NONPARAMETRIC ESTIMATION FOR STOCHASTIC DIFFERENTIAL EQUATIONS WITH RANDOM EFFECTS

F. COMTE[(1)], V. GENON-CATALOT[(1)] & A. SAMSON[(1)]

ABSTRACT. We consider N independent stochastic processes $(X_j(t), t \in [0,T])$, $j = 1, \dots, N$, defined by a one-dimensional stochastic differential equation with coefficients depending on a random variable ϕ_j and study the nonparametric estimation of the density of the random effect ϕ_j in two kinds of mixed models. A multiplicative random effect and an additive random effect are successively considered. In each case, we build kernel and deconvolution estimators and study their L^2-risk. Asymptotic properties are evaluated as N tends to infinity for fixed T or for $T = T(N)$ tending to infinity with N. For $T(N) = N^2$, adaptive estimators are built. Estimators are implemented on simulated data for several examples.

December 5, 2012

Keywords and phrases: Diffusion process, mixed models, nonparametric estimation, random effects.

AMS Classification. 62G07–62M05.

Running title: Estimation for SDE with random effects.

[(1)]: Sorbonne Paris Cité, MAP5, UMR 8145 CNRS, Université Paris Descartes, FRANCE,
email: fabienne.comte@parisdescartes.fr,
valentine.genon-catalot@parisdescartes.fr,
adeline.samson@parisdescartes.fr.

1. Introduction

Random effects models are increasingly used in the biomedical field and have proved to be adequate tools for the study of repeated measurements collected on a series of subjects (see *e.g.* Davidian and Giltinian, 1995, Pinheiro and Bates, 2000, Nie and Yang, 2005, Nie, 2006, 2007).

Such mixed models are defined through a hierarchical structure where, first, a distribution describes the intra-individual variability and, second, a distribution describes the inter-individual variability. In stochastic differential equations (SDEs) with random effects, the dynamics of each individual is given by a SDE with drift and/or diffusion coefficient depending on parameters, and the parameters of each SDE are random variables thus taking into account the variability between individuals. To be more precise, consider the following one-dimensional SDE modeling a continuous evolution for N subjects:

$$dX_j(t) = b(X_j(t), \phi_j)dt + \sigma(X_j(t), \phi_j)dW_j(t), \quad X_j(0) = x^j, \ \ j = 1, \ldots, N. \tag{1}$$

Here, $x^1, \ldots, x^N$ are known values, $(W_1, \ldots, W_N)$ are independent standard Brownian motions, $\phi_1, \ldots, \phi_N$ are *i.i.d.* random variables and $(W_1, \ldots, W_N)$ and $(\phi_1, \ldots, \phi_N)$ are independent. Each process $(X_j(t))$ represents an individual and the random variable ϕ_j represents the random effect of individual j. The functions $b(x, \varphi), \sigma(x, \varphi)$ are supposed to be known and statistical inference is mainly concerned with the common unknown distribution of the random effects ϕ_j. In several recent contributions, a parametric model is proposed for the latter distribution and various numerical as well as theoretical results have been obtained concerning maximum likelihood estimation (see *e.g.* Ditlevsen and De Gaetano, 2005, Overgaard *et al.*, 2005, Donnet and Samson, 2008, Picchini *et al.*, 2010, Picchini and Ditlevsen, 2011, Donnet and Samson, 2010, Delattre *et al.*, 2012).

Nonparametric estimation in the context of random effects models has been recently investigated for linear discrete time models (see *e.g.* Comte and Samson, 2012). To our knowledge, the problem of nonparametric estimation of the population distribution (distribution of ϕ_j) has not been yet investigated in the context of SDEs with random effects. In this paper, we study this topic in two special cases of (1). First, a multiplicative random effect in the drift:

$$dX_j(t) = \phi_j \, b(X_j(t))dt + \sigma(X_j(t))dW_j(t), \quad X_j(0) = x^j, \ \ j = 1, \ldots, N. \tag{2}$$

Second, an additive random effect in the drift:

$$dX_j(t) = (\phi_j + \ b(X_j(t)))dt + \sigma(X_j(t))dW_j(t), \quad X_j(0) = x^j, \ \ j = 1, \ldots, N. \tag{3}$$

The functions $b : \mathbb{R} \to \mathbb{R}$ and $\sigma : \mathbb{R} \to \mathbb{R}$ are known. We assume that the random variables $\phi_1, \ldots, \phi_N$ have a common density f on $\mathbb{R}$ and that the processes $(X_j(t))$ are continuously observed on a time interval $[0, T]$ with $T > 0$ given. We build a series of nonparametric estimators of the density f from the observations $\{X_j(t), 0 \le t \le T, j = 1, \ldots, N\}$ which are either kernel or deconvolution estimators. We study the L^2-risk of the proposed estimators and discuss their asymptotic properties as N tends to infinity. We distinguish the case of fixed T and the case of $T = T(N)$ tending to infinity with N. In view of practical implementation, the case of discrete observations $(X_j(k\delta), 0 \le k \le n, j = 1, \ldots, N)$, with $n\delta = T$ is studied. Numerical simulation results using discretized sample paths for various models are given.

Section 2 concerns model (2). For large T, we propose a kernel estimator $\hat{f}_h^{(1)}$ of f. The bound of its L^2-risk shows that $T = T(N) \ge N^3$ ensures the possibility of obtaining the usual optimal rates on Nikol'ski classes of regularity for f. Section 2.2 is devoted to the special case $b = \sigma$. Using an approach analogous to the one in Comte and Samson (2012), we build a deconvolution estimator $\hat{f}_m^{(2)}$ and discuss its properties for fixed and large T. Then, under the condition $T = T(N) = N^2$, a special class of deconvolution estimators $(\hat{f}_\tau, \tau \le N^2)$ is introduced.

We propose a data-driven selection of the parameter τ which is non standard and leads to an adaptive estimator (Theorem 1).

Section 3 concerns model (3). We follow an analogous scheme. The model assumptions and the constraints on T are weaker. First, a kernel estimator $\hat{f}_h^{(4)}$ is proposed which attains usual optimal rates for $T = T(N) \geq N^{5/2}$. Then, for $T = T(N) = N^2$ a class of deconvolution estimators $(\tilde{f}_\tau, \tau \leq N^2))$ and a data-driven selection of τ are studied.

In view of practical implementation, we consider the case of discrete observations with small sampling interval δ and check that the estimators can be adapted to this case (Section 4). This is used indeed in Section 5, devoted to numerical simulation on various models: the kernel and the adaptive deconvolution estimators are implemented. The results are satisfactory, even when the condition $T(N) = N^2$ is not fulfilled. In Section 6, model properties are investigated in the general framework of (1) or in the special cases of (2) and (3). Some concluding remarks are given in Section 7. Proofs are gathered in Section 8.

2. Multiplicative random effect in the drift

We consider in this section model (2) where φ belongs to $\mathbb{R}$ and b and σ are known functions satisfying:

(H1) b and σ are Lipschitz:

$$\exists L > 0, \forall x, y \in \mathbb{R}, \quad |b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leq L|x - y|.$$

(H2) $\displaystyle\int_0^T \frac{b^2}{\sigma^2}(X_j(s))ds < +\infty,\ j = 1, \ldots, N,\ \ a.s.$

The processes $(W_1, \ldots, W_N)$ and the r.v.'s $(\phi_1, \ldots, \phi_N)$ are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Consider the filtration $(\mathcal{F}_t, t \geq 0)$ defined by $\mathcal{F}_t = \sigma(\phi_j, W_j(s), s \leq t, j = 1, \ldots, N)$. As $\mathcal{F}_t = \sigma(W_i(s), s \leq t) \vee \mathcal{F}_t^i$, with $\mathcal{F}_t^i = \sigma(\phi_i, (\phi_j, W_j(s), s \leq t), j \neq i)$ independent of W_i, each process W_i is a $(\mathcal{F}_t, t \geq 0)$-Brownian motion. The random variables ϕ_i are $\mathcal{F}_0$-measurable. Under (H1), equation (2) admits a strong solution adapted to the filtration $(\mathcal{F}_t, t \geq 0)$ (see Proposition 8). Assumption (H2) is clearly fulfilled, for instance, if σ is lower bounded. Following Delattre *et al.* (2012), we may now introduce statistics which have a central role in the estimation procedure. For $j = 1, \ldots, N$, we denote

$$U_j(t) = \int_0^t \frac{b}{\sigma^2}(X_j(s))dX_j(s), \ \ V_j(t) = \int_0^t \frac{b^2}{\sigma^2}(X_j(s))ds. \tag{4}$$

Then, equation (2) yields:

$$U_j(t) = \phi_j V_j(t) + M_j(t),\ j = 1, \ldots, N, \tag{5}$$

with $M_j(t) = \int_0^t (b/\sigma)(X_j(s))dW_j(s)$.

2.1. **Large observation time.** In this paragraph, we consider the asymptotic framework where both N and T tend to infinity. And, in addition to (H1)-(H2), we assume:

(H3) $\displaystyle\int_0^{+\infty} \frac{b^2}{\sigma^2}(X_j(s))ds = +\infty,\ j = 1, \ldots, N,\ \ a.s.$

Let

$$A_{j,T} := \frac{U_j(T)}{V_j(T)}.$$

The statistic $A_{j,T}$ coincides with the maximum in φ of the conditional likelihood of (2) given $\phi_j = \varphi$. From (5), we see that $A_{j,T} = \phi_j + M_j(T)/V_j(T)$ (see *e.g.* Delattre *et al.* (2012)). The second term is the ratio of a martingale divided by its quadratic variation, which under (H3), tends to zero a.s. when T tends to infinity. Thus, $A_{j,T}$ is a consistent estimator of the random variable ϕ_j. To deal with expectations, we need the following stronger assumption implying (H3):

(H4) There exist positive constants c_0, c_1 such that

$$c_0^2 \leq b^2(x)/\sigma^2(x) \leq c_1^2, \quad \forall x \in \mathbb{R}.$$

As $A_{j,T}$ approximates ϕ_j, we introduce a kernel estimator for the unknown density f of ϕ_j:

$$\hat{f}_h^{(1)}(x) = \frac{1}{N}\sum_{j=1}^{N} K_h(x - A_{j,T}), \quad K_h(x) = \frac{1}{h}K\left(\frac{x}{h}\right), \tag{6}$$

where K is integrable with $\int K(u)du = 1$, C^1 and satisfies

$$\|K\|^2 := \int K^2(u)du < +\infty, \ \|K'\|^2 = \int (K')^2(u)du < +\infty, \tag{7}$$

where K' denotes the derivative of K and $\|.\|$ the L^2-norm. We define $f_h = f \star K_h$ where $\star$ denotes the convolution product.

Proposition 1. *Consider estimator $\hat{f}_h^{(1)}$ given by (6) under* (H1), (H2), (H4) *and (7). Then*

$$\mathbb{E}(\|\hat{f}_h^{(1)} - f\|^2) \leq 2\|f - f_h\|^2 + \frac{\|K\|^2}{Nh} + \frac{2c_1^2\|K'\|^2}{c_0^4}\frac{1}{Th^3}. \tag{8}$$

The proof of Proposition 1, as all other proofs, is relegated to Section 8.

Under weak regularity assumptions on f, $\|f - f_h\|^2$ tends to zero when h tends to zero. The estimating method is consistent, as soon as $1/(Nh) + 1/(Th^3)$ tends to zero. Since N is the number of *i.i.d.* observed trajectories, we now look at the rate of the L^2-risk expressed as a function of N and thus let T and h be expressed as functions of N. We recall that a kernel of order ℓ satisfies $\int x^k K(x)dx = 0$ for $k = 1, \ldots, \ell$. For constants $\beta > 0$ and $L > 0$, the Nikol'ski class $\mathcal{N}(\beta, L)$ is defined by:

$$\mathcal{N}(\beta, L) = \left\{ f : \mathbb{R} \mapsto \mathbb{R} : \left[\int \left(f^{(\ell)}(x+t) - f^{(\ell)}(x)\right)^2 dx\right\}^{1/2} \leq L|t|^{\beta-\ell}, \quad \forall t \in \mathbb{R} \right\}.$$

The following Corollary holds

Corollary 1. *Assume that f belongs to $\mathcal{N}(\beta, L)$ and that the kernel K has order $\ell = \lfloor \beta \rfloor$ with $\int |x|^\beta |K(x)|dx < +\infty$. Under the condition*

$$h \propto N^{-1/(2\beta+1)} \quad \text{and} \quad T = T(N) \geq N^{(2\beta+3)/(2\beta+1)}, \tag{9}$$

we have

$$\mathbb{E}(\|\hat{f}_h^{(1)} - f\|^2) \lesssim N^{-2\beta/(2\beta+1)}. \tag{10}$$

We give here the classical discussion implying Corollary 1. If f belongs to $\mathcal{N}(\beta, L)$ and if the kernel K has order ℓ, then $\|f - f_h\|^2 \leq C^2 h^{2\beta}$ with $C = L\int |u|^\beta |K(u)|du/\ell!$ (see Tsybakov, 2009). For the first two terms in the r.h.s. of (8), the classical rate-optimal compromise imposes that $h \propto N^{-1/(2\beta+1)}$ thus implying $h^{2\beta} + 1/(Nh) \propto N^{-2\beta/(2\beta+1)}$. Fitting the last term with this rate requires that $1/(Th^3) \leq N^{-2\beta/(2\beta+1)}$. This holds for $T \geq N^{(2\beta+3)/(2\beta+1)}$. The conditions in (9) yield the rate (10).

The constraint on T is satisfied for any nonnegative β if $T \geq N^3$. If $\beta \geq \beta_0$ for some known β_0, we only require $T \geq N^{(2\beta_0+3)/(2\beta_0+1)}$. Even if β is very large, we always need T/N tends to infinity when N tends to infinity.

