_{k-1}^n\right] = \sum_{u=0}^{v-1} \left\{(\mathcal{L}^\varepsilon)^u - \left(\mathcal{L}^0\right)^u\right\} (\phi_k^i)(X_{t_{k-1}})\frac{n^{-u}}{u!} + \mathrm{R}(n^{-v}, X_{t_{k-1}}).$$

Now, Lemma 2 below and an appropriate choice of v (in view of [B]) give

$$\left|\mathbb{E}\left[\tilde{P}_k^i(\alpha_0) \mid \mathcal{G}_{k-1}^n\right]\right| = \mathrm{R}(\varepsilon^2 n^{-1}, X_{t_{k-1}}) + \mathrm{R}(n^{-v}, X_{t_{k-1}}) = \mathrm{R}(\varepsilon^2 n^{-1}, X_{t_{k-1}}).$$

To prove the second part of the theorem, we proceed analogously: now set $\phi_{k,i,j} = \{y^i - X_{t_{k-1}}^i - \tilde{\delta}_n^i(X_{t_{k-1}}, \alpha_0)\}\{y^j - X_{t_{k-1}}^j - \tilde{\delta}_n^j(X_{t_{k-1}}, \alpha_0)\}$. Then we have in analogy with (15):

$$\mathbb{E}\left[\tilde{P}_k^i(\alpha_0)\tilde{P}_k^j(\alpha_0) \mid \mathcal{G}_{k-1}^n\right] = \mathbb{E}\left[\phi_{k,i,j}(X_{t_k}) \mid X_{t_{k-1}}\right] = \sum_{u=0}^{v-1} (\mathcal{L}^\varepsilon)^u (\phi_{k,i,j})(X_{t_{k-1}})\frac{n^{-u}}{u!} + \mathrm{R}(n^{-v}, X_{t_{k-1}}).$$

Using that $\phi_{k,i,j}(\xi_{1/n}(X_{t_{k-1}}, \alpha_0)) = 0$, we obtain by subtracting a Taylor expansion similar to (16),

$$\mathbb{E}\left[\tilde{P}_k^i(\alpha_0)\tilde{P}_k^j(\alpha_0) \mid \mathcal{G}_{k-1}^n\right] = \left\{(\mathcal{L}^\varepsilon) - \left(\mathcal{L}^0\right)\right\} (\phi_{k,i,j})(X_{t_{k-1}})n^{-1} + \sum_{u=2}^{v-1} \left\{(\mathcal{L}^\varepsilon)^u - \left(\mathcal{L}^0\right)^u\right\} (\phi_{k,i,j})(X_{t_{k-1}})\frac{n^{-u}}{u!} + \mathrm{R}(n^{-v}, X_{t_{k-1}}).$$

Now simple computations shows that $\left\{(\mathcal{L}^\varepsilon) - \left(\mathcal{L}^0\right)\right\} (\phi_{k,i,j}) = \varepsilon^2 [\sigma\sigma^*]^{i,j}$ and 2) then follows by Lemma 2.

To prove 3)–5), we first show the following expansion

$$\tilde{P}_k(\alpha_0) = \varepsilon\sigma(X_{t_{k-1}}, \beta_0)(W_{t_k} - W_{t_{k-1}}) + E_k \tag{17}$$

where the remainder term E_k satisfies that for all $M \geq 2$, $\mathbb{E}\left[|E_k|^M \mid \mathcal{G}_{k-1}^n\right] \leq c(M)\varepsilon^M n^{-M}$. By (1) and (5), we can write $\tilde{P}_k(\alpha_0) = \varepsilon\sigma(X_{t_{k-1}}, \beta_0)(W_{t_k} - W_{t_{k-1}}) +$

$E_{k,1} + E_{k,2}$ with:

$$E_{k,1} = \varepsilon \int_{t_{k-1}}^{t_k} \left\{\sigma(X_s, \beta_0) - \sigma(X_{t_{k-1}}, \beta_0)\right\} \mathrm{d}W_s,$$
$$E_{k,2} = \tilde{\delta}_n(X_{t_{k-1}}, \alpha_0) - \int_{t_{k-1}}^{t_k} b(X_s, \alpha_0)\mathrm{d}s.$$

Using the Burkhölder–Davis–Gundy inequality and then Jensen's inequality, we have

$$\begin{aligned} &\mathbb{E}\left[|E_{k,1}|^M \mid \mathcal{G}_{k-1}^n\right] \\ &\leq \; c(M)\varepsilon^M \mathbb{E}\left[\left|\int_{t_{k-1}}^{t_k} \left|\sigma(X_s, \beta_0) - \sigma(X_{t_{k-1}}, \beta_0)\right|^2 \mathrm{d}s\right|^{\frac{M}{2}} \mid \mathcal{G}_{k-1}^n\right] \\ &\leq \; c(M)\varepsilon^M n^{-M/2+1} \int_{t_{k-1}}^{t_k} \mathbb{E}\left[\left|\sigma(X_s, \beta_0) - \sigma(X_{t_{k-1}}, \beta_0)\right|^M \mid \mathcal{G}_{k-1}^n\right] \mathrm{d}s \end{aligned}$$

Using the Ito formula and [A3'], we obtain $\mathbb{E}\left[\left|\sigma(X_s, \beta_0) - \sigma(X_{t_{k-1}}, \beta_0)\right|^M \mid \mathcal{G}_{k-1}^n\right]$ $\leq c\,|s - t_{k-1}|^{M/2}$, and deduce that $\mathbb{E}\left[|E_{k,1}|^M \mid \mathcal{G}_{k-1}^n\right] \leq c(M)\varepsilon^M n^{-M}$.

To evaluate $E_{k,2}$, remark that by (3)–(4), we can write

$$E_{k,2} = \int_0^{1/n} b(\xi_s(X_{t_{k-1}}, \alpha_0), \alpha_0)\mathrm{d}s - \int_{t_{k-1}}^{t_k} b(X_s, \alpha_0)\mathrm{d}s.$$

Then, the function b being Lipschitz, we deduce

$$\mathbb{E}\left[|E_{k,2}|^M \mid \mathcal{G}_{k-1}^n\right] \leq cn^{-M+1} \int_0^{1/n} \mathbb{E}\left[\left|\xi_s(X_{t_{k-1}}, \alpha_0) - X_{t_{k-1}+s}\right|^M \mid \mathcal{G}_{k-1}^n\right] \mathrm{d}s.$$

Direct application of Lemma 3 gives $\mathbb{E}\left[|E_{k,2}|^M \mid \mathcal{G}_{k-1}^n\right] \leq \mathrm{R}(\varepsilon^M n^{-(3/2)M}, X_{t_{k-1}})$ and (17) follows.