With $T = T(N) \geq N^3$, an adaptive selection of the bandwidth h could be done relying on the method described in Goldenshluger and Lepski (2011).

2.2. **The special case $b = \sigma$ for fixed T.** The following approach is inspired by the discrete time model studied in Comte and Samson, 2012. Indeed, when $b = \sigma$, the variables in (4) and equation (5) reduce to:

$$U_j(t) = \int_0^t \frac{dX_j(s)}{\sigma(X_j(s))}, \;\; V_j(t) = t, \;\; U_j(t) = \phi_j t + W_j(t), \; j = 1, \ldots, N.$$

Assume that T is fixed and let Δ be a sampling interval to be discussed later. For $k = 1, \ldots, K$, $t_k = k\Delta$, $T = K\Delta$, we consider the following random variables based on observations

$$U_{j,k} := \frac{U_j(t_k) - U_j(t_{k-1})}{\Delta} = \phi_j + \frac{1}{\sqrt{\Delta}} \frac{W_j(t_k) - W_j(t_{k-1})}{\sqrt{\Delta}}. \tag{11}$$

Using a Gaussian deconvolution, we define the estimator:

$$\hat{f}_m^{(2)}(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-iux} \frac{\Delta}{NT} \sum_{j=1}^{N} \sum_{k=1}^{K} e^{iuU_{j,k}} e^{u^2/(2\Delta)} du, \tag{12}$$

where m is a cutoff which should be chosen adaptively.

Recall some standard notations. For possibly complex-valued functions $g, h \in L^2(\mathbb{R}) \cap L^1(\mathbb{R})$, we denote by $g^*(u) = \int e^{iux} g(x) dx$ the Fourier transform of g, by $\langle g, h \rangle = \int g(x) \overline{h}(x) dx$ the L^2-scalar product. The Plancherel theorem yields $\langle g, h \rangle = \langle g^*, h^* \rangle / 2\pi$. Let now

$$f_m(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-iux} f^*(u) du, \;\; i.e. \;\; f_m^*(u) = f^*(u) \mathbb{1}_{[-\pi m, \pi m]}(u)$$

so that $\hat{f}_m^{(2)}(x)$ is, for all x, an unbiased estimator of $f_m(x)$, *i.e.* $\mathbb{E}(\hat{f}_m^{(2)}(x)) = f_m(x)$. We have the following result.

Proposition 2. *Consider the estimator $\hat{f}_m^{(2)}$ given by (12). We have*

$$\begin{aligned} \mathbb{E}(\|f - \hat{f}_m^{(2)}\|^2) &\leq \|f - f_m\|^2 + \frac{\Delta}{2\pi NT} \int_{-\pi m}^{\pi m} e^{u^2/\Delta} du + \frac{m}{N} \\ &\leq \frac{1}{2\pi} \int_{|u| \geq \pi m} |f^*(u)|^2 du + \frac{\Delta}{\pi NT} \left(\frac{\Delta e^{\pi^2 m^2/\Delta}}{\pi m} + \frac{e}{2}\sqrt{\Delta} \right) + \frac{m}{N}. \end{aligned} \tag{13}$$

The bound of the L^2-risk is classically split into the sum of a bias term $\|f - f_m\|^2$ and a variance term, itself the sum of two expressions (last two terms in the r.h.s. of (13)). Let us discuss the possible best choice of Δ for fixed T.
Do we have interest to take small Δ? The function $\Delta \mapsto \Delta^2 e^{(\pi m)^2/\Delta}$ decreases for $\Delta \leq \pi^2 m^2/2$, increases for $\Delta \geq \pi^2 m^2/2$ and thus admits a unique minimum reached at $\pi^2 m^2/2$. As m has to be large (for the bias to be small), the choice of Δ minimizing $\Delta^2 e^{(\pi m)^2/\Delta}$ is excluded. Consequently, with fixed $T \geq 1$, the best choice of $\Delta \leq 1$ is just $\Delta = 1$. It follows that the dominating variance term is $\propto (e^{(\pi m)^2}/m)/N$. This is the standard order of variance in nonparametric deconvolution when measurement errors are Gaussian. We refer to *e.g.* Comte *et al.* (2006) for defining the adequate adaptive choice of the cut-off m.

In view of the next section, let us consider the case of a large T. Then, we can choose a large Δ if $T \geq (\pi m)^2/2$. With $\Delta = (\pi m)^2/2$ and $T \geq (\pi m)^2/2$,

$$\frac{\Delta^2 e^{(\pi m)^2/\Delta}}{\pi^2 m\, N\, T} \leq \frac{e^2}{2}\frac{m}{N}.$$

The two variance terms thus have the same order m/N which is standard in density estimation without noise. For instance, for $\Delta = T, K = 1, \pi m = \sqrt{T}$, the estimator (12) is equal to:

$$\frac{1}{2\pi}\int_{-\sqrt{T}}^{\sqrt{T}} e^{-iux}\frac{1}{N}\sum_{j=1}^{N} e^{iuA_{j,T}} e^{u^2/(2T)} du, \text{ with here } \quad A_{j,T} = \frac{U_j(T)}{T}.$$

The variance term in its L^2-risk bound is simply $(3e/2)\sqrt{T}/N + m/N$. This suggests the choice $T(N) \propto N^2$ as adequate.

2.3. **Estimation in the case $b = \sigma$ for large T.** In the case of large T, with $b = \sigma$, the previous discussion indicates that $T = T(N)$ should be proportional to N^2. Let us assume, for simplicity, that

$$T = T(N) = N^2.$$

The discussion also suggests to introduce a new class of estimators defined by

$$\hat{f}_\tau(x) = \frac{1}{2\pi}\int_{-\sqrt{\tau}}^{\sqrt{\tau}} e^{-iux}\frac{1}{N}\sum_{j=1}^{N} e^{iuA_{j,\tau}} e^{u^2/(2\tau)} du,$$

based on the observed processes

$$A_{j,\tau} = \frac{U_j(\tau)}{\tau}, \tau \in [0, N^2]. \tag{14}$$

Then, we proceed to build an adaptive choice of τ, among the integer values $\{1, \dots, N^2\}$. Note that the class of estimators $(\hat{f}_\tau)$ presents a non standard feature as the parameter τ appears not only as a cut-off parameter for deconvolution but also inside the integral. The following risk bound holds when τ is fixed.

Proposition 3. *Assume* (H1), $b = \sigma$ *and* $\tau \leq T(N) = N^2$. *Then,*

$$\mathbb{E}(\|\hat{f}_\tau - f\|^2) \leq \frac{1}{2\pi}\int_{|u|\geq\sqrt{\tau}} |f^*(u)|^2 du + \frac{\sqrt{\tau}}{\pi N}\left(1 + \int_0^1 e^{v^2} dv\right).$$

Let f_τ be such that $f_\tau^* = f^*\mathbf{1}_{[-\sqrt{\tau},\sqrt{\tau}]}$. The bias term above is:

$$\|f - f_\tau\|^2 = \frac{1}{2\pi}\int_{|u|\geq\sqrt{\tau}} |f^*(u)|^2 du.$$

Regularity spaces usually considered in deconvolution setting are Sobolev balls defined by:

$$\mathcal{C}(a, L) = \{f \in (\mathbb{L}^1 \cap \mathbb{L}^2)(\mathbb{R}), \int (1+u^2)^a |f^*(u)|^2 du \leq L\}. \tag{15}$$

Clearly, if f belongs to $\mathcal{C}(a, L)$:

$$\frac{1}{2\pi}\int_{|u|\geq\sqrt{\tau}} |f^*(u)|^2 du \leq (L/2\pi)\tau^{-a}.$$

Thus, the bias-variance compromise is obtained for τ of order $N^{2/(2a+1)}$. The resulting rate is proportional to $N^{-2a/(2a+1)}$. Evidently, as a is unknown, this choice of τ cannot be done.

Therefore, we now define the data-driven selection of τ. With κ a constant to be specified later, let

$$\text{pen}(k) = \kappa \frac{\sqrt{k}}{N}, \tag{16}$$

and

$$\hat{\Gamma}_\tau = \max_{1\le k\le N^2} \left(\|\hat{f}_{k\wedge\tau} - \hat{f}_k\|^2 - \text{pen}(k)\right)_+ = \max_{\tau\le k\le N^2} \left(\|\hat{f}_\tau - \hat{f}_k\|^2 - \text{pen}(k)\right)_+, \tag{17}$$

where $(x)_+ = \max(x, 0)$.
Now define

$$\hat{\tau} = \arg\min_{1\le\tau\le N^2} \left\{\hat{\Gamma}_\tau + \text{pen}(\tau)\right\}. \tag{18}$$

As in Goldenschluger-Lepski method (see reference [9]), $\hat{\Gamma}_\tau$ is a data-driven estimation of the bias term $\|f - f_\tau\|^2$ while $\text{pen}(\tau)$ corresponds to the variance term. The choice of $\hat{\tau}$ is thus done to minimise the bias-variance compromise.

Recall that f_τ is defined by $f_\tau^* = f^* \mathbb{1}_{[-\sqrt{\tau},\sqrt{\tau}]}$.

Theorem 1. *Assume* (H1), $b = \sigma$ *and* $T(N) = N^2$. *For* $\kappa \ge 12 \int_0^1 e^{v^2} dv/\pi (\simeq 5.58)$, *we get*

$$\mathbb{E}(\|\hat{f}_{\hat{\tau}} - f\|^2 \le C \inf_{1\le\tau\le N^2} (\|f_\tau - f\|^2 + \text{pen}(\tau)) + \frac{C'}{N},$$

where C is a numerical constant ($C = 7$) and C' is a constant.

Theorem 1 states that the bias-variance compromise is automatically achieved by the adaptive estimator $\hat{f}_{\hat{\tau}}$ without knowledge of the regularity of f.

The condition $T(N) = N^2$ may appear rather strong, but can be weakened if some knowledge on the regularity of f is available. For instance, assume that $f \in \mathcal{C}(a, L)$ with $a \ge 2$. Then the optimal τ has order $N^{2/(2a+1)} \le N^{2/5}$. Therefore, the condition $T(N) = N^2$ can be replaced by $T(N) = N^{2/5}$.

2.4. **Miscellaneous remarks.** The two strategies described sections 2.1 and 2.2 have links. Under (H1) and (H4), f can be as well estimated using deconvolution as follows:

$$\hat{f}_m^{(3)}(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-iux} \frac{1}{N} \sum_{j=1}^N e^{iuA_{j,T}} du = \frac{1}{\pi N} \sum_{j=1}^N \frac{\sin(\pi m(A_{j,T} - x))}{A_{j,T} - x}. \tag{19}$$

This estimator corresponds to a kernel estimator as (6) with $K(x) = \sin(\pi x)/(\pi x)$ (not fulfilling (7)) and bandwidth $h = 1/m$ and is also a possible estimator of f in the special case $b = \sigma$. The risk is then bounded as follows.

Proposition 4. *Consider the estimator given by (19) under* (H1) *and* (H4). *We have*

$$\mathbb{E}(\|\hat{f}_m^{(3)} - f\|^2) \le \frac{1}{2\pi} \int_{|u|\ge\pi m} |f^*(u)|^2 du + \frac{\pi^2 c_1^2}{3c_0^4} \frac{m^3}{T} + \frac{m}{N}. \tag{20}$$

If $b = \sigma$,

$$\mathbb{E}(\|\hat{f}_m^{(3)} - f\|^2) \le \frac{1}{2\pi} \int_{|u|\ge\pi m} |f^*(u)|^2 du + \frac{1}{8\pi T^2} \int_{-\pi m}^{\pi m} |u|^4 |f^*(u)|^2 du + \frac{m}{N}. \tag{21}$$

Compared with Proposition 1, the bound in (20) is analogous to (8). In the special case $b = \sigma$, the risk bound of $\hat{f}_h^{(1)}$ is unchanged.
On the contrary, the bound of $\hat{f}_m^{(3)}$ in (21) contains an additional bias term (middle term in the r.h.s. of (21)) which allows to weaken the constraints on T. For instance, if $\int |u|^4 |f^*(u)|^2 du < +\infty$, which is true if $f \in \mathcal{C}(a, L)$ for $a \geq 2$, the new term is negligible as soon as $T \geq \sqrt{N}$. Thus, if the regularity index of f is larger than 2, rates of density estimation without noise can be attained.
If we have no knowledge at all on the behaviour of $|f^*(u)|^2$, we can only use that $|f^*(u)| \leq 1$ and get the order m^5/T^2 which is better than m^3/T.

3. Additive random effect in the drift

We study in this section model (3), under (H1).

3.1. **Large time strategy.** Relying on Equation (3), we set:

$$Z_{j,T} := \frac{X_j(T) - X_j(0) - \int_0^T b(X_j(s))ds}{T} = \phi_j + \frac{1}{T}\int_0^T \sigma(X_j(s))dW_j(s),$$

and define the estimator

$$\hat{f}_h^{(4)}(x) = \frac{1}{N}\sum_{j=1}^N K_h(x - Z_{j,T}), \tag{22}$$

where $K_h(x) = (1/h)K(x/h)$, K is C^2, integrable, with $\int K(u)du = 1$ and

$$\int K^2(u)du < +\infty, \quad \int (K")^2(u)du < +\infty. \tag{23}$$

The risk of $\hat{f}_h^{(4)}$ is slightly different than the risk of $\hat{f}_h^{(1)}$ of Section 2.

Proposition 5. *Consider estimator $\hat{f}_h^{(4)}$ given by (22) under* (H1) *and (23). Assume moreover that $\sigma^2(x) \leq \sigma_1^2$, $\forall x \in \mathbb{R}$. Then*

$$\mathbb{E}(\|\hat{f}_h^{(4)} - f\|^2) \leq 2\|f - f_h\|^2 + \frac{\|K\|^2}{Nh} + c\sigma_1^4 \|K"\|^2 \frac{1}{T^2 h^5} \tag{24}$$

where c is a numerical constant.

The discussion of Corollary 1 can be done here. Consider that f belongs to the Nikol'ski class $\mathcal{N}(\beta, L)$ and that the kernel has order $\ell = \lfloor \beta \rfloor$ with $\int |x|^\beta |K(x)|dx < +\infty$. If we impose that $h \propto N^{-1/(2\beta+1)}$, we get a rate proportional to $N^{-2\beta/(2\beta+1)}$ plus a term of order $N^{5/(2\beta+1)}/T^2$. If $T^2 \geq N^{(2\beta+5)/(2\beta+1)}$, then $N^{5/(2\beta+1)}/T^2 \leq N^{-2\beta/(2\beta+1)}$. This constraint holds for any nonnegative β if $T \geq N^{5/2}$. Note that this constraint is weaker than for $\hat{f}_h^{(1)}$ and the result holds under weaker assumptions for the model.

3.2. **Fixed step strategy using increments.** The method using increments with a sampling interval Δ is possible here. Using (3) yields,

$$Y_{j,k} := \frac{X_j(t_k) - X_j(t_{k-1}) - \int_{t_{k-1}}^{t_k} b(X_j(s))ds}{t_k - t_{k-1}} = \phi_j + \frac{1}{t_k - t_{k-1}}\int_{t_{k-1}}^{t_k} \sigma(X_j(s))dW_j(s)$$

where $t_k = k\Delta, k = 1, \dots, K$, $T = K\Delta$. Let:

$$\hat{f}_m^{(5)}(x) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{-iux}\frac{\Delta}{NT}\sum_{j=1}^{N}\sum_{k=1}^{K} e^{iuY_{j,k}} e^{\frac{u^2}{2\Delta^2}\int_{t_{k-1}}^{t_k}\sigma^2(X_j(s))ds} du.$$

Note that, in the above formula, we use a "stochastic deconvolution" to build the estimator. The following risk decomposition holds.

Proposition 6. *Consider estimator $\hat{f}_m^{(5)}$ given above under* (H1). *Assume moreover that $\sigma^2(x) \le \sigma_1^2$, $\forall x \in \mathbb{R}$. Then*

$$\mathbb{E}(\|\hat{f}_m^{(5)} - f\|^2) \le \|f - f_m\|^2 + \frac{\Delta}{\pi NT}\int_{-\pi m}^{\pi m} e^{u^2\sigma_1^2/\Delta} du + \frac{m}{N}.$$

The same discussion as after Proposition 2 can be done here. As above, we can define another class of estimators depending on a cutoff parameter τ and build an adaptive choice of it. Let

$$\tilde{f}_\tau(x) = \frac{1}{2\pi}\int_{-\sqrt{\tau}}^{\sqrt{\tau}} e^{-iux}\frac{1}{N}\sum_{j=1}^{N} e^{iuZ_{j,\tau}} e^{\frac{u^2}{2\tau^2}\int_0^\tau \sigma^2(X_j(s))ds} du,$$

$$Z_{j,\tau} := \frac{X_j(\tau) - X_j(0) - \int_0^\tau b(X_j(s))ds}{\tau}.$$

We define

$$\tilde{\Gamma}_\tau = \max_{1\le k\le N^2}\left(\|\tilde{f}_{k\wedge\tau} - \tilde{f}_k\|^2 - \widetilde{\mathrm{pen}}(k)\right)_+ \text{ with } \widetilde{\mathrm{pen}}(k) = \tilde{\kappa}\log(N)\frac{\sqrt{k}}{N}, \tag{25}$$

Let

$$\tilde{\tau} = \arg\min_{1\le\tau\le N^2}\left\{\tilde{\Gamma}_\tau + \widetilde{\mathrm{pen}}(\tau)\right\}. \tag{26}$$

Theorem 2. *Assume* (H1) *and $T(N) = N^2$. Assume moreover that $\sigma^2(x) \le \sigma_1^2$, $\forall x \in \mathbb{R}$. Then there exists a numerical constant $\tilde{\kappa}$ such that:*

$$\mathbb{E}(\|\tilde{f}_{\tilde{\tau}} - f\|^2 \le C\inf_{1\le\tau\le N^2}(\|f_\tau - f\|^2 + \widetilde{\mathrm{pen}}(\tau)) + \frac{C'}{N}, \tag{27}$$

where C is a numerical constant and C' is a constant.

The additional factor $\log(N)$ in the penalty $\widetilde{\mathrm{pen}}$ is due to the fact that $\sigma(x)$ is not constant. It implies a logarithmic loss in the rate. More precisely, if $f \in \mathcal{C}(a, L)$, the infimum in the right-hand-side of (27) is of order $(N/\log(N))^{-2a/(2a+1)}$ instead of $N^{-2a/(2a+1)}$.

If $\sigma(x)$ is constant, this factor $\log(N)$ in the penalty is not needed.

Remark 1. *The study of Section 3 can be done under weaker assumptions on σ including the case where a random effect is present in the diffusion coefficient too, say $\sigma(x, \psi)$ provided that* $\sup_{t\ge0}\mathbb{E}(\sigma^2(X_j(t), \psi_j) = \sigma_1^2 < +\infty$.

4. Discretizations

In practice, only discrete time observations of $(X_j(t))$ are available. We investigate now this context and assume that for a small sampling interval δ, the observations are $(X_j(k\delta))_{1\le k\le n}$, with $T = n\delta$. In estimators of Sections 2 and 3, we replace all ordinary and stochastic integrals based on the continuous time observations by their usual discrete approximations using the discrete data. This procedure is classical. To avoid unnecessary repetitions, we only deal with the estimators built in Section 3.

We replace the $Z_{j,T}$'s by the usual approximation:

$$\hat{Z}_{j,T} = \frac{1}{T}\left(X_j(T) - X_j(0) - \delta \sum_{k=1}^{n} b(X_j((k-1)\delta))\right),$$

and define the estimator based on discrete observations:

$$\hat{f}^{(4)}_{h,\delta}(x) = \frac{1}{N}\sum_{j=1}^{N} K_h\left(x - \hat{Z}_{j,T}\right).$$

Analogously, for $\tau \leq T$, set:

$$\hat{Z}_{j,\tau} = \frac{1}{\tau}\left(X_j(\delta[\tau/\delta]) - X_j(0) - \delta \sum_{k=1}^{[\tau/\delta]} b(X_j((k-1)\delta))\right),$$

$$\tilde{f}_{\tau,\delta}(x) = \frac{1}{2\pi}\int_{-\sqrt{\tau}}^{\sqrt{\tau}} e^{-iux}\frac{1}{N}\sum_{j=1}^{N} e^{iu\hat{Z}_{j,\tau}} e^{\frac{u^2}{2\tau^2}\sum_{k=1}^{[\tau/\delta]}\delta\sigma^2(X_j((k-1)\delta))} du.$$

The following proposition holds:

Proposition 7. *Assume* (H1). *Assume moreover that* $\mathbb{E}(\phi_j^2) < +\infty$, $\sigma^2(x) \leq \sigma_1^2$ *for all* x *and* $\sup_{s\geq 0}\mathbb{E}[b^2(X_j(s))] < +\infty$.

(1) *Let the kernel satisfy (7) and (23). Then,*

$$\mathbb{E}(\|\hat{f}^{(4)}_{h,\delta} - f\|^2) \leq 2\|f - f_h\|^2 + \frac{\|K\|^2}{Nh} + c\,\sigma_1^4\|K"\|^2\frac{1}{T^2h^5} + C\frac{\delta}{h^3}\|K'\|^2, \tag{28}$$

where c, C *are constants not depending on* T, h, δ.

(2) *For* $0 < \delta < 1 \leq \tau$,

$$\mathbb{E}(\|\tilde{f}_{\tau,\delta} - f\|^2) \leq C(\|f_\tau - f\|^2 + \sqrt{\tau}(\frac{1}{N} + \delta)), \tag{29}$$

where C *is a constant not depending on* τ, δ.

Let us give some comments on the assumption $\sup_{s\geq 0}\mathbb{E}[b^2(X_j(s))] < +\infty$. It holds obviously if the drift b is bounded. Otherwise, it may be linked with the existence of a stationary distribution for model (3), but it will be checked directly in the examples below. Repeating all steps of Theorem 2, we can define a data-driven choice $\tilde{\tau}_\delta$ of τ and prove that the corresponding estimator $\tilde{f}_{\tilde{\tau}_\delta,\delta}$ satisfies an inequality analogous to (27).

5. Numerical simulation results.

In this section, we consider three models of multiplicative or additive form.

Model (1). Geometric Brownian motion:

$$dX_j(t) = X_j(t)(\phi_j dt + \sigma dW_j(t)), X_j(0) = x^j > 0. \tag{30}$$

Then, $X_j(t) = \exp(Y_j(t))$ with $dY_j(t) = (\phi_j - \sigma^2/2)dt + \sigma dW_j(t)$.
For this model, we use exact simulations for $N = 50$ and $T = 300$ or $N = 200$ and $T = 300$, with $\sigma = 1$, $x^j = 1$. The distributions of the random effects are: Gamma $\Gamma(3,\ 0.1)$, where 3 is the shape parameter and 0.1 the scale parameter, mixed gamma $0.3\ \Gamma(3, 0.1) + 0.7\ \Gamma(15, 0.1)$, and Gaussian $\mathcal{N}(1, 0.2^2)$.

The computation of the kernel estimator is fast as we only need the terminal variable $U_j(T)$, which has an explicit form for model (1). The bandwidth is selected by the R-function *density*,

Kernel estimator $\hat{f}_h^{(1)}$ **Deconvolution estimator** $\hat{f}_{\hat{\tau}}$

$N = 50,\ T = 300$ $N = 200,\ T = 300$ $N = 200,\ T = 300$

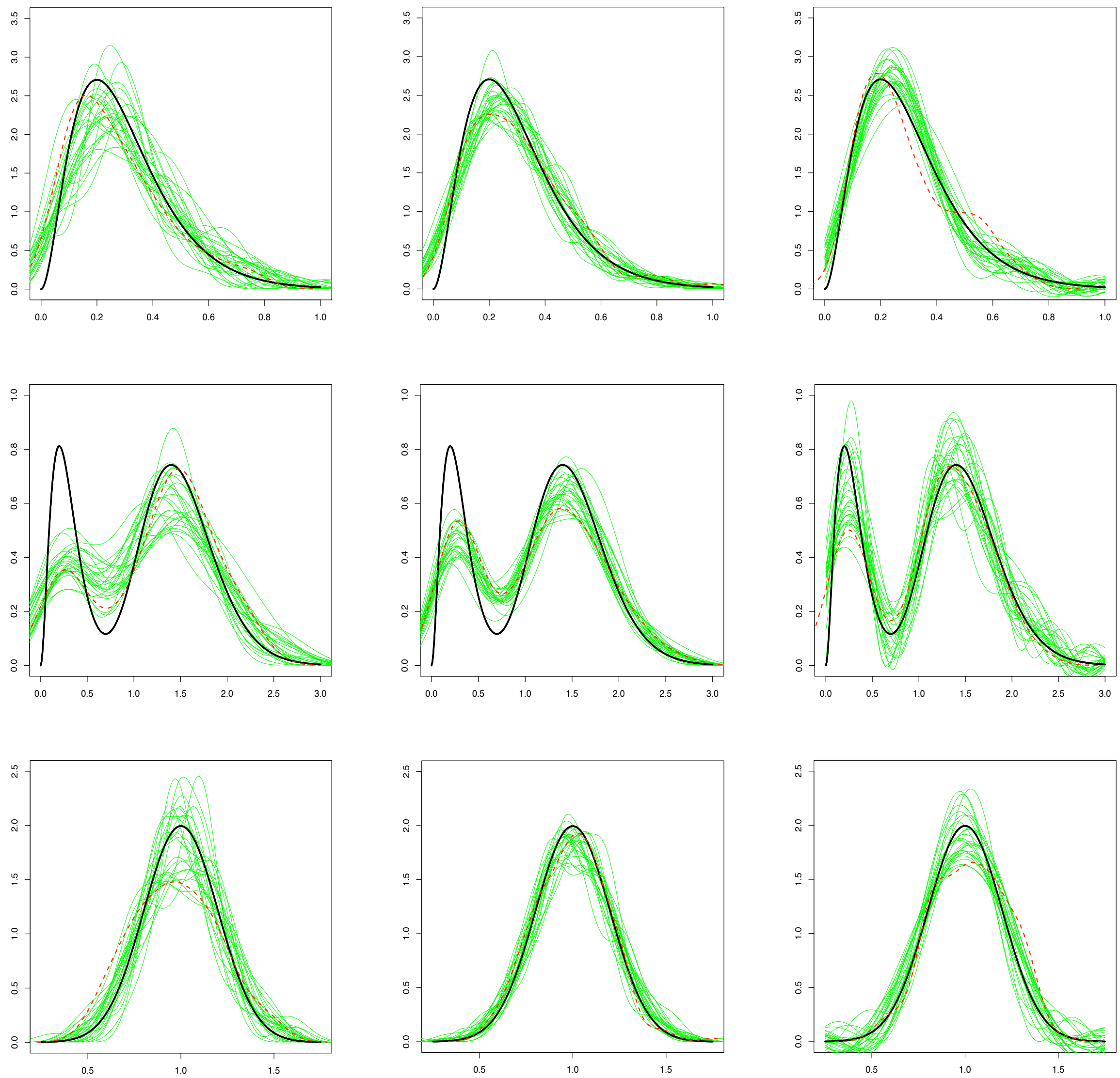

Figure 1. Model (1), Geometric Brownian Motion with distribution of the random effects: Gamma (first line), mixed gamma (second line), and Gaussian (third line). Estimated density for 25 independent samples in thin green: kernel estimator (first two columns) and deconvolution estimator (last column). True density in bold black. Estimated density for one sample of (ϕ_j)'s directly observed with standard kernel density estimator in bold dotted red.