Now, 3)–4) are deduced from (17), using the expressions for the moments of order ≤ 4 of Gaussian variables with covariance matrix given by (6), and by application of the Cauchy–Schwarz inequality to the remainder terms. Finally, 5) is immediate by (17). □

Remark that point 5 in Lemma 1 still holds true without the condition [B].

Lemma 2. *Assume [A2], [A3'] and let $f \in \mathcal{C}^{\infty}_{\uparrow}(\mathbb{R}^d, \mathbb{R})$. Then,*

$$\left(\mathcal{L}^{\varepsilon}\right)^0(f) - \left(\mathcal{L}^0\right)^0(f) = 0$$
$$\forall u \geq 1, \quad \exists c(u), \forall x, \left|\left(\mathcal{L}^{\varepsilon}\right)^u(f)(x) - \left(\mathcal{L}^0\right)^u(f)(x)\right| \leq \varepsilon^2 c(u)(1 + |x|^{c(u)})$$

Proof. The first property is immediate since $\left(\mathcal{L}^{\varepsilon}\right)^0 = \left(\mathcal{L}^0\right)^0 = \mathcal{I}d$. The second one follows easily because, by induction on u, that we have,

$$\forall u \geq 1, \quad \left(\mathcal{L}^{\varepsilon}\right)^u(f) = \left(\mathcal{L}^0\right)^u(f) + \varepsilon^2 g_u$$

where g_u is some element of $\mathcal{C}^{\infty}_{\uparrow}(\mathbb{R}^d, \mathbb{R})$. □

Lemma 3. *Assume [A1], [A2], [A3']. Then for all $M \geq 1$, for all $k \in \{1, \ldots, n\}$ and $t \in [0, 1]$,*

$$\mathbb{E}\left[\left|X_t - \xi_{t-t_{k-1}}(X_{t_{k-1}}, \alpha_0)\right|^M \mid \mathcal{G}^n_{k-1}\right] \leq R(\varepsilon^M \left|t - t_{k-1}\right|^{M/2}, X_{t_{k-1}}).$$

Proof. The result is obtained in the proof of Theorem 1.2 p.45–47 of Freidlin & Wentzell (1998) for the case $M = 2$. The proof extends classically to any $M \geq 1$. □

Lemma 4. *Assume [A1], [A2], [A3'], [A5], [C1'] and let $f_n, f \in \mathcal{C}^{\infty}_{\uparrow}\left(\mathbb{R}^d \times \overline{\Theta}, \mathbb{R}\right)$ such that the sequence f_n converges uniformly on any compact subset of $\mathbb{R}^d \times \overline{\Theta}$ to f. Further, assume that the two following conditions hold for some constant c:*

$$\forall \theta, n, x, |f_n(x, \theta)| \leq c(1 + |x|^c) \tag{18}$$
$$\forall \theta, \theta', n, x, \left|f_n(x, \theta) - f_n(x, \theta')\right| \leq \left|\theta - \theta'\right| c(1 + |x|^c). \tag{19}$$

Then,

1) $n^{-1}\sum_{k=1}^n f_n(X_{t_{k-1}}, \theta) \xrightarrow{n \to \infty, \varepsilon \to 0} \int_0^1 f(X^0_s, \theta)$ uniformly in $\mathbb{P}$-probability.

2) Under the additional condition [B], the following sequence is bounded in $\mathbb{P}$-probability

$$\left(\sup_{\theta \in \Theta} \varepsilon^{-1} \sum_{k=1}^n f_n(X_{t_{k-1}}, \theta) P_k(\alpha_0)\right)_{n \geq 1}.$$

Proof. 1) Using [A2] and [A5], X takes values on some compact set with any probability arbitrary closed to 1. Hence the uniform convergence property for f_n implies that $\sup_{\theta \in \overline{\Theta}}\left|n^{-1}\sum_{k=1}^n f_n(X_{t_{k-1}}, \theta) - n^{-1}\sum_{k=1}^n f(X_{t_{k-1}}, \theta)\right|$ converges to 0 in $\mathbb{P}$-probability. Then the convergence of $n^{-1}\sum_{k=1}^{n} f(X_{t_{k-1}},\theta)$ is obtained as in Sørensen & Uchida (2003).

2) We set $C_n(\theta) = \varepsilon^{-1}\sum_{k=1}^{n} f_n(X_{t_{k-1}},\theta)\mathbb{E}\left[P_k(\alpha_0) \mid \mathcal{G}_{k-1}^n\right]$ and $M_n(\theta) = \sum_{k=1}^{n} \varpi_{k,n}(\theta)$ with $\varpi_{k,n}(\theta) = \varepsilon^{-1} f_n(X_{t_{k-1}},\theta)\left\{P_k(\alpha_0) - \mathbb{E}\left[P_k(\alpha_0) \mid \mathcal{G}_{k-1}^n\right]\right\}$. First, using (18) and Lemma 1 1), we have $|C_n(\theta)| \le (\varepsilon n^{-1} + n^{-3/2})\sum_{k=1}^{n} c(1+\left|X_{t_{k-1}}\right|^c)$ which converges to 0 in $\mathbb{P}$-probability.

Finally, it remains to prove the tightness of $M_n(.)$. For this it is sufficient to show that (see Theorem 20 in Appendix I of Ibragimov & Has'minskii (1981) or Lemma 3.1 of Yoshida (1992)):

$$\mathbb{E}\left[\left|\sum_{k=1}^{n} \varpi_{k,n}(\theta)\right|^{2l}\right] \le c \tag{20}$$

$$\mathbb{E}\left[\left|\sum_{k=1}^{n} \varpi_{k,n}(\theta_1) - \sum_{k=1}^{n} \varpi_{k,n}(\theta_2)\right|^{2l}\right] \le c\left|\theta_1 - \theta_2\right|^{2l} \tag{21}$$

for any $\theta, \theta_1, \theta_2 \in \overline{\Theta}$ and $2l$ an even integer greater than the dimension $p+q$ of the parameter space Θ.