with a Gaussian kernel. The bandwidth h is selected by cross-validation. The deconvolution-type estimator $\hat{f}_{\hat{\tau}}$ is computed, with constant κ calibrated through preliminary simulations to $\kappa = 150$. Results are illustrated in Figure 1 which represents variability bands for the two estimators. The

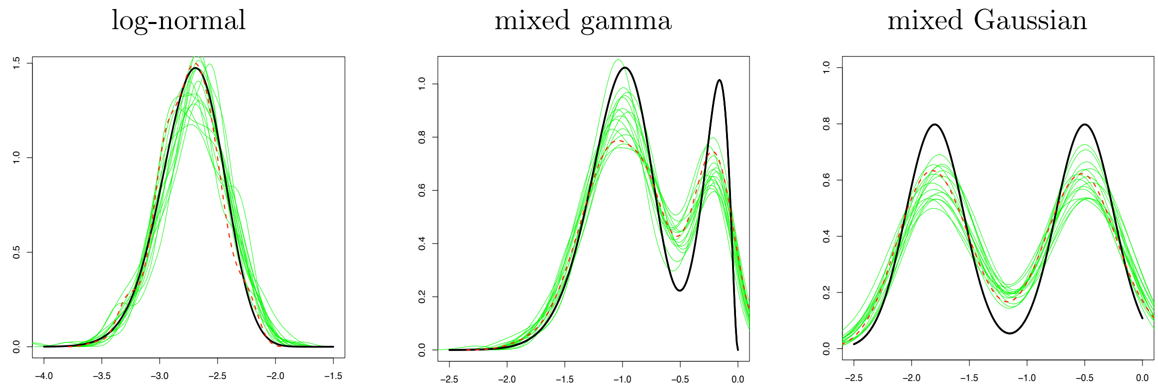

FIGURE 2. Model (2), multiplicative Ornstein-Uhlenbeck process with distribution of the random effects: log-normal (first column), mixed gamma (second column), and mixed Gaussian (third column). Kernel density estimators ($\hat{f}_h^{(1)}$) for 25 independent samples in thin green. True density in bold black. Estimated density for one sample of (ϕ_j)'s directly observed with standard kernel density estimator in bold dotted red. $N = 200,\ T = 400$.

first column gives 25 estimated densities with $N = 50,\ T = 300$ for the kernel estimator. In the last two columns, we have $N = 200,\ T = 300$, for the kernel (column 2) and the deconvolution (last column) estimators. Clearly, increasing N improves the result as expected. Note that the methods work well even for T/N not very large. Looking at the last two columns, we see that the deconvolution method seems less biased but with a greater variability than the kernel method. Especially, for a bimodal distribution, the deconvolution methods works well, even better than when the individual parameter ϕ_j are directly observed (dotted line). Nevertheless, the deconvolution method is more time consuming as we need to compute all the $U_j(\tau)$ for $1 \leq \tau \leq T$.

Model (2). Ornstein-Uhlenbeck process with multiplicative random effect:

$$dX_j(t) = \phi_j X_j(t)dt + \sigma dW_j(t), X_j(0) = x^j. \tag{31}$$

This model satisfies (H1)-(H2) but not (H4). Nevertheless, it is a classical model. We use the explicit solution:

$$X_j(t) = x^j e^{\phi_j t} + \sigma e^{\phi_j t} \int_0^t e^{-\phi_j s} dW_j(s).$$

We use exact simulations of the $X_j(k\delta)$ with $\delta = 0.01$, for $N = 200$ and $T = 400$, with $\sigma = 1$ and $x^j = 0$. The distributions of the opposite of the random effect $-\phi_j$ are: log-normal from normal distribution $\mathcal{N}(1, 0.1^2)$, mixed gamma $0.3\,\Gamma(3, 0.08){+}0.7\,\Gamma(15, 0.07)$, and mixed Gaussian $0.5\,\mathcal{N}(0.5, 0.25^2) + 0.5\,\mathcal{N}(1.8, 0.25^2)$.

To compute the kernel estimator $\hat{f}_h^{(1)}$, we replace $U_j(T)$ and $V_j(T)$ by their usual discrete approximations using discrete data $X_j(k\delta)$. The method works well as illustrated by Figure 2 where 25 estimated densities are plotted. Indeed, the graphs are comparable to those obtained from a sample of directly observed ϕ_j (bold dotted curves). Again, T/N need not be very large.

Model (3): Ornstein-Uhlenbeck process with additive random effect:

$$dX_j(t) = (\phi_j - X_j(t))dt + \sigma dW_j(t), X_j(0) = x^j. \tag{32}$$

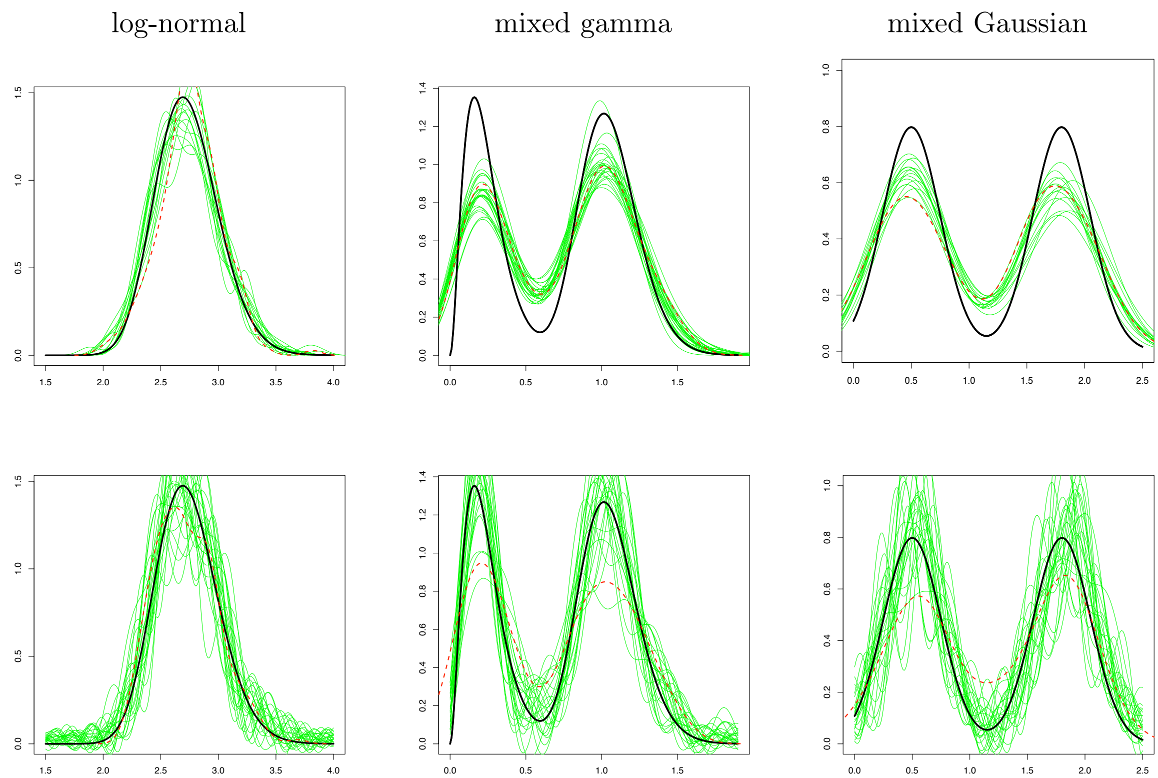

FIGURE 3. Model (3), additive Ornstein-Uhlenbeck process with distribution of the random effects: log-normal (first column), mixed gamma (second column), and mixed Gaussian (third column). Estimated density for 25 independent samples in thin green: kernel estimator ($\hat{f}^{(4)}_{h,\delta}$, first line) and deconvolution estimator ($\tilde{f}_{\tilde{\tau}_\delta,\delta}$, second line). True density in bold black. Estimated density for one sample of (ϕ_j)'s directly observed with standard kernel density estimator in bold dotted red. $N = 200,\ T = 400$.

The solution is:

$$X_j(t) = x^j e^{-t} + \phi_j(1 - e^{-t}) + \sigma e^{-t} \int_0^t e^s dW_j(s).$$

Hence,

$$\mathbb{E}(X_j^2(t)|\phi_j = \varphi) \leq 3(x^j)^2 + \varphi^2 + \frac{\sigma^2}{2t}.$$

Provided that $\mathbb{E}\phi_j^2 < +\infty$, $\mathbb{E}(X_j^2(t)) \leq 3((x^j)^2 + \mathbb{E}\phi_j^2 + \sigma^2)$. All assumptions of Section 3 are satisfied. For this model, we use exact simulations of the $X_j(k\delta)$ with $\delta = 1$ for $N = 200$ and $T = 400$, with $\sigma = 1$ and $x^j = 0$. The distributions of the random effects are: log-normal $\log\mathcal{N}(1,\ 0.1^2)$, mixed gamma $0.4\ \Gamma(3, 0.08) + 0.6\ \Gamma(30, 0.035)$, and mixed Gaussian $0.5\,\mathcal{N}(0.5, 0.25^2) + 0.5\,\mathcal{N}(1.8, 0.25^2)$.

The computation of the kernel estimator $\hat{f}^{(4)}_{h,\delta}$ only requires the discretized terminal variable $\hat{Z}_{j,T}$ while the deconvolution-type estimator $\tilde{f}_{\tilde{\tau}_\delta,\delta}$ requires all the variables $\hat{Z}_{j,\tau}$ for $1 \leq \tau \leq N^2$, see Proposition 7. The deconvolution-type estimator $\tilde{f}_{\tilde{\tau}_\delta,\delta}$ is computed with constant $\tilde{\kappa}$ calibrated through preliminary simulation experiments to $\tilde{\kappa} = 10$.

Figure 3 (25 estimated densities) illustrates the kernel method (first line) and the deconvolution method (second line). The deconvolution method has a great variability but captures well the two modes of the bimodal distributions. On the contrary, the variability band for the kernel estimators is thiner but seems to miss the height of the two modes.

6. Model properties.

In this section, we detail some properties of model (1), give some interpretations of our assumptions and look at links between model (2) and (3).

6.1. **Existence and uniqueness of strong solutions.** We consider N real valued stochastic processes $(X_j(t), t \geq 0)$, $j = 1, \ldots, N$, with dynamics ruled by (1) where ϕ_j are $\mathbb{R}^d$-valued. Consider the assumptions

(A) The functions $(x,\varphi) \to b(x,\varphi)$ and $(x,\varphi) \to \sigma(x,\varphi)$ are C^1 on $\mathbb{R} \times \mathbb{R}^d$, and such that:

$$\exists K > 0, \forall (x,\varphi) \in \mathbb{R} \times \mathbb{R}^d, \quad |b(x,\varphi)| + |\sigma(x,\varphi)| \leq K(1 + |x| + |\varphi|).$$

(B) The functions $(x,\varphi) \to b(x,\varphi)$ and $(x,\varphi) \to \sigma(x,\varphi)$ are C^1 on $\mathbb{R} \times \mathbb{R}^d$, and such that:

$$|b'_x(x,\varphi)| + |\sigma'_x(x,\varphi)| \leq L(\varphi), \quad |b'_\varphi(x,\varphi)| + |\sigma'_\varphi(x,\varphi)| \leq L(\varphi)(1 + |x|),$$

with $\varphi \to L(\varphi)$ continuous.

Under (A) or (B), for all φ, and all $x^j \in \mathbb{R}$, the stochastic differential equation

$$d\,X_j^{\varphi,x^j}(t) = b(X_j^{\varphi,x^j}(t),\varphi)dt + \sigma(X_j^{\varphi,x^j}(t),\varphi)\,dW_j(t), \quad X_j^{\varphi,x^j}(0) = x^j \tag{33}$$

admits a unique strong solution process $(X_j^{\varphi,x^j}(t), t \geq 0)$ adapted to the filtration $(\mathcal{F}_t, t \geq 0)$. Let $C(\mathbb{R}^+,\mathbb{R})$ be the space of continuous functions on $\mathbb{R}^+$, endowed with the Borel σ-field associated with the topology of uniform convergence on compact sets. The distribution of $X_j^{\varphi,x^j}(.)$ is uniquely defined on this space. Now, we can state:

Proposition 8.