We only give details for the proof of (21) since the other one may be proved similarly. Using Rosenthal's inequality for martingales (see Burkhölder (1973), Hall & Heyde (1980)), we have:

$$\mathbb{E}\left[\left|\sum_{k=1}^{n} \varpi_{k,n}(\theta_1) - \varpi_{k,n}(\theta_2)\right|^{2l}\right] \le \tag{22}$$

$$c\mathbb{E}\left[\left(\sum_{k=1}^{n} \mathbb{E}\left[\left|\varpi_{k,n}(\theta_1) - \varpi_{k,n}(\theta_2)\right|^2 \mid \mathcal{G}_{k-1}^n\right]\right)^l + \left(\sum_{k=1}^{n} \left|\varpi_{k,n}(\theta_1) - \varpi_{k,n}(\theta_2)\right|^{2l}\right)\right]$$

Using Lemma 1 5) and the Lipschitz condition (19), we can prove that

$$\forall M, \quad \mathbb{E}\left[\left|\varpi_{k,n}(\theta_1) - \varpi_{k,n}(\theta_2)\right|^M \mid \mathcal{G}_{k-1}^n\right] \le c n^{-M/2}\left|\theta_1 - \theta_2\right|^M (1 + \left|X_{t_{k-1}}\right|^c).$$

Using this bound, in (22) with $M = 2$ and $M = 2l$, and the fact that X_t has finite moments, gives (21). □

Lemma 5. *Assume [A1], [A2], [A3'], [A5], [B], [C1'] and let* $f \in \mathcal{C}^{\infty}_{\uparrow}\left(\mathbb{R}^d \times \overline{\Theta}, \mathbb{R}\right)$. *Then we have the convergence, uniform with respect to* θ, *in* $\mathbb{P}$*-probability:*

$$\varepsilon^{-2} \sum_{k=1}^{n} f(X_{t_{k-1}}, \theta) P_k^i(\alpha_0) P_k^j(\alpha_0) \xrightarrow{n\to\infty, \varepsilon\to 0} \int_0^1 f(X_s^0, \theta)[\sigma\sigma^*]^{i,j}(X_s^0, \beta_0)\, ds.$$

Proof. We follow the scheme of proof of Lemma 3 in Sørensen & Uchida (2003) (see Lemma 9 from Genon–Catalot & Jacod (1993) too). It is sufficient to prove the three following facts:

$$\sum_{k=1}^{n} \mathbb{E}\left[\varepsilon^{-2} f(X_{t_{k-1}}, \theta) P_k^i(\alpha_0) P_k^j(\alpha_0) \mid \mathcal{G}_{k-1}^n\right] \xrightarrow{\mathbb{P}} \int_0^1 f(X_s^0, \theta)[\sigma\sigma^*]^{i,j}(X_s^0, \beta_0) \mathrm{d}s,$$

$$\sum_{k=1}^{n} \mathbb{E}\left[\varepsilon^{-4} f^2(X_{t_{k-1}}, \theta) \left(P_k^i(\alpha_0) P_k^j(\alpha_0)\right)^2 \mid \mathcal{G}_{k-1}^n\right] \xrightarrow{\mathbb{P}} 0,$$

$$\sup_{\varepsilon, n} \mathbb{E}\left[\sup_{\theta} \left| \frac{\partial}{\partial\theta} \sum_{k=1}^{n} \varepsilon^{-2} f(X_{t_{k-1}}, \theta) P_k^i(\alpha_0) P_k^j(\alpha_0) \right|\right] < \infty.$$

The first point is shown by using first Lemma 1 2) and then Lemma 4 1). The second and third points follows from Lemma 1 5). □

We end this section by the following lemma used in the proof of Proposition 1.

Lemma 6. *Let* K *and* $\mathcal{U}$ *be as in the proof of Proposition 1. Then there exist smooth functions* b' *and* σ' *such that:*

1. $\forall x \in K, \forall \alpha \in \overline{\Theta}_\alpha$, $b'(x,\alpha) = b(x,\alpha)$ *and* b' *has compact support*
2. $\forall x \in K, \forall \beta \in \overline{\Theta}_\beta$, $\sigma'(x,\beta) = \sigma(x,\beta)$; $\inf_{x\in\mathbb{R}^d, \beta\in\overline{\Theta}_\beta} \det \sigma'\sigma'^*(x,\beta) > 0$ *and* σ' *is constant except on some compact set.*

Proof. The construction of b' is immediate by multiplication of b by a smooth function $\psi_K(x)$ equal to 1 on K with compact support. For 2), using that $\mathcal{U} \times \overline{\Theta}_\beta \subset \mathbb{R}^d \times \overline{\Theta}_\beta$ are two convex sets, there exists a smooth retraction $\phi(t,x,\beta) : [0,1] \times \mathbb{R}^d \times \overline{\Theta}_\beta$ such that $\phi(1,x,\beta) = (x,\beta)$ and $\phi(0,x,\beta) = (\overline{x}, \overline{\theta})$ is some fixed element of $\mathcal{U} \times \overline{\Theta}_\beta$ and for all $t \in [0,1]$, $\phi(t, \mathcal{U} \times \overline{\Theta}_\beta) \subset \mathcal{U} \times \overline{\Theta}_\beta$. Let $\psi_{K,\mathcal{U}}$ be a smooth function with compact support on $\mathbb{R}^d$, equal to 1 on K, and vanishing on $\mathbb{R}^d - \mathcal{U}$. We set $\sigma'(x,\beta) = \sigma(\phi(\psi_{K,\mathcal{U}}(x), x, \beta))$, which by [A3] satisfies 2). □

4.3. **Properties of $\tilde{\delta}_n$.** First, we compare $n\tilde{\delta}_n(x,\alpha)$ with $b(x,\alpha)$.

Proposition 2. *1) Assume [A2], then the flow $(\xi_t(x,\alpha))_{t\geq 0}$ is well defined and is smooth on $\mathbb{R}^d \times \overline{\Theta}_\alpha$. Further the functions $n\tilde{\delta}_n$ and all their derivatives with respect to α are bounded independently of n on compact subsets of $\mathbb{R}^d \times \overline{\Theta}_\alpha$.*

2) Assume [A2] and [A3'], then the sequence $n\tilde{\delta}_n$ converges to b uniformly:

$$\sup_{x\in\mathbb{R}^d,\alpha\in\overline{\Theta}_\alpha} \left| n\tilde{\delta}_n(x,\alpha) - b(x,\alpha) \right| \leq cn^{-1},$$

and a similar bound holds for all derivatives of order ≤ 2 with respect to α of $\tilde{\delta}_n$ and b.

Proof. Using well known results on the dependence of solution of an ordinary differential equation on a parameter (see for instance Walter (1998) p.151), all the derivatives of $\xi(x,\alpha)_t$ with respect to α and x exist and they satisfy the differential equation obtain by formal differentiation of (3). Since, using (3), we have

$$n\tilde{\delta}_n(x,\alpha) = n\int_0^{\frac{1}{n}} b(\xi(x,\alpha)_s,\alpha)\mathrm{d}s, \tag{23}$$

point 1) follows.