- *Under (A) or (B), for $j = 1, \ldots, N$, equation (1) admits a unique solution process $(X_j(t), t \geq 0)$, adapted to the filtration $(\mathcal{F}_t = \sigma(\phi_j, W_j(s), s \leq t, j = 1, \ldots, N), t \geq 0)$ such that the joint process $((X_j(t),\phi_j), t \geq 0)$ is Markov and there exists a measurable functional*

$$(\varphi, x, w_.) \in (\mathbb{R}^d \times \mathbb{R} \times C(\mathbb{R}^+,\mathbb{R})) \to F_.(\varphi, x, w_.) \in C(\mathbb{R}^+,\mathbb{R}) \tag{34}$$

  such that $X_j(.) = F_.(\phi_j, x^j, W_j(.))$.
  Given that $\phi_i = \varphi$, the conditional distribution of $(X_j(t), t \geq 0)$ is identical to the distribution of the process $(X_j^{\varphi,x^j}(t), t \geq 0)$.
  The processes $(X_j(t), t \geq 0), j = 1, \ldots, N$ are independent.
- *Under (A), if moreover, for $k \geq 1$, $\mathbb{E}(|\phi_i|^{2k}) < \infty$, then, for all $T > 0$, $\sup_{t\in[0,T]} \mathbb{E}[X_j(t)]^{2k} < \infty$.*

6.2. **Assumption** (H4)**.** Let us discuss some implications of Assumption (H4). Under (H4), the function b^2/σ^2 is non null. If we combine (H4) with the assumption that $0 < \sigma_0^2 \leq \sigma^2(x) \leq \sigma_1^2, \forall x \in \mathbb{R}$, then, we can assume that both b and σ are positive and satisfy:

(H5) $(i)\quad 0 < c_0 \leq b(x)/\sigma(x) \leq c_1, \quad (ii) \quad 0 < \sigma_0 \leq \sigma(x) \leq \sigma_1, \forall x \in \mathbb{R}$.

Proposition 9. *Let X_i be given by (2), $X_j(0) = x_j$, $j = 1, \dots, N$ under* (H5)*. Then, the process X_i is transient in the sense that:*

$$\mathbb{P}(\lim_{t\to+\infty} X_i(t) = +\infty) = \mathbb{P}(\phi_i > 0), \qquad \mathbb{P}(\lim_{t\to+\infty} X_i(t) = -\infty) = \mathbb{P}(\phi_i < 0).$$

6.3. **Links between the multiplicative and the additive model.** In some cases, the multiplicative model can be transformed into an additive model by a change of function.

Proposition 10. *Assume that the drift function in model (2) is C^1 and satisfies, for some positive constant b_0, $b(x) \geq b_0$ for all x, and set $F(x) = \int_0^x du/b(u)$. Then the process $Y_j(t) = F(X_j(t))$ satisfies*

$$dY_j(t) = (\phi_j - \frac{1}{2}\frac{b' \circ F^{-1}(Y_j(t))}{b^2 \circ F^{-1}(Y_j(t))})dt + \frac{\sigma \circ F^{-1}(Y_j(t))}{b \circ F^{-1}(Y_j(t))} dW_j(t)$$

The result follows from a standard application of the Itô formula. Consequently, (2) can be treated as (3) after the change of function when $b(\cdot)$ is lower bounded b y a positive constant.

7. Concluding remarks

In this paper, we consider N i.i.d. processes $(X_j(t), t \in [0,T])$, $j = 1, \dots, N$, where the dynamics of X_j is described by a SDE including a random effect ϕ_j. The nonparametric density estimation of ϕ_j is investigated in two specific models ((2)-(3)) where the diffusion coefficient does not contain the random effect. The drift term is either multiplicative $(b(x,\varphi) = \varphi b(x))$ or additive $(b(x,\varphi) = \phi + b(x))$. We build kernel and adaptive deconvolution-type estimators and study their L^2-risk as both N and $T = T(N)$ tend to infinity. Under the theoretical condition that $T(N)/N$ tends to infinity rather fast, the estimators attain the usual optimal rates. Nevertheless, numerical simulation results on several models show that, in practice, $T(N)/N$ needs not be very large. Extensions of the present work to models with a more general drift are on going work.

8. Proofs

8.1. **Proof of Proposition 1.** Classically,

$$\begin{aligned} \mathbb{E}(\|f - \hat{f}_h^{(1)}\|^2) &= \|f - \mathbb{E}(\hat{f}_h^{(1)})\|^2 + \mathbb{E}(\|\hat{f}_h^{(1)} - \mathbb{E}(\hat{f}_h^{(1)})\|^2) \\ &\leq 2\|f - f_h\|^2 + 2\|f_h - \mathbb{E}(\hat{f}_h^{(1)})\|^2 + \mathbb{E}(\|\hat{f}_h^{(1)} - \mathbb{E}(\hat{f}_h^{(1)})\|^2). \end{aligned} \tag{35}$$

The last term is the usual variance term and is bounded by

$$\mathbb{E}(\|\hat{f}_h^{(1)} - \mathbb{E}(\hat{f}_h^{(1)})\|^2) = \frac{1}{N}\int \mathrm{Var}(K_h(x - A_{1,T}))dx \leq \frac{\int K^2}{Nh}. \tag{36}$$

The first term is a usual bias term. The specific term is the middle one. We note that $f_h(x) = \mathbb{E}(K_h(x - \phi_1))$ and we apply the Taylor Formula with integral remainder:

$$K_h(x - A_{1,T}) - K_h(x - \phi_1) = \frac{(\phi_1 - A_{1,T})}{h^2}\int_0^1 K'\left(\frac{1}{h}(x - \phi_1 + u(\phi_1 - A_{1,T}))\right)du$$

which yields

$$\begin{aligned}
\|f_h - \mathbb{E}(\hat{f}_h^{(1)})\|^2 &= \int (\mathbb{E}(K_h(x-A_{1,T}) - K_h(x-\phi_1)))^2\, dx \\
&\le \int \mathbb{E}\left[(K_h(x-A_{1,T}) - K_h(x-\phi_1))^2\right] dx \\
&= \mathbb{E}\left[\int (K_h(x-A_{1,T}) - K_h(x-\phi_1))^2\, dx\right] \\
&\le \frac{1}{h^3}\int (K')^2(y)dy\mathbb{E}[(\phi_1 - A_{1,T})^2] = \frac{1}{h^3}\int (K')^2(y)dy\mathbb{E}\left(\left(\frac{M_1(T)}{V_1(T)}\right)^2\right) \\
&\le \frac{1}{h^3}\int (K')^2(y)dy\frac{c_1^2}{Tc_0^4}
\end{aligned}$$

since under (H4),

$$\mathbb{E}\left(\frac{(M_1(T))^2}{(V_1(T))^2}\right) \le \frac{1}{T^2c_0^4}\mathbb{E}[(M_1(T))^2] \le \frac{1}{T}\frac{c_1^2}{c_0^4}.$$

8.2. **Proof of Proposition 2.** Since $\mathbb{E}(\hat{f}_m^{(2)}) = f_m$, we have

$$\mathbb{E}(\|f - \hat{f}_m^{(2)}\|^2) = \|f - f_m\|^2 + \mathbb{E}(\|f_m - \hat{f}_m^{(2)}\|^2). \tag{37}$$

As $\|f_m - \hat{f}_m^{(2)}\|^2 = (2\pi)^{-1}\|f_m^* - (\hat{f}_m^{(2)})^*\|^2$, and the random vectors $(U_{j,k})_{1\le k\le K}$ are independent and identically distributed for $j = 1, \dots, N$ (see (11)), we get

$$\begin{aligned}
\mathbb{E}(\|f_m - \hat{f}_m^{(2)}\|^2) &= \frac{1}{2\pi}\int_{-\pi m}^{\pi m} \mathbb{E}\left(\left|\frac{1}{NK}\sum_{j=1}^{N}\sum_{k=1}^{K}(e^{iuU_{j,k}}e^{u^2/\Delta} - \mathbb{E}(e^{iuU_{j,k}}e^{u^2/(2\Delta)}))\right|^2\right) du \\
&= \frac{1}{2\pi N}\int_{-\pi m}^{\pi m} \mathrm{Var}\left(\frac{1}{K}\sum_{k=1}^{K} e^{iuU_{1,k}}e^{u^2/(2\Delta)}\right) du \\
&= \frac{1}{2\pi NK^2}\int_{-\pi m}^{\pi m}\sum_{k,k'=1}^{K} \mathrm{cov}(e^{iuU_{1,k}}, e^{iuU_{1,k'}})e^{u^2/\Delta}du. \qquad (38)
\end{aligned}$$

Looking at (11), we can see that, for $k = k'$,

$$\mathrm{cov}(e^{iuU_{1,k}}, e^{iuU_{1,k'}}) = 1 - |f^*(u)|^2 e^{-u^2/\Delta}$$

and for $k \ne k'$,

$$\mathrm{cov}(e^{iuU_{1,k}}, e^{iuU_{1,k'}}) = (1 - |f^*(u)|^2)e^{-u^2/\Delta}.$$

Plugging this in (38) yields

$$\mathbb{E}(\|f_m - \hat{f}_m^{(2)}\|^2) \le \frac{\Delta(m)}{NK} + \frac{m}{n} \tag{39}$$

where

$$\Delta(m) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{u^2/\Delta}du.$$

Plugging (39) into (37) gives (13).

We now bound $\Delta(m)$. First

$$\Delta(m) = \frac{\sqrt{\Delta}}{\pi}\int_0^{\pi m/\sqrt{\Delta}} e^{v^2}dv \leq \frac{\sqrt{\Delta}}{\pi}\left(e + \int_1^{\pi m/\sqrt{\Delta}} e^{v^2}dv\right).$$

Integrating by part and using that $v \mapsto e^{v^2}/v^2$ is nondecreasing for $v \geq 1$ imply

$$\begin{aligned}\int_1^{\pi m/\sqrt{\Delta}} e^{v^2}dv &= \left[\frac{e^{v^2}}{2v}\right]_1^{\pi m/\sqrt{\Delta}} + \frac{1}{2}\int_1^{\pi m/\sqrt{\Delta}} e^{v^2}\frac{dv}{v^2} \\ &\leq \frac{\sqrt{\Delta}}{\pi m}e^{\pi^2 m^2/\Delta} - \frac{e}{2}.\end{aligned}$$

Therefore we get

$$\Delta(m) \leq \frac{\sqrt{\Delta}}{\pi}\left(\frac{\sqrt{\Delta}}{\pi m}e^{\pi^2 m^2/\Delta} + \frac{e}{2}\right).$$

As $f_m^* = f^*\mathbf{1}_{[-\pi m, \pi m]}$, $\|f - f_m\|^2 = (2\pi)^{-1}\int_{|u|\geq \pi m}|f^*(u)|^2du$, which ends the proof of (13) and thus of Proposition 2. □

8.3. **Proof of Proposition 3.** The proof is close to the one of Proposition 2 and simpler. The following decomposition holds:

$$\|\hat{f}_\tau - f\|^2 = \|\hat{f}_\tau - f_\tau\|^2 + \|f_\tau - f\|^2.$$

Then,

$$\mathbb{E}(\|\hat{f}_\tau - f_\tau\|^2) = \frac{1}{2\pi N}\int_{-\sqrt{\tau}}^{\sqrt{\tau}} \mathrm{Var}\left(e^{iuA_{1,\tau}}e^{u^2/(2\tau)}\right)du,$$

where

$$\mathrm{Var}(e^{iuA_{1,\tau}}) = 1 - |f^*(u)|^2 e^{-u^2/\tau}.$$

Hence,

$$\mathbb{E}(\|\hat{f}_\tau - f_\tau\|^2) = \frac{1}{2\pi N}\int_{-\sqrt{\tau}}^{\sqrt{\tau}}(1 - |f^*(u)|^2e^{-u^2/\tau})e^{u^2/\tau}du \leq \frac{\sqrt{\tau}}{\pi N}\int_0^1 e^{v^2}dv. \tag{40}$$

8.4. **Proof of Theorem 1.** We have the successive decompositions

$$\|\hat{f}_{\hat{\tau}} - f\|^2 \leq 3(\|\hat{f}_{\hat{\tau}} - \hat{f}_{\hat{\tau}\wedge\tau}\|^2 + \|\hat{f}_{\hat{\tau}\wedge\tau} - \hat{f}_\tau\|^2 + \|\hat{f}_\tau - f\|^2)$$

and

$$\begin{aligned}\|\hat{f}_{\hat{\tau}} - \hat{f}_{\hat{\tau}\wedge\tau}\|^2 &= \|\hat{f}_{\hat{\tau}} - \hat{f}_\tau\|^2\mathbf{1}_{\hat{\tau}\geq\tau} \\ &= \left(\|\hat{f}_{\hat{\tau}} - \hat{f}_\tau\|^2 - \mathrm{pen}(\hat{\tau})\right)\mathbf{1}_{\hat{\tau}\geq\tau} + \mathrm{pen}(\hat{\tau})\mathbf{1}_{\hat{\tau}\geq\tau} \\ &\leq \hat{\Gamma}_\tau + \mathrm{pen}(\hat{\tau}) \\ \|\hat{f}_{\hat{\tau}\wedge\tau} - \hat{f}_\tau\|^2 &= \|\hat{f}_{\hat{\tau}} - \hat{f}_\tau\|^2\mathbf{1}_{\hat{\tau}\leq\tau} \\ &= \left(\|\hat{f}_{\hat{\tau}} - \hat{f}_\tau\|^2 - \mathrm{pen}(\tau)\right)\mathbf{1}_{\hat{\tau}\leq\tau} + \mathrm{pen}(\tau)\mathbf{1}_{\hat{\tau}\leq\tau} \\ &\leq \hat{\Gamma}_{\hat{\tau}} + \mathrm{pen}(\tau).\end{aligned}$$

This yields

$$\begin{aligned}\|\hat{f}_{\hat{\tau}} - f\|^2 &\leq 3(\hat{\Gamma}_\tau + \mathrm{pen}(\hat{\tau}) + \hat{\Gamma}_{\hat{\tau}} + \mathrm{pen}(\tau) + \|\hat{f}_\tau - f\|^2) \\ &\leq 6(\hat{\Gamma}_\tau + \mathrm{pen}(\tau)) + 3\|\hat{f}_\tau - f\|^2.\end{aligned}$$

Then, for $k \geq \tau$:

$$\|\hat{f}_k - \hat{f}_\tau\|^2 \leq 3(\|\hat{f}_\tau - f_\tau\|^2 + \|\hat{f}_k - f_k\|^2 + \|f_k - f_\tau\|^2)$$