Now under [A3'], we have $\sup_{x,\alpha}|\xi(x,\alpha)_s - x| \leq cs$, and we deduce from (23) that

$$\sup_{x\in\mathbb{R}^d,\alpha\in\Theta_\alpha} \left| n\tilde{\delta}_n(x,\alpha) - b(x,\alpha) \right| \leq cn^{-1}.$$

To show that an analogous bound hold for the derivatives, we write

$$\frac{\partial(n\tilde{\delta}_n)}{\partial\alpha}(x,\alpha) = n\int_0^{\frac{1}{n}} \frac{\partial b}{\partial\alpha}(\xi(x,\alpha)_s,\alpha)\mathrm{d}s + n\int_0^{\frac{1}{n}} \frac{\partial b}{\partial x}(\xi(x,\alpha)_s,\alpha)\frac{\partial\xi}{\partial\alpha}(x,\alpha)_s\mathrm{d}s.$$

Then using that $\frac{\partial\xi}{\partial\alpha}(x,\alpha)_0 = 0$, and hence that $\left|\frac{\partial\xi}{\partial\alpha}(x,\alpha)_s\right| \leq cs$ we deduce

$$\sup_{x\in\mathbb{R}^d,\alpha\in\overline{\Theta}_\alpha} \left| \frac{\partial(n\tilde{\delta}_n)}{\partial\alpha}(x,\alpha) - \frac{\partial b}{\partial\alpha}(x,\alpha) \right| \leq cn^{-1}.$$

By differentiating (23) twice, we deduce similarly the approximation for the second order derivatives. $\square$

Remark that by 1) in the previous proposition, we see that the choice $\delta_n = \tilde{\delta_n}$ satisfies the condition [C2]. Note too that if we choose δ_n so that it satisfies [C1'], then, by 2) of the above proposition, we can deduce the following lemma.

Lemma 7. *Assume [A2], [A3'], [C1'] then the sequence $n\delta_n$ converges uniformly to b: $\sup_{x\in\mathbb{R}^d,\alpha\in\Theta_\alpha} |n\delta_n(x,\alpha) - b(x,\alpha)| \leq c(n^{-1} + \varepsilon n^{-1/2})$, and a similar bound holds for all derivatives of order ≤ 2 with respect to α.*

4.4. **Proof of Theorem 1.** Let us introduce the following quantities that we will use in the proof:

$$U_1(\alpha,\alpha_0,\beta) = \int_0^1 (b(X_s^0,\alpha) - b(X_s^0,\alpha_0))^*[\sigma\sigma^*]^{-1}(X_s^0,\beta)(b(X_s^0,\alpha) - b(X_s^0,\alpha_0))\mathrm{d}s$$

$$\begin{aligned} U_2(\beta,\beta_0) = & \int_0^1 \log\det[\sigma\sigma^*](X_s^0,\beta)[\sigma\sigma^*]^{-1}(X_s^0,\beta_0)\mathrm{d}s \\ & + \int_0^1 \mathrm{tr}\left[[\sigma\sigma^*](X_s^0,\beta)[\sigma\sigma^*]^{-1}(X_s^0,\beta_0)\right]\mathrm{d}s - d. \end{aligned}$$

We rewrite the expression of the contrast function (7) in a more convenient form (and use the fact that under [A3'] we can suppress $\mathbb{1}_{\{Z>0\}}$):

$$\begin{aligned} U_{\varepsilon,n}(\theta) = & \sum_{k=1}^n \log\det\Xi_{k-1}(\beta) + \varepsilon^{-2} n \sum_{k=1}^n P_k^*(\alpha_0)\Xi_{k-1}(\beta)^{-1}P_k(\alpha_0) \\ & + \varepsilon^{-2} n \sum_{k=1}^n (\delta_n(X_{t_{k-1}},\alpha_0) - \delta_n(X_{t_{k-1}},\alpha))^*\Xi_{k-1}^{-1}(\beta)(\delta_n(X_{t_{k-1}},\alpha_0) \\ & \qquad - \delta_n(X_{t_{k-1}},\alpha)) \\ & + 2\varepsilon^{-2} n \sum_{k=1}^n (\delta_n(X_{t_{k-1}},\alpha_0) - \delta_n(X_{t_{k-1}},\alpha))^*\Xi_{k-1}^{-1}(\beta)P_k(\alpha_0). \end{aligned}$$

4.4.1. *Consistency of the estimator.* $1^{\mathbf{st}}$ **step.** We prove the consistency for the drift parameter. For this, repeating the arguments of Theorem 1 in Sørensen & Uchida (2003), it is sufficient to show the following convergence uniformly with respect to (α,β):

$$\varepsilon^2\{U_{\varepsilon,n}(\alpha,\beta) - U_{\varepsilon,n}(\alpha_0,\beta)\} \xrightarrow[\mathbb{P}]{n\to\infty,\varepsilon\to 0} U_1(\alpha,\alpha_0,\beta).$$

By the expression of the contrast function, we have

$$\varepsilon^2\{U_{\varepsilon,n}(\alpha,\beta)-U_{\varepsilon,n}(\alpha_0,\beta)\} =$$
$$n^{-1}\sum_{k=1}^{n}(n\delta_n(X_{t_{k-1}},\alpha_0)-n\delta_n(X_{t_{k-1}},\alpha))^*\Xi_{k-1}^{-1}(\beta)(n\delta_n(X_{t_{k-1}},\alpha_0)-n\delta_n(X_{t_{k-1}},\alpha))$$
$$+2\sum_{k=1}^{n}(n\delta_n(X_{t_{k-1}},\alpha_0)-n\delta_n(X_{t_{k-1}},\alpha))^*\Xi_{k-1}^{-1}(\beta)P_k(\alpha_0).$$

Using Lemma 7 and [C2'], we may apply the results of Lemma 4 to the two sums above. This yields $\varepsilon^2\{U_{\varepsilon,n}(\alpha,\beta)-U_{\varepsilon,n}(\alpha_0,\beta)\}\to U_1(\alpha,\alpha_0,\beta)$.

$2^{\mathbf{nd}}$ **step.** We prove that $\varepsilon^{-1}(\hat\alpha_{\varepsilon,n}-\alpha_0)$ is tight. This is needed before proving the consistency for the parameter β. By consistency the probability of the event $\{\hat\alpha_{\varepsilon,n}\in\Theta_\alpha\}$ tends to 1, and on this event, by a first order expansion around $(\alpha_0,\hat\beta_{\varepsilon,n})$, we have $0=\frac{\partial}{\partial\alpha}U_{\varepsilon,n}(\hat\alpha_{\varepsilon,n},\hat\beta_{\varepsilon,n})=D_{\varepsilon,n}+N_{\varepsilon,n}(\hat\alpha_{n,\varepsilon}-\alpha_0)$, where $D_{\varepsilon,n}=\frac{\partial}{\partial\alpha}U_{\varepsilon,n}(\alpha_0,\hat\beta_{\varepsilon,n})$, and $N_{\varepsilon,n}$ is the symmetric matrix $N_{\varepsilon,n}=\int_0^1\frac{\partial^2}{\partial\alpha^2}U_{\varepsilon,n}(\alpha_0+t(\hat\alpha_{\varepsilon,n}-\alpha_0),\hat\beta_{\varepsilon,n})\mathrm{d}t$. By simple computations,

$$D_{\varepsilon,n}=-2\varepsilon^{-2}\sum_{k=1}^{n}P_k^*(\alpha_0)\Xi_{k-1}^{-1}(\hat\beta_{\varepsilon,n})\frac{\partial(n\delta_n)}{\partial\alpha}(X_{t_{k-1}},\alpha_0).$$

Using Lemma 4 2),

$$\sup_\beta\varepsilon^{-1}\left|\sum_{k=1}^{n}P_k^*(\alpha_0)\Xi_{k-1}^{-1}(\beta)\frac{\partial(n\delta_n)}{\partial\alpha}(X_{t_{k-1}},\alpha_0)\right|$$

is bounded in $\mathbb{P}$-probability. Thus $(\varepsilon D_{\varepsilon,n})_{\varepsilon,n}$ is a tight sequence.