Note that

$$\begin{aligned}\|f_k - f_\tau\|^2 &= \frac{1}{2\pi}\|f_k^* - f_\tau^*\|^2 = \frac{1}{2\pi}\int_{\sqrt{\tau}\leq|u|\leq\sqrt{k}} |f^*(u)|^2 du \\ &\leq \frac{1}{2\pi}\int_{|u|\geq\sqrt{\tau}} |f^*(u)|^2 du = \|f - f_\tau\|^2.\end{aligned}$$

Moreover, as the Fourier transforms of $\hat{f}_\tau - f_\tau$ and $f - f_\tau$ have disjoint supports,

$$\|\hat{f}_\tau - f\|^2 = \|\hat{f}_\tau - f_\tau\|^2 + \|f_\tau - f\|^2.$$

It follows that, for $k \geq \tau$:

$$\|\hat{f}_k - \hat{f}_\tau\|^2 \leq 3(\|\hat{f}_k - f_k\|^2 + \|\hat{f}_\tau - f\|^2),$$

which in turn implies that

$$\hat{\Gamma}_\tau \leq 3 \max_{\tau\leq k\leq N^2}(\|\hat{f}_k - f_k\|^2 - \mathrm{pen}(k)/3) + 3\|\hat{f}_\tau - f\|^2.$$

Therefore, gathering inequalities, we get

$$\|\hat{f}_{\hat{\tau}} - f\|^2 \leq 18 \max_{\tau\leq k\leq N^2}(\|\hat{f}_k - f_k\|^2 - \mathrm{pen}(k)/3) + 6\|\hat{f}_\tau - f\|^2 + 6\mathrm{pen}(\tau)$$

and

$$\begin{aligned}\mathbb{E}(\|\hat{f}_{\hat{\tau}} - f\|^2) &\leq 18\sum_{k=\tau}^{N^2}\mathbb{E}(\|\hat{f}_k - f_k\|^2 - \mathrm{pen}(k)/3) + 6\mathbb{E}(\|\hat{f}_\tau - f\|^2) + 6\mathrm{pen}(\tau) \\ &\leq 18\sum_{k=\tau}^{N^2}\mathbb{E}(\|\hat{f}_k - f_k\|^2 - \mathrm{pen}(k)/3) + 6\|f_\tau - f\|^2 \\ &\quad +6\frac{1+\int_0^1 e^{v^2}dv}{\pi}\frac{\sqrt{\tau}}{N} + 6\mathrm{pen}(\tau).\end{aligned}$$

Consequently

$$\mathbb{E}(\|\hat{f}_{\hat{\tau}} - f\|^2) \leq 6\|f_\tau - f\|^2 + 6\left(1 + \frac{1+\int_0^1 e^{v^2}dv}{\pi\kappa}\right)\mathrm{pen}(\tau) + 18\sum_{k=\tau}^{N^2}\mathbb{E}(\|\hat{f}_k - f_k\|^2 - \frac{\mathrm{pen}(k)}{3}).$$

Now, we use the Talagrand Inequality to prove the Lemma

Lemma 1. *Under the Assumptions of Theorem 1*

$$\sum_{k=\tau}^{N^2}\mathbb{E}(\|\hat{f}_k - f_k\|^2 - \mathrm{pen}(k)/3) \leq \frac{C}{N}.$$

This and Proposition 3 gives the result of Theorem 1. □

Proof of Lemma 1. We have that $\|\hat{f}_k - f_k\|^2 = \sup_{t\in S_{\sqrt{k}},\|t\|=1}|\nu_N(t)|^2$ where $S_m = \{t \in L^2, \mathrm{supp}(t^*) = [-m, m]\}$ and

$$\nu_N(t) = \frac{1}{2\pi}\langle t^*, (\hat{f}_k - f_k)^*\rangle = \frac{1}{2\pi N}\sum_{j=1}^{N}\int_{-\sqrt{k}}^{\sqrt{k}} t^*(-u)(e^{iuA_{j,k}}e^{u^2/(2k)} - f^*(u))du.$$

Note that:

$$\nu_N(t) = \frac{1}{N}\sum_{j=1}^{N}(\psi_t(A_{j,k}) - \mathbb{E}(\psi_t(A_{j,k}))) \quad \text{with} \quad \psi_t(x) = \frac{1}{2\pi}\int_{-\sqrt{k}}^{\sqrt{k}} t^*(-u)e^{iux}e^{u^2/(2k)}du.$$

The class of functions $\{\psi_t, t \in S_{\sqrt{k}}, \|t\| = 1\}$ is uniformly bounded as follows.

$$\|\psi_t\|_\infty \leq \frac{1}{\sqrt{2\pi}}\|t\|\left(\int_{-\sqrt{k}}^{\sqrt{k}} e^{u^2/k}du\right)^{1/2} \leq \sqrt{e/\pi}\, k^{1/4} := M.$$

By inequality (40), we get that:

$$\mathbb{E}\left(\sup_{t\in S_{\sqrt{k}}, \|t\|=1} |\nu_N(t)|^2\right) = \mathbb{E}(\|\hat{f}_k - f_k\|^2) \leq c\sqrt{k}/N := H^2,$$

with $c = \int_0^1 e^{v^2}dv/\pi$. At last, let us determine the bound v.

$$\begin{aligned}
& 4\pi^2 \sup_{t\in S_{\sqrt{k}}, \|t\|=1} \mathrm{Var}\psi_t(A_{j,k}) \\
\leq\ & \sup_{t\in S_{\sqrt{k}}, \|t\|=1} \mathbb{E}\left(\iint t^*(u)t^*(-v)e^{i(u-v)A_{j,k}}e^{(u^2+v^2)/(2k)}dudv\right) \\
\leq\ & \sup_{t\in S_{\sqrt{k}}, \|t\|=1} \left(\iint t^*(u)t^*(-v)f^*(u-v)e^{(u^2+v^2+(u-v)^2)/(2k)}dudv\right) \\
\leq\ & \left(\int_{-\sqrt{k}}^{\sqrt{k}}\int_{-\sqrt{k}}^{\sqrt{k}} |f^*(u-v)|^2 e^{3(u^2+v^2)/k}dudv\right)^{1/2} \\
\leq\ & \left(2\sqrt{k}e^6\int_{-2\sqrt{k}}^{2\sqrt{k}} |f^*(z)|^2dz\right)^{1/2} \\
\leq\ & 2\sqrt{\pi}e^3k^{1/4}\|f\| := 4\pi^2\, v
\end{aligned}$$

Thus applying the Talagrand Inequality yields with $\epsilon^2 = 1/2$ that:

$$\mathbb{E}\left(\sup_{t\in S_{\sqrt{k}}, \|t\|=1} |\nu_N(t)|^2 - 4H^2\right) \leq C_1\left(\frac{k^{1/4}}{N}e^{-C_2k^{1/4}} + \frac{k^{1/2}}{N^2}e^{-C_3\sqrt{N}}\right),$$

where C_1, C_2, C_3 are positive constants. Thus for $\mathrm{pen}(k)/3 = 4H^2$, we get

$$\begin{aligned}
\sum_{k=\tau}^{N^2}\mathbb{E}(\|\hat{f}_k - f_k\|^2 - \mathrm{pen}(k)/3) &= \sum_{k=\tau}^{N^2}\mathbb{E}\left(\sup_{t\in S_{\sqrt{k}}, \|t\|=1} |\nu_N(t)|^2 - 2H^2\right) \\
&\leq \frac{C_1}{N}\sum_{k=\tau}^{N^2}\left(k^{1/4}e^{-C_2k^{1/4}} + e^{-C_3k^{1/4}}\right) \leq \frac{C_4}{N}.
\end{aligned}$$

We can choose the constant κ such that $\kappa \geq 12c$. □

8.5. **Proof of Proposition 4.** We have

$$\begin{aligned}\mathbb{E}(\|\hat{f}_m^{(3)} - f\|^2) &= \mathbb{E}(\|\hat{f}_m^{(3)} - \mathbb{E}(\hat{f}_m^{(3)}) + \mathbb{E}(\hat{f}_m^{(3)}) - f_m + f_m - f\|^2) \\ &= \|f_m - f\|^2 + \|\mathbb{E}(\hat{f}_m^{(3)}) - f_m\|^2 + \mathbb{E}(\|\hat{f}_m^{(3)} - \mathbb{E}(\hat{f}_m^{(3)})\|^2)\end{aligned}$$

The term $\|f_m - f\|^2 = \frac{1}{2\pi}\int_{|u|\geq \pi m} |f^*(u)|^2 du$ is the usual bias term. The additional bias term $\|\mathbb{E}(\hat{f}_m^{(3)}) - f_m\|^2$ is bounded by

$$\begin{aligned}\|\mathbb{E}(\hat{f}_m^{(3)}) - f_m\|^2 &= \frac{1}{2\pi}\int_{-\pi m}^{\pi m} |\frac{1}{N}\sum_{j=1}^{N}\left(\mathbb{E}(e^{iuA_{j,T}}) - \mathbb{E}(e^{iu\phi_j})\right)|^2 du \\ &= \frac{1}{2\pi}\int_{-\pi m}^{\pi m} |\left(\mathbb{E}(e^{iuA_{1,T}} - e^{iu\phi_1})\right)|^2 du\end{aligned}$$

By the Taylor formula,

$$e^{iuA_{1,T}} - e^{iu\phi_1} = e^{iu\phi_1}\left(iu\frac{M_1(T)}{V_1(T)}e^{iu\xi_{j,T}}\right).$$

Thus

$$\|\mathbb{E}(\hat{f}_m^{(3)}) - f_m\|^2 \leq \frac{1}{2\pi}\int_{-\pi m}^{\pi m} u^2\mathbb{E}\left(\left(\frac{M_1(T)}{V_1(T)}\right)^2\right) du$$

Then, under (H4),

$$\mathbb{E}\left(\frac{(M_j(T))^2}{(V_j(T))^2}\right) \leq \frac{1}{T^2c_0^4}\mathbb{E}[(M_j(T))^2] \leq \frac{1}{T}\frac{c_1^2}{c_0^4}.$$

We obtain

$$\|\mathbb{E}(\hat{f}_m^{(3)}) - f_m\|^2 \leq \frac{\pi^2 c_1^2}{3c_0^4}\frac{m^3}{T}$$

Using the Parseval equality, the variance term can be bounded as follows

$$\begin{aligned}\mathbb{E}(\|\hat{f}_m^{(3)} - \mathbb{E}(\hat{f}_m^{(3)})\|^2) &= \frac{1}{2\pi}\mathbb{E}\left(\int_{-\pi m}^{\pi m} |\frac{1}{N}\sum_{j=1}^{N}(e^{iuA_{j,T}} - \mathbb{E}(e^{iuA_{j,T}}))|^2 du\right) \\ &= \frac{1}{2\pi}\int_{-\pi m}^{\pi m} Var\left(\frac{1}{N}\sum_{j=1}^{N} e^{iuA_{j,T}}\right) du \leq \frac{m}{N}\end{aligned}$$

Finally, we obtain the bound (20).

If $b = \sigma$,

$$\mathbb{E}(\hat{f}_m^{(3)}(x)) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{-iux} f^*(u)e^{-u^2/(2T)}du.$$

The risk of the estimator, using repeatedly Parseval equality, is

$$\begin{aligned}\mathbb{E}(\|f - \hat{f}_m^{(3)}\|^2) &= \|f - f_m\|^2 + \|f_m - \mathbb{E}(\hat{f}_m^{(3)})\|^2 + \mathbb{E}(\|\hat{f}_m^{(3)} - \mathbb{E}(\hat{f}_m^{(3)})\|^2) \\ &\leq \|f - f_m\|^2 + \frac{1}{2\pi}\int_{-\pi m}^{\pi m} |f^*(u)|^2|e^{-u^2/(2T)} - 1|^2 du + \frac{m}{N} \\ &\leq \frac{1}{2\pi}\int_{|u|\geq \pi m} |f^*(u)|^2 du + \frac{1}{8\pi T^2}\int_{-\pi m}^{\pi m} |u|^4|f^*(u)|^2 du + \frac{m}{N}, \qquad (41)\end{aligned}$$

which gives the second bound of Proposition 4. □

8.6. **Proof of Proposition 5.** Equations (35) and (36) hold with $\hat{f}_h^{(4)}$ instead of $\hat{f}_h^{(1)}$ and $Z_{1,T}$ instead of $A_{1,T}$. Now we study $\|f_h - \mathbb{E}(\hat{f}_h^{(4)})\|^2$, still considering that $f_h(x) = \mathbb{E}(K_h(x-\phi_1))$. We apply the Taylor Formula with integral remainder:

$$\begin{aligned} K_h(x - Z_{1,T}) - K_h(x-\phi_1) &= \frac{(\phi_1 - Z_{1,T})}{h^2} K'\left(\frac{x-\phi_1}{h}\right) \\ &\quad + \frac{(\phi_1 - Z_{1,T})^2}{h^3}\int_0^1 (1-u)K"\left(\frac{1}{h}(x-\phi_1+u(\phi_1 - Z_{1,T}))\right) du. \end{aligned}$$