We now focus on $N_{\varepsilon,n}$. The second order derivative with respect to α of the contrast function is given by

$$\begin{aligned}
&\frac{\partial^2}{\partial\alpha_i\partial\alpha_j}U_{\varepsilon,n}(\alpha,\beta) = \\
&\qquad -2\varepsilon^{-2}n^{-1}\sum_{k=1}^{n}\frac{\partial^2(n\delta_n)}{\partial\alpha_i\partial\alpha_j}(X_{t_{k-1}},\alpha)^*\Xi_{k-1}^{-1}(\beta)\left\{n\delta_n(X_{t_{k-1}},\alpha_0)-n\delta_n(X_{t_{k-1}},\alpha)\right\}\\
&\qquad +2\varepsilon^{-2}n^{-1}\sum_{k=1}^{n}\frac{\partial(n\delta_n)}{\partial\alpha_i}(X_{t_{k-1}},\alpha)^*\Xi_{k-1}^{-1}(\beta)\frac{\partial(n\delta_n)}{\partial\alpha_j}(X_{t_{k-1}},\alpha)\\
&\qquad -2\varepsilon^{-2}\sum_{k=1}^{n}P_k(\alpha_0)^*\Xi_{k-1}^{-1}(\beta)\frac{\partial^2(n\delta_n)}{\partial\alpha_i\partial\alpha_j}(X_{t_{k-1}},\alpha).
\end{aligned}$$

Again, application of Lemma 4 gives that $\varepsilon^2\frac{\partial^2 U_{\varepsilon,n}}{\partial\alpha_i\partial\alpha_j}(\alpha,\beta)$ converges to

$$\begin{aligned}
&-2\int_0^1\frac{\partial^2 b}{\partial\alpha_i\partial\alpha_j}(X_s^0,\alpha)^*[\sigma\sigma^*]^{-1}(X_s^0,\beta)\left\{b(X_s^0,\alpha_0)-b(X_s^0,\alpha)\right\}\mathrm{d}s\\
&\qquad\qquad +2\int_0^1\frac{\partial b}{\partial\alpha_i}(X_s^0,\alpha)^*[\sigma\sigma^*]^{-1}(X_s^0,\beta)\frac{\partial b}{\partial\alpha_j}(X_s^0,\alpha)\mathrm{d}s,
\end{aligned}$$

and by [A3'] and the positivity of $I_b(\theta_0)$, we deduce that

$$\begin{aligned}
(24)\qquad &\inf_{\beta\in\overline{\Theta}_\beta}\det\left(\left[\int_0^1\frac{\partial b}{\partial\alpha_i}(X_s^0,\alpha_0)^*[\sigma\sigma^*]^{-1}(X_s^0,\beta)\frac{\partial b}{\partial\alpha_j}(X_s^0,\alpha_0)\mathrm{d}s\right]_{1\le i,j\le p}\right)\\
&\qquad\geq c\det\left[\left(\int_0^1\frac{\partial b}{\partial\alpha_i}(X_s^0,\alpha_0)^*\frac{\partial b}{\partial\alpha_j}(X_s^0,\alpha_0)\mathrm{d}s\right]_{1\le i,j\le p}\right)>0.
\end{aligned}$$

Then the consistency of $\hat\alpha_{\varepsilon,n}$ implies that $\mathbb{P}(\det(\varepsilon^2 N_{\varepsilon,n})>0)$ tends to one. Thus, we get, on some event with arbitrarily large probability, that $\varepsilon^{-1}(\hat\alpha_{\varepsilon,n}-\alpha) = -(\varepsilon^2N_{\varepsilon,n})^{-1}\varepsilon D_{\varepsilon,n}$, and hence the sequence is tight.

3^{rd} **step.** We prove the consistency for the diffusion parameter. Again by a repetition of the arguments in Sørensen & Uchida (2003) it is sufficient to show the following convergence uniformly with respect to β:

$$n^{-1}\{U_{\varepsilon,n}(\hat\alpha_{\varepsilon,n},\beta)-U_{\varepsilon,n}(\hat\alpha_{\varepsilon,n},\beta_0)\}\xrightarrow[\mathbb{P}]{n\to\infty,\varepsilon\to 0}U_2(\beta,\beta_0).$$

Using the expression of the contrast function, we have:

$$n^{-1}\{U_{\varepsilon,n}(\hat{\alpha}_{\varepsilon,n},\beta)-U_{\varepsilon,n}(\hat{\alpha}_{\varepsilon,n},\beta_0)\} = n^{-1}\sum_{k=1}^{n}\log\det\Xi_{k-1}(\beta)\Xi_{k-1}^{-1}(\beta_0)$$
$$+\varepsilon^{-2}\sum_{k=1}^{n}P_k(\alpha_0)^*\Xi_{k-1}^{-1}(\beta)P_k(\alpha_0)-\varepsilon^{-2}\sum_{k=1}^{n}P_k(\alpha_0)^*\Xi_{k-1}^{-1}(\beta_0)P_k(\alpha_0)$$
$$+\Lambda^{(1)}(\hat{\alpha}_{\varepsilon,n},\alpha_0,\beta)+\Lambda^{(2)}(\hat{\alpha}_{\varepsilon,n},\alpha_0,\beta)-\Lambda^{(1)}(\hat{\alpha}_{\varepsilon,n},\alpha_0,\beta_0)-\Lambda^{(2)}(\hat{\alpha}_{\varepsilon,n},\alpha_0,\beta_0)$$

where

$$\Lambda^{(1)}(\alpha,\alpha_0,\beta)=\varepsilon^{-2}n^{-2}\sum_{k=1}^{n}\left\{n\delta_n(X_{t_{k-1}},\alpha_0)-n\delta_n(X_{t_{k-1}},\alpha)\right\}^*\Xi_{k-1}^{-1}(\beta)\left\{n\delta_n(X_{t_{k-1}},\alpha_0)-n\delta_n(X_{t_{k-1}},\alpha)\right\}, \tag{25}$$

and

$$\Lambda^{(2)}(\alpha,\alpha_0,\beta)= 2\varepsilon^{-2}n^{-1}\sum_{k=1}^{n}P_k(\alpha_0)^*\Xi_{k-1}^{-1}(\beta)\left\{n\delta_n(X_{t_{k-1}},\alpha_0)-n\delta_n(X_{t_{k-1}},\alpha)\right\}. \tag{26}$$

Using that $n\delta_n$ is Lipschitz and [A3'], we have $\left|\Lambda^{(1)}(\alpha,\alpha_0,\beta)\right|\leq cn^{-1}\varepsilon^{-2}\left|\alpha-\alpha_0\right|^2$ and $\left|\Lambda^{(2)}(\alpha,\alpha_0,\beta)\right|\leq cn^{-1}\varepsilon^{-2}\left|\alpha-\alpha_0\right|\sum_{k=1}^{n}|P_k(\alpha_0)|$. Thus using the tightness of $\varepsilon^{-1}\left|\alpha-\hat{\alpha}_{\varepsilon,n}\right|$ and Lemma 1 5), we deduce that the four last terms in the expansion of $n^{-1}\{U_{\varepsilon,n}(\hat{\alpha}_{\varepsilon,n},\beta)-U_{\varepsilon,n}(\hat{\alpha}_{\varepsilon,n},\beta_0)\}$ tends to 0 uniformly. Now using Lemmas 4 1) and 5, we deduce that the three first terms in this expansion converge to $U_2(\beta,\beta_0)$.

4.4.2. *Asymptotic normality of the estimator.* We consider the derivatives of the contrast function:

$$\Gamma_{\varepsilon,n}(\theta_0)=\begin{pmatrix}-\varepsilon\left(\frac{\partial}{\partial\alpha_i}U_{\varepsilon,n}(\theta)\right)_{1\leq i\leq p}\\ -\frac{1}{\sqrt{n}}\left(\frac{\partial}{\partial\beta_i}U_{\varepsilon,n}(\theta)\right)_{1\leq i\leq q}\end{pmatrix}$$

and

$$C_{\varepsilon,n}(\theta)=\begin{pmatrix}\varepsilon^2\left(\frac{\partial^2}{\partial\alpha_i\partial\alpha_j}U_{\varepsilon,n}(\theta)\right)_{1\leq i,j\leq p} & \varepsilon\frac{1}{\sqrt{n}}\left(\frac{\partial^2}{\partial\alpha_i\partial\beta_j}U_{\varepsilon,n}(\theta)\right)_{1\leq i\leq p,1\leq j\leq q}\\ \varepsilon\frac{1}{\sqrt{n}}\left(\frac{\partial^2}{\partial\beta_i\partial\alpha_j}U_{\varepsilon,n}(\theta)\right)_{1\leq i\leq p,1\leq j\leq q} & \frac{1}{n}\left(\frac{\partial^2}{\partial\beta_i\partial\beta_j}U_{\varepsilon,n}(\theta)\right)_{1\leq i,j\leq q}\end{pmatrix}.$$

Following the proof of Theorem 1 in Sørensen & Uchida (2003), the asymptotic normality follows from the two following properties:

$$\Gamma_{\varepsilon,n}(\theta_0) \xrightarrow{\text{law}} \mathcal{N}(0, 4I(\theta_0)), \tag{27}$$

$$\sup_{t\in[0,1]} \left| C_{\varepsilon,n}(\theta_0 + t(\hat{\theta}_{\varepsilon,n} - \theta_0)) - 2I(\theta_0) \right| \xrightarrow{\mathbb{P}} 0. \tag{28}$$

First, we prove (27). For this we compute,

$$-\varepsilon \frac{\partial U_{\varepsilon,n}}{\partial \alpha_i}(\theta_0) = 2\varepsilon^{-1} \sum_{k=1}^{n} P_k^*(\alpha_0) \Xi_{k-1}^{-1}(\beta_0) \frac{\partial (n\delta_n)}{\partial \alpha_i}(X_{t_{k-1}}, \alpha_0).$$

Using Lemma 7 and Lemma 1 5), we may write $-\varepsilon \frac{\partial U_{\varepsilon,n}}{\partial \alpha_i}(\theta_0) = \sum_{k=1}^n \xi_k^i(\theta_0) + \mathrm{O}_{L^1}\left(\varepsilon + n^{-1/2}\right)$ where

$$\xi_k^i(\theta_0) = 2\varepsilon^{-1} P_k^*(\alpha_0) \Xi_{k-1}^{-1}(\beta_0) \frac{\partial b}{\partial \alpha_i}(X_{t_{k-1}}, \alpha_0).$$

Differentiation of the contrast function with respect to β_j yields, $-n^{-1/2} \frac{\partial U_{\varepsilon,n}}{\partial \beta_j}(\theta_0) = \eta_k^j(\theta_0)$ where

$$\eta_k^j(\theta_0) = n^{-1/2} \operatorname{tr}\left(\Xi_{k-1}^{-1}(\beta_0) \frac{\partial \Xi_{k-1}}{\partial \beta_j}(\beta_0) \right) - \\ \varepsilon^{-2} n^{1/2} P_k^*(\alpha_0) \Xi_{k-1}^{-1}(\beta_0) \frac{\partial \Xi_{k-1}}{\partial \beta_j}(\beta_0) \Xi_{k-1}^{-1}(\beta_0) P_k(\alpha_0)$$

We know by Theorem 3.2 and 3.4 in Hall & Heyde (1980) that to obtain (27) it is sufficient show the following results on convergence in $\mathbb{P}$-probability

$$\sum_{k=1}^{n} \mathbb{E}\left[\xi_k^i(\theta_0) \mid \mathcal{G}_{k-1}^n \right] \to 0$$

$$\sum_{k=1}^{n} \mathbb{E}\left[\eta_k^j(\theta_0) \mid \mathcal{G}_{k-1}^n \right] \to 0$$

$$\sum_{k=1}^{n} \mathbb{E}\left[\xi_k^{i_1}(\theta_0) \xi_k^{i_2}(\theta_0) \mid \mathcal{G}_{k-1}^n \right] \to 4I_b^{i_1,i_2}(\theta_0)$$

$$\sum_{k=1}^{n} \mathbb{E}\left[\eta_k^{j_1}(\theta_0) \eta_k^{j_2}(\theta_0) \mid \mathcal{G}_{k-1}^n \right] \to 4I_\sigma^{j_1,j_2}(\theta_0)$$

$$\sum_{k=1}^{n} \mathbb{E}\left[\xi_k^i(\theta_0)\eta_k^j(\theta_0) \mid \mathcal{G}_{k-1}^n\right] \to 0$$

$$\sum_{k=1}^{n} \mathbb{E}\left[(\xi_k^i(\theta_0))^4 \mid \mathcal{G}_{k-1}^n\right] \to 0$$

$$\sum_{k=1}^{n} \mathbb{E}\left[(\eta_k^j(\theta_0))^4 \mid \mathcal{G}_{k-1}^n\right] \to 0$$

Theses seven properties follow from the expressions of $\xi_k^i(\theta_0)$, $\eta_k^i(\theta_0)$, Lemma 1 and Lemma 4 1); we omit the detailed proof.