Recall that ϕ_1 is $\mathcal{F}_0$-measurable, thus $\mathbb{E}((\phi_1 - Z_{1,T})K'(\frac{x-\phi_1}{h})) = 0$, and we obtain, using successively the Cauchy Schwarz and the Burkholder-Davis-Gundy inequalities,

$$\begin{aligned} \|f_h - \mathbb{E}(\hat{f}_h^{(4)})\|^2 &= \int (\mathbb{E}(K_h(x-Z_{1,T}) - K_h(x-\phi_1)))^2\, dx \\ &\leq \int \mathbb{E}\left[(K_h(x-Z_{1,T}) - K_h(x-\phi_1))^2\right] dx \\ &= \mathbb{E}\left[\int (K_h(x-Z_{1,T}) - K_h(x-\phi_1))^2\, dx\right] \\ &\leq \frac{1}{3h^5}\int (K")^2(y)dy\, \mathbb{E}\left[\frac{\left(\int_0^T \sigma(X_1(s))dW(s)\right)^4}{T^4}\right] \\ &\leq \frac{c}{h^5T^4}\|K"\|^2 \mathbb{E}\left[\left(\int_0^T \sigma^2(X_1(s))ds\right)^2\right] \\ &\leq \frac{c}{h^5T^2}\|K"\|^2\sigma_1^4. \end{aligned}$$

8.7. **Proof of Proposition 6.** We know that $\exp(U_t - \langle U\rangle_t/2)$ is a martingale if U is a martingale and $\mathbb{E}(e^{\langle U\rangle_t/2}) < \infty$. As $\sigma^2(x) \leq \sigma_1^2$ for all x, this implies that

$$\exp\left(i\frac{u}{\Delta}\int_{(k-1)\Delta}^{k\Delta} \sigma(X_j(s))dW(s) + \frac{u^2}{2\Delta^2}\int_{(k-1)\Delta}^{k\Delta}\sigma^2(X_j(s))ds\right)$$

has conditional expectation 1 given $\mathcal{F}_{(k-1)\Delta}$. Since $e^{iu\phi_j}$ is $\mathcal{F}_{(k-1)\Delta}$-measurable, we obtain that

$$\begin{aligned} &\mathbb{E}\left(e^{iuY_{j,k}} e^{\frac{u^2}{2\Delta^2}\int_{(k-1)\Delta}^{k\Delta}\sigma^2(X_j(s))ds}\right) \\ =\ &\mathbb{E}\left(e^{iu\phi_j}\mathbb{E}\left(e^{i\frac{u}{\Delta}\int_{(k-1)\Delta}^{k\Delta}\sigma(X_j(s))dW(s) + \frac{u^2}{2\Delta^2}\int_{(k-1)\Delta}^{k\Delta}\sigma^2(X_j(s))ds}|\mathcal{F}_{(k-1)\Delta}\right)\right) \\ =\ &\mathbb{E}(e^{iu\phi_1}) = f^*(u). \end{aligned}$$

Therefore $\mathbb{E}(\hat{f}_m^{(5)}(x)) = f_m(x)$ and we decompose the risk in the two usual terms

$$\mathbb{E}(\|\hat{f}_m^{(5)} - f\|^2) = \|f - f_m\|^2 + \mathbb{E}(\|\hat{f}_m^{(5)} - f_m\|^2). \tag{42}$$

Then we have to compute

$$\mathbb{E}(\|\hat{f}_m^{(5)} - f_m\|^2) = \frac{1}{2\pi}\mathbb{E}\left(\int_{-\pi m}^{\pi m}\left|\frac{1}{NK}\sum_{j=1}^N\sum_{k=1}^K (e^{iuY_{j,k}}e^{\frac{u^2}{2\Delta^2}\int_{(k-1)\Delta}^{k\Delta}\sigma^2(X_j(s))ds} - f^*(u))\right|^2 du\right).$$

Then using that the terms under expectation are centered, we compute a variance with independent variables with respect to the index j:

$$\mathbb{E}(\|\hat{f}_m^{(5)} - f_m\|^2) = \frac{1}{2\pi N}\int_{-\pi m}^{\pi m} \mathbb{E}\left(\left|\frac{1}{K}\sum_{k=1}^{K}(e^{iuY_{1,k}}e^{\frac{u^2}{2\Delta^2}\int_{(k-1)\Delta}^{k\Delta}\sigma^2(X_1(s))ds} - f^*(u))\right|^2\right) du.$$

Let us denote by $M_k = \int_{(k-1)\Delta}^{k\Delta}\sigma(X_1(s))dW(s)$ and $\langle M\rangle_k = \int_{(k-1)\Delta}^{k\Delta}\sigma^2(X_1(s))ds$. Now, for $k<\ell$, by conditioning as follows:

$$\begin{aligned} &\operatorname{cov}\left(e^{iu(\phi_1+M_k)+u^2\langle M\rangle_k/2}, e^{iu(\phi_1+M_\ell)+u^2\langle M\rangle_\ell/2}\right)\\ =\;&\mathbb{E}\left(e^{iuM_k+u^2\langle M\rangle_k/2}\mathbb{E}(e^{iuM_\ell+u^2\langle M\rangle_\ell/2}|\mathcal{F}_{(\ell-1)\Delta})\right) - |f^*(u)|^2\end{aligned}$$

we get that

$$\operatorname{cov}\left(e^{iu(\phi_1+M_k)+u^2\langle M\rangle_k/2}, e^{iu(\phi_1+M_\ell)+u^2\langle M\rangle_\ell/2}\right) = 1-|f^*(u)|^2.$$

We obtain

$$\begin{aligned}\mathbb{E}(\|\hat{f}_m^{(5)} - f_m\|^2) &= \frac{1}{2\pi N}\int_{-\pi m}^{\pi m}\frac{1}{K^2}\left[\sum_{k=1}^{K}(e^{\frac{u^2}{\Delta^2}\int_{(k-1)\Delta}^{k\Delta}\sigma^2(X_1(s))ds} - |f^*(u)|^2) + \sum_{k\neq\ell}(1-|f^*(u)|^2)\right] du\\ &\leq \frac{1}{2\pi NK}\int_{-\pi m}^{\pi m} e^{\frac{u^2\sigma_1^2}{\Delta^2}}du + \frac{m}{N}.\end{aligned}$$

Gathering the last inequality and (42) gives the result of Proposition 6. □

8.8. **Proof of Theorem 2.** The proof follows the steps of Theorem 1. We prove that

$$\mathbb{E}(\|\tilde{f}_\tau - f\|^2) \leq \|f-f_\tau\|^2 + \frac{\sqrt{\tau}}{\pi N}\int_0^1 e^{\sigma_1^2 v^2}dv.$$

Then we obtain

$$\mathbb{E}(\|\tilde{f}_{\tilde{\tau}} - f\|^2) \leq 6\|f_\tau - f\|^2 + 6\left(1+\frac{1+\int_0^1 e^{\sigma_1^2v^2}dv}{\pi\tilde{\kappa}}\right)\widetilde{\mathrm{pen}}(\tau) + 18\sum_{k=\tau}^{N^2}\mathbb{E}(\|\tilde{f}_k - f_k\|^2 - \frac{\widetilde{\mathrm{pen}}(k)}{3}).$$

Then we have to prove

$$\mathbb{E}\left(\sum_{k=\tau}^{N^2}\mathbb{E}(\|\tilde{f}_k - f_k\|^2 - \frac{\widetilde{\mathrm{pen}}(k)}{3})\right) \leq \frac{C'}{N}.$$

To this end, we define

$$\tilde{\nu}_N(t) = \frac{1}{N}\sum_{j=1}^{N}(\psi_t(\tilde{A}_{j,k}) - \mathbb{E}(\psi_t(\tilde{A}_{j,k}))) \quad\text{with}\quad \tilde{A}_{j,k} = (Z_{j,k}, \int_0^k \sigma^2(X_j(s))ds)$$

and

$$\psi_t(x,y) = \frac{1}{2\pi}\int_{-\sqrt{k}}^{\sqrt{k}} t^*(-u)e^{iux+\frac{u^2}{2k^2}y}du\;\mathbb{1}_{0\leq y\leq \sigma_1^2 k}.$$

We find $\|\psi_t\|_\infty \leq \tilde{M} := \sqrt{e^{\sigma_1^2}/\pi}k^{1/4}$,

$$\left(\sup_{t\in S_{\sqrt{k}}, \|t\|=1}|\tilde{\nu}_N(t)|^2\right) \leq \tilde{c}\sqrt{k}/N := \tilde{H}^2$$

with $\tilde{c} = \int_0^1 e^{\sigma_1^2 v^2} dv/\pi$. The difference with the proof of Theorem 1 is that, here, we get nothing better than $\tilde{v} := N\tilde{H}^2$. This is why we take $\epsilon^2 = 3\log(N)/K_1$ in the Talagrand Inequality. Then, choosing $\widetilde{\text{pen}}(k)/3 \geq (1+2\epsilon^2)\tilde{H}^2$ gives the result and the value of the constant $\tilde{\kappa}$. □

8.9. **Proof of Proposition 7.** For the fisrt point, the proof is very close to the one of Proposition 5. The only difference is that the bias term includes an additional term due to the approximation of $Z_{j,T}$ by $\hat{Z}_{j,T}$. We have:

$$K_h(x - \hat{Z}_{j,T}) - K_h(x - Z_{j,T}) = \frac{Z_{j,T} - \hat{Z}_{j,T}}{h^2} \int_0^1 K'(\frac{x - \hat{Z}_{j,T} + u(Z_{j,T} - \hat{Z}_{j,T})}{h})du.$$

Using the Cauchy-Schwarz inequality, we get:

$$\int dx \left(\mathbb{E}(K_h(x - \hat{Z}_{j,T}) - K_h(x - Z_{j,T}))\right)^2 \leq \frac{\mathbb{E}(Z_{j,T} - \hat{Z}_{j,T})^2}{h^3} \int (K'(y))^2 dy.$$

It remains to study

$$Z_{j,T} - \hat{Z}_{j,T} = \frac{1}{T}\sum_{k=1}^{n} \int_{(k-1)\delta}^{k\delta} (b(X_j(s)) - b(X_j((k-1)\delta)))ds.$$

Using again the Cauchy-Schwarz inequality and the fact that b is Lipschitz, say with constant L, yields:

$$\mathbb{E}(Z_{j,T} - \hat{Z}_{j,T})^2 \leq \frac{L^2}{T}\sum_{k=1}^{n} \int_{(k-1)\delta}^{k\delta} \left(X_j(s) - X_j((k-1)\delta)\right)^2 ds.$$

As

$$X_j(s) - X_j((k-1)\delta) = \int_{(k-1)\delta}^{s} (\phi_j + b(X_j(u))\, du + \int_{(k-1)\delta}^{s} \sigma(X_j(u))dW_j(u),$$

we obtain:

$$\mathbb{E}\left(X_j(s) - X_j((k-1)\delta)\right)^2 \leq 4\delta^2(\mathbb{E}\phi_j^2 + \sup_{s\geq 0}\mathbb{E}[b^2(X_j(s))] + 2\delta\sigma_1^2.$$

Thus, $\mathbb{E}(Z_{j,T} - \hat{Z}_{j,T})^2 \leq C\delta$ for some constant C which depends neither on T nor on δ.

Second, we have:

$$\mathbb{E}(\|\tilde{f}_{\tau,\delta} - f\|^2) \leq 2\mathbb{E}(\|\tilde{f}_{\tau,\delta} - \tilde{f}_\tau\|^2) + 2\mathbb{E}(\|\tilde{f}_\tau - f\|^2).$$

Thus, we only deal with the additional term:

$$\mathbb{E}(\|\tilde{f}_{\tau,\delta} - \tilde{f}_\tau\|^2) \leq \frac{1}{2\pi}\int_{-\sqrt{\tau}}^{\sqrt{\tau}} \mathbb{E}|e^{iu\hat{Z}_{1,\tau} + \frac{u^2}{2\tau^2}\sum_{k=1}^{[\tau/\delta]}\delta\sigma^2(X_1((k-1)\delta))} - e^{iuZ_{1,\tau} + \frac{u^2}{2\tau^2}\int_0^\tau \sigma^2(X_1(s)ds}|^2 du.$$

For $\delta < 1 \leq \tau$, we can prove as above:

$$\mathbb{E}(Z_{1,\tau} - \hat{Z}_{1,\tau})^2 + \frac{1}{\tau^2}\mathbb{E}(\sum_{k=1}^{[\tau/\delta]} \delta\sigma^2(X_1((k-1)\delta)) - \int_0^\tau \sigma^2(X_1(s))ds)^2 \leq C\delta$$

for some constant C which depends neither on τ nor on δ. Therefore,

$$\mathbb{E}|e^{iu\hat{Z}_{1,\tau} + \frac{u^2}{2\tau^2}\sum_{k=1}^{[\tau/\delta]}\delta\sigma^2(X_1((k-1)\delta)))} - e^{iuZ_{1,\tau} + \frac{u^2}{2\tau^2}\int_0^\tau \sigma^2(X_1(s))ds}|^2 \leq C\delta(e^{u^2\sigma_1^2/\tau} + \frac{u^4}{4\tau^2}e^{2u^2\sigma_1^2/\tau}),$$

which, integrating with respect to u, gives the result. □

8.10. **Proof of Proposition 8.** . Consider the two-dimensional SDE:

$$\begin{aligned} dX_j(t) &= b(X_j(t),\phi_j(t))dt + \sigma(X_j(t),\phi_j(t))dW_i(t), \quad X_j(0)=x^j, \\ d\phi_j(t) &= 0, \quad \phi_i(0)=\phi_j. \end{aligned}$$

This clarifies the Markov property of the joint process $(X_j(t),\phi_j)$ once we have proved existence and unicity of a strong solution. Moreover, the random effect ϕ_j thus appears as an unobserved initial condition.