Finally we show (28). For this, note that we have already shown in the 2$^{\text{nd}}$ step of Section 4.4.1 that $\varepsilon^2 \frac{\partial^2 U_{\varepsilon,n}}{\partial\alpha_i\partial\alpha_j}(\theta)$ converges uniformly to the quantity $2I_b^{i,j}(\theta) - 2\int_0^1 \frac{\partial^2 b}{\partial\alpha_i\partial\alpha_j}(X_s^0,\alpha)^*[\sigma\sigma^*]^{-1}(X_s^0,\beta)\left\{b(X_s^0,\alpha_0) - b(X_s^0,\alpha)\right\}\mathrm{d}s$. Hence consistency of the estimator yields

$$\sup_{t\in[0,1]}\left|C_{\varepsilon,n}(\theta_0 + t(\hat{\theta}_{\varepsilon,n} - \theta_0))^{i,j} - 2I(\theta_0)^{i,j}\right| \xrightarrow{\mathbb{P}} 0, \text{ for } 1 \le i,j \le p.$$

We now focus on the mixed term, for which we need to show that

$$\sup_{t\in[0,1]}\left|\frac{\varepsilon}{\sqrt{n}}\frac{\partial^2}{\partial\alpha_i\partial\beta_j}U_{\varepsilon,n}(\theta_0 + t(\hat{\theta}_{\varepsilon,n} - \theta_0))\right| \to 0. \tag{29}$$

However, by the expression of the contrast, we compute

$$\begin{aligned}\frac{\varepsilon}{\sqrt{n}}\frac{\partial^2 U_{\varepsilon,n}}{\partial\alpha_i\partial\beta_j}(\theta) = &- 2\varepsilon^{-1}n^{-\frac{1}{2}}\sum_{k=1}^{n} P_k(\alpha_0)^*\frac{\partial(\Xi_{k-1}^{-1})}{\partial\beta_j}(\beta)\frac{\partial(n\delta_n)}{\partial\alpha_i}(X_{t_{k-1}},\alpha)\\ &-2\varepsilon^{-1}n^{-\frac{3}{2}}\sum_{k=1}^{n}\frac{\partial(n\delta_n)}{\partial\alpha_i}(X_{t_{k-1}},\alpha)^*\frac{\partial(\Xi_{k-1}^{-1})}{\partial\beta_j}(\beta)\{n\delta_n(X_{t_{k-1}},\alpha_0) - n\delta_n(X_{t_{k-1}},\alpha)\}\end{aligned}$$

Using Lemma 4 2), the first sum tends to 0 in $\mathbb{P}$-probability, uniformly with respect to θ. Using the Lipschitz condition on $n\delta_n$, the second sum above is bounded by $cn^{-1/2}\left|\alpha - \alpha_0\right|\varepsilon^{-1}$. Thus its contribution is negligible by the tightness of $\left|\hat{\alpha}_{\varepsilon,n} - \alpha_0\right|\varepsilon^{-1}$. Hence (29) follows.

For the derivatives with respect to β, direct computation and application of Lemmas 4 1) and 5 gives (recall (25)–(26) too):

$$\frac{1}{n}\frac{\partial^2 U_{\varepsilon,n}}{\partial\beta_i\partial\beta_j}(\theta) = C(\beta,\beta_0)^{i,j} + o_{\mathbb{P}}(1) + \frac{\partial^2\Lambda^{(1)}}{\partial\beta_i\partial\beta_j}(\alpha,\alpha_0,\beta) + \frac{\partial^2\Lambda^{(2)}}{\partial\beta_i\partial\beta_j}(\alpha,\alpha_0,\beta)$$

where, with $\gamma = \sigma\sigma^*$,

$$\begin{aligned}
C(\beta,\beta_0)^{i,j} &= \int_0^1 \operatorname{tr}\Big(\gamma^{-1}\partial^2_{\beta_i,\beta_j}(\gamma)(X^0_s,\theta)\Big)\,\mathrm{d}s \\
&- \int_0^1 \operatorname{tr}\left(\gamma^{-1}\partial_{\beta_i}(\gamma)\gamma^{-1}\partial_{\beta_j}(\gamma)(X^0_s,\theta)\right)\mathrm{d}s \\
&- \int_0^1 \operatorname{tr}\Big((\gamma^{-1}\partial^2_{\beta_i,\beta_j}(\gamma)\gamma^{-1})(X^0_s,\theta)\gamma(X^0_s,\theta_0)\Big)\,\mathrm{d}s \\
&+ \int_0^1 \operatorname{tr}\left((\gamma^{-1}\partial_{\beta_i}(\gamma)\gamma^{-1}\partial_{\beta_j}(\gamma)\gamma^{-1})(X^0_s,\theta)\gamma(X^0_s,\theta_0)\right)\mathrm{d}s \\
&+ \int_0^1 \operatorname{tr}\left((\gamma^{-1}\partial_{\beta_j}(\gamma)\gamma^{-1}\partial_{\beta_i}(\gamma)\gamma^{-1})(X^0_s,\theta)\gamma(X^0_s,\theta_0)\right)\mathrm{d}s
\end{aligned}$$

Note that $C(\beta_0,\beta_0)^{i,j} = 2I^{i,j}_\sigma(\theta_0)$. Moreover, using [A3'], [C2'] and (25), we obtain the bound $\left|\frac{\partial^2\Lambda^{(1)}}{\partial\beta_i\partial\beta_j}(\alpha,\alpha_0,\beta)\right| \leq cn^{-1}\varepsilon^{-2}\left|\alpha-\alpha_0\right|^2$; and using Lemma 1 5), we have $\left|\frac{\partial^2\Lambda^{(2)}}{\partial\beta_i\partial\beta_j}(\alpha,\alpha_0,\beta)\right| \leq O_{L^1}(n^{-1/2})\varepsilon^{-1}\left|\alpha-\alpha_0\right|$. This is sufficient, with the tightness of $\varepsilon^{-1}\left|\hat{\alpha}_{\varepsilon,n}-\alpha_0\right|$, to conclude (28).

Acknowledgement. The authors would like to thank a referee for his very useful comments and in particular for suggesting remark 2.

References

Azencott R. (1982). Formule de Taylor stochastique et développement asympotique d'intégrales de Feynmann. *Séminaire de Probabilité XVI; Supplément: Géometrie Différentielle Stochastique. Lecture Notes In Math.* **921**, 237–285. Springer–Verlag, Berlin.

Bibby B.M., Jacobsen M. & Sørensen M. (2004). Estimating functions for discretely sampled diffusion-type models. Preprint 2004-4, Department of Applied Mathematics and Statistics, University of Copenhagen. To appear in Aït-Sahalia Y. & Hansen L.P. (eds.): *Handbook of Financial Econometrics*. North-Holland, Amsterdam.

Burkhölder D. (1973). Distribution function inequalities for martingales. *Ann. Probab.* **1**, 19-42.

Chan K.C., Andrew Karolyi G., Longstaff F.A. & Sanders A.B. (1992). An empirical comparison of alternative models of the short–term interest rate. *The Journal of Finance* **47**(3), 1209–1227.

Freidlin M.I. & Wentzell A.D. (1998). *Random pertubations of dynamical systems. Second edition* Springer–Verlag, New York.

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

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

Gihman I.I. & Skorohod A.V. (1972). *Stochastic differential equations.* Springer–Verlag, New York.

Gobet E. (2002). LAN property for ergodic diffusions with discrete observations. *Ann. I. H. Poincaré - PR*, **38**, 711–737.

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

Iacus S.M. & Kutoyants Yu. (2001). Semiparametric hypotheses testing for dynamical systems with small noise. *Math. Methods Statist.* **10**(1), 105–120.

Ibragimov I.A. & Has'minskii R.Z. (1981). *Statistical estimation. Asymptotic theory.* Springer–Verlag, New York.

Kessler M. (1997). Estimation of an ergodic diffusion from discrete observation. *Scand. J. Statist.* **24**, 211–239.

Kutoyants Yu. (1994). *Identification of dynamical systems with small noise.* Mathematics and its Applications, 300. Kluwer Academic Publishers Group.

Kutoyants Yu. (1998). Semiparametric estimation for dynamical systems with small noise. *Math. Methods Statist.* **7**(4), 457–465.

Longstaff F. & Schwartz E. (1995). A simple approach to valuing risky fixed and floating rate debt. *The journal of finance* **1**, 789–819.

Overhaus M., Bermúdez A., Buehler H., Ferraris A., Jordinson C., Lamnouar A. (2006). *Equity hybrid derivatives.* Wiley finance series.

Picard J. (1986). Nonlinear filtering of one-dimensional diffusions in the case of a high signal-to-noise ratio. *SIAM J. Appl. Math.* **46**(3), 1098–1125.

Picard J. (1991). Efficiency of the extended Kalman filter for nonlinear system with small noise. *SIAM J. Appl. Math.* **51**(3), 843–885.

Sørensen M. (2000). Small dispersion asymptotics for diffusion martingale estimating functions. Preprint 2000-2, Department of Statistics and Operation Research, University of Copenhagen, Copenhagen.

Sørensen M. & Uchida M. (2003). Small diffusion asymptotics for discretely sampled stochastic differential equations. *Bernoulli* **9**, 1051–1069.

Uchida M. (2004). Estimation for discretely observed small diffusions based on approximate martingale estimating functions. *Scand. J. Statist.* **31**, 553–566.

Uchida M. & Yoshida N. (2004a). Information criteria for small diffusions via the theory of Malliavin-Watanabe. *Stat. Inference Stoch. Process.* **7**, 35–67.

Uchida M. & Yoshida N. (2004b). Asymptotic expansion for small diffusions applied to option pricing. *Stat. Inference Stoch. Process.* **7**, 189–223

Walter W. (1998). *Ordinary differential equations.* Springer–Verlag, New York.

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

Arnaud Gloter
Université Paris-Est
Laboratoire d'Analyse et de Mathématiques Appliquées, UMR 8050
5, boulevard Descartes
Cité Descartes, Champs-sur-Marne
77454 Marne-la-Vallée cedex 2, France

Michael Sørensen
Department of Mathematics Sciences
University of Copenhagen
Universitetsparken 5
DK-2100 København Ø, Denmark

	$n=10$	$n=20$	$n=50$	$n=100$
$\hat{\mu}_1$	1.000 (0.01)	1.000 (0.01)	1.000 (0.01)	1.007 (0.01)
$\hat{\mu}_2$	1.006 (0.13)	0.998 (0.13)	1.005 (0.13)	0.997 (0.12)
$\hat{m}$	0.996 (0.04)	0.996 (0.04)	0.997 (0.04)	0.993 (0.04)
$\hat{\kappa}_1^2$	0.94 (0.43)	0.95 (0.30)	0.98 (0.19)	0.99 (0.14)
$\hat{\kappa}_2^2$	0.73 (0.27)	0.87 (0.29)	0.93 (0.18)	0.96 (0.13)
$\hat{\rho}$	0.392 (0.25)	0.34 (0.18)	0.31 (0.12)	0.30 (0.08)

TABLE 1. mean (s.d.) of the simulated estimator values ($\mu_1 = \mu_2 = m = \kappa_1^2 = \kappa_2^2 = 1$, $\rho = 0.3$, $\varepsilon = 0.01$)

	$n=10$	$n=20$	$n=50$	$n=100$
$\hat{\mu}_1$	0.994 (0.10)	0.998 (0.10)	1.000 (0.10)	0.999 (0.10)
$\hat{\mu}_2$	1.72 (1.19)	1.71 (1.26)	1.83 (1.21)	1.79 (1.25)
$\hat{m}$	0.94 (0.35)	0.93 (0.36)	0.97 (0.34)	0.96 (0.34)
$\hat{\kappa}_1^2$	0.92 (0.43)	0.95 (0.30)	0.98 (0.19)	0.99 (0.14)
$\hat{\kappa}_2^2$	0.73 (0.36)	0.85 (0.28)	0.92 (0.19)	0.96 (0.13)
$\hat{\rho}$	0.38 (0.24)	0.34 (0.19)	0.31 (0.12)	0.31 (0.09)

TABLE 2. mean (s.d.) of the simulated estimator values ($\mu_1 = \mu_2 = m = \kappa_1^2 = \kappa_2^2 = 1$, $\rho = 0.3$, $\varepsilon = 0.1$)

$\hat{\mu}_1$	-0.065 (0.53)	$\hat{\mu}_2$	10.02 (4.9)	$\hat{m}$	2.06 (0.56)
$\hat{\kappa}_1^2$	24.9 (2.1)	$\hat{\kappa}_2^2$	437 (33)	$\hat{\rho}$	0.013 (0.06)

TABLE 3. mean (s.d.) of the simulated estimator values ($\mu_1 = -0.125$, $\mu_2 = 5.7$, $m = 2$, $\kappa_1^2 = 25$, $\kappa_2^2 = 450$, $\rho = 0$)