- Assumption (A) standardly implies that the above two-dimensional SDE admits a unique strong solution and that there exists a functional F such that $X_j(.) = F_.(\phi_j, x^j, W_j(.))$ where $F_. : \mathbb{R}^d \times \mathbb{R} \times C(\mathbb{R}^+,\mathbb{R}) \to C(\mathbb{R}^+,\mathbb{R})$ is measurable (see *e.g.* Karatzas and Shreve (1997) p.310).
  The joint process $((X_j(t),\phi_j(t)\equiv\phi_j), t\geq 0)$ is Markov. By the Markov property, for all φ, x^j, the conditional distribution of $(X_j(t),\phi_j), t\geq 0)$ given $\phi_j=\varphi$ is the distribution of
  $$X_j^{\varphi,x^j}(.) = F_.(\varphi, x^j, W_j(.)).$$
  As $(\phi_j, W_j(.))$ are independent, the processes $(X_j(.))$ are independent. As (ϕ_j, x^j) is the initial condition, the moment result follows.

- Assumption (A) does not cover the case of (2). This is why we also consider (B). Under (B), we proceed in several steps which are classically used to prove regularity w.r.t. an initial condition.
  Under (B), for all φ, equation (33) admits a unique strong solution. Therefore, there exists a measurable functional
  $$(x, w_.) \in \mathbb{R}\times C(\mathbb{R}^+,\mathbb{R}) \to F_.(\varphi, x, w_.) \in C(\mathbb{R}^+,\mathbb{R}) \tag{43}$$
  such that $X_j^{\varphi,x^j}(.) = F_.(\varphi, x^j, W_j)$ is the unique strong solution of (33). Moreover, as the initial condition x^j is deterministic, it holds that, for all integer $k\geq 1$ and all $T>0$,
  $$\mathbb{E}\sup_{t\in[0,T]}(X_j^{\varphi,x^j}(t))^{2k} < +\infty.$$
  We now prove that (43) is measurable as a function of $(\varphi, x, w.)$.
  Step 1
  Let H be a compact subset of $\mathbb{R}^d$, we prove that for all $x^j\in\mathbb{R}$ and all $T>0$,
  $$\sup_{\varphi\in H}\mathbb{E}(\sup_{u\leq T}(X_j^{\varphi,x^j}(u))^{2k} := C(T,H) < +\infty. \tag{44}$$
  By equation (33), for all $\varphi\in H$ and $t\leq T$,
  $$X_j^{\varphi,x^j}(t) = x^j + \int_0^t b(X_j^{\varphi,x^j}(s),\varphi)ds + \int_0^t \sigma(X_j^{\varphi,x^j}(s),\varphi)dW_j(s)$$
  By (B), we have, for all φ, x and $k\geq 1$,
  $$b^{2k}(x,\varphi) + \sigma^{2k}(x,\varphi) \leq L(k,\varphi)\ (1+x^{2k}),$$
  where $\varphi\to L(k,\varphi)$ is continuous. To ease notations, let us set, $X(t) = X_j^{\varphi,x^j}(t), x = x^j$. We proceed as in Ikeda and Watanabe (1981,Theorem 2.4, p.163). Using the Hölder inequality, the Burholder-Davis-Gundy inequality, we get
  $$\mathbb{E}\sup_{u\leq t}(X(u))^{2k} \leq C(k)(x^{2k} + L_T(k,\varphi)\int_0^t (1+\mathbb{E}\sup_{u\leq s}(X(u))^{2k})ds,$$

where $C(k)$ is a constant and $\varphi \to L_T(k,\varphi)$ is continuous. We conclude by the Gronwall lemma that, for all $t \leq T$,

$$\mathbb{E} \sup_{u \leq t} (X(u))^{2k} \leq C_T(k,\varphi),$$

where $\varphi \to C_T(k,\varphi)$ is continuous. Thus, we get (44).

Step 2

We prove that

$$\varphi \to X_j^{\varphi,x^j}(.) \tag{45}$$

is continuous as a function $\mathbb{R}^d \to C(\mathbb{R}^+,\mathbb{R})$.

Let H be a compact convex subset of $\mathbb{R}^d$ and set, for $\varphi,\varphi' \in H$,

$$S_t(\varphi,\varphi') = \mathbb{E} \sup_{u \leq t} \left(X_j^{\varphi,x^j}(u) - X_j^{\varphi',x^j}(u) \right)^{2k}.$$

We have:

$$X_j^{\varphi,x^j}(t) - X_j^{\varphi',x^j}(t) = \int_0^t \left(b(X_j^{\varphi,x^j}(s),\varphi) - b(X_j^{\varphi',x^j}(s),\varphi') \right) ds$$

$$+ \int_0^t \left(\sigma(X_j^{\varphi,x^j}(s),\varphi) - \sigma(X_j^{\varphi',x^j}(s),\varphi') \right) dW_j(s).$$

By (B), we have for $s \leq T$,

$$|b(X_j^{\varphi,x^j}(s),\varphi) - b(X_j^{\varphi',x^j}(s),\varphi')| + |\sigma(X_j^{\varphi,x^j}(s),\varphi) - \sigma(X_j^{\varphi',x^j}(s),\varphi')|$$

$$\leq L(\varphi) \left(|X_j^{\varphi,x^j}(s) - X_j^{\varphi',x^j}(s)| + |\varphi - \varphi'|(1 + \sup_{u \leq T} |X_j^{\varphi',x^j}(u)|) \right)$$

Now, we proceed as in Step 1. We use the Hölder inequality, the Burkholder-Davis-Gundy inequality, the result (44) of Step 1 and finally the Gronwall lemma to obtain that, for all $t \leq T$,

$$S_t(\varphi,\varphi') \leq |\varphi - \varphi'|^{2k} C_T(k,H),$$

for (another) constant $C_T(k,H)$. Now, choosing $2k > d$, we can conclude by the Kolmogorov continuity of sample paths theorems (see *e.g.* Revuz and Yor, 1991, Theorem 2.1, p.25), that

$$\omega_T(\delta) = \sup_{\varphi,\varphi' \in H} \sup_{t \leq T} |X_j^{\varphi,x^j}(t) - X_j^{\varphi',x^j}(t)|$$

tends to 0 *a.s.* as δ tends to 0. This gives (45).

Step 3

As (43) is measurable for all φ, Step 2 standardly implies that

$$(\varphi, x, w_.) \in \mathbb{R}^d \times \mathbb{R} \times C(\mathbb{R}^+,\mathbb{R}) \to F_.(\varphi, x, w_.) \in C(\mathbb{R}^+,\mathbb{R})$$

is measurable.

The conclusion of Proposition 8 follows.

□

8.11. **Proof of Proposition 9.** For the proof, we omit the index j in the notations. Consider a fixed value $\varphi > 0$ and introduce the process $Y_0(t)$ given by:

$$dY_0(t) = \varphi\sigma_0 c_0 dt + \sigma(Y_0(t))dW(t), \quad Y_0(0) = x.$$

As $\varphi > 0$, $\varphi b(x) \geq \varphi\sigma_0 c_0$ for all x. By the comparison theorem for one-dimensional SDEs (see *e.g.* Ikeda and Watanabe 1981, p.352), it holds that $X^{\varphi,x}(t) \geq Y_0(t)$ for all $t \geq 0$. Thus,

$$\begin{aligned} \frac{X^{\varphi,x}(t)}{t} &\geq \frac{x}{t} + \varphi\sigma_0 c_0 + \frac{1}{t}\int_0^t \sigma(Y_0(s))dW(s) \\ &= \frac{x}{t} + \varphi\sigma_0 c_0 + \frac{\int_0^t \sigma(Y_0(s))dW(s)}{\int_0^t \sigma^2(Y_0(s))ds}\,\frac{\int_0^t \sigma^2(Y_0(s))ds}{t} \end{aligned}$$

As $\int_0^t \sigma^2(Y_0(s))ds \geq \sigma_0^2 t$, $\int_0^{+\infty} \sigma^2(Y_0(s))ds = +\infty$, so

$$\frac{\int_0^t \sigma(Y_0(s))dW(s)}{\int_0^t \sigma^2(Y_0(s))ds} \to 0, \quad \text{a.s..}$$

As $0 < \frac{\int_0^t \sigma^2(Y_0(s))ds}{t} \leq \sigma_1^2$, we deduce that, a.s., $\liminf_{t\to+\infty} \frac{X^{\varphi,x}(t)}{t} \geq \varphi\sigma_0 c_0$, hence $\lim_{t\to+\infty} X^{\varphi,x}(t) = +\infty$.
Analogously, for $\varphi < 0$, $\mathbb{P}(\lim_{t\to+\infty} X^{\varphi,x}(t) = -\infty) = 1$. For $\varphi = 0$, $X^{0,x}(.)$ is a martingale such that $< X^{0,x} >_{+\infty} = \int_0^{+\infty} \sigma(X^{0,x}(s))ds = +\infty$ a.s.. Hence, $\mathbb{P}(\liminf_{t\to+\infty} X^{0,x}(t) = -\infty, \limsup_{t\to+\infty} X^{0,x}(t) = +\infty) = 1$. Noting that

$$\mathbb{P}(\lim_{t\to+\infty} X(t) = +\infty | \phi = \varphi) = \mathbb{P}(\lim_{t\to+\infty} X^{\varphi,x}(t) = +\infty) = 1_{(\varphi>0)},$$

we get that

$$\mathbb{P}(\lim_{t\to+\infty} X(t) = +\infty) = \mathbb{P}(\phi > 0).$$

The other property follows analogously. □

8.12. **Talagrand Inequality.** The following result follows from the Talagrand concentration inequality given in Klein and Rio (2005) and arguments in Birgé and Massart (1998) (see the proof of their Corollary 2 page 354).

Lemma 2. *(Talagrand Inequality) Let $Y_1, \ldots, Y_n$ be independent random variables, let $\nu_{n,Y}(\psi) = (1/n)\sum_{i=1}^n [\psi(Y_i) - \mathbb{E}(\psi(Y_i))]$ and let $\mathcal{F}$ be a countable class of uniformly bounded measurable functions. Then for $\epsilon^2 > 0$*

$$\mathbb{E}\Big[\sup_{\psi\in\mathcal{F}} |\nu_{n,Y}(\psi)|^2 - 2(1+2\epsilon^2)H^2\Big]_+ \leq \frac{4}{K_1}\left(\frac{v}{n}e^{-K_1\epsilon^2\frac{nH^2}{v}} + \frac{98M^2}{K_1 n^2 C^2(\epsilon^2)}e^{-\frac{2K_1 C(\epsilon^2)\epsilon}{7\sqrt{2}}\frac{nH}{M}}\right),$$

with $C(\epsilon^2) = \sqrt{1+\epsilon^2} - 1$, $K_1 = 1/6$, and

$$\sup_{\psi\in\mathcal{F}} \|\psi\|_\infty \leq M, \quad \mathbb{E}\Big[\sup_{\psi\in\mathcal{F}} |\nu_{n,Y}(\psi)|\Big] \leq H, \ \sup_{\psi\in\mathcal{F}} \frac{1}{n}\sum_{k=1}^n \mathrm{Var}(\psi(Y_k)) \leq v.$$

By standard density arguments, this result can be extended to the case where $\mathcal{F}$ is a unit ball of a linear normed space, after checking that $f \mapsto \nu_n(\psi)$ is continuous and $\mathcal{F}$ contains a countable dense family.

References

[1] Birgé, L., and Massart, P. (1998) Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli* **4**, 329-375.

[2] Comte, F., Rozenholc, Y. and Taupin, M.-L. (2006) Penalized contrast estimator for adaptive density deconvolution. *Canadian Journal of Statistics* **34**, 3, 431-452.

[3] Comte, F. and Samson, A. (2012). Nonparametric estimation of random effects densities in linear mixed-effects model, *Journal of Nonparametric Statistics*, **24**, 4, 951-975.

[4] Davidian, M., and Giltinan, D.M. (1995). *Nonlinear Models for Repeated Measurement Data*, New York: Chapman and Hall.

[5] Delattre M., Genon-Catalot V. and Samson A. (2012). Maximum likelihood estimation for stochastic differential equations with random effects. *Scandinavian Journal of Statistics*, to appear.

[6] Ditlevsen, S. and De Gaetano, A. (2005). Mixed effects in stochastic differential equation models. *REVSTAT* **3**, 137-153.

[7] Donnet, S. and Samson, A. (2008) Parametric inference for mixed models defined by stochastic differential equations. *ESAIM Probab. Stat.* **12**, 196-218.

[8] Donnet S, Samson A. (2010). EM algorithm coupled with particle filter for maximum likelihood parameter estimation of stochastic differential mixed-effects models. *Preprint MAP5 2010-24*.

[9] Goldenshluger, A., Lepski, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39**, 1608-1632.

[10] Ikeda, N. and Watanabe, S. (1989) *Stochastic differential equations and diffusion processes.* Second edition. North-Holland Mathematical Library, 24. North-Holland Publishing Co., Amsterdam; Kodansha, Ltd., Tokyo.

[11] Karatzas, I. and Shreve, S. (1997). *Brownian motion and stochastic calculus.* Springer-Verlag, New-York.

[12] Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33**, 1060-1077.

[13] Nie, L. (2006) Strong consistency of the maximum likelihood estimator in generalized linear and nonlinear mixed-effects models. *Metrika* **63**, 123-143.

[14] Nie, L. (2007). Convergence rate of the mle in generalized linear and nonlinear mixed-effects models: Theory and applications. *J. Statist. Plann Inf.* **137**, 1787-1804.

[15] Nie, L. and Yang, M. (2005). Strong consistency of MLE in nonlinear mixed-effects models with large cluster size. *Sankhya* **67**, 736-763.

[16] Overgaard, R., Jonsson, N., Tornøe, C. and Madsen, H. (2005). Non-linear mixed effects models with stochastic differential equations: Implementation of an estimation algorithm. *J. Pharmacokinet. Pharmacodyn.* **32**, 85-107.

[17] Picchini, U., De Gaetano, A. and Ditlevsen, S. (2010). Stochastic differential mixed-effects models. *Scand. J. Statist.* **37**, 67-90.

[18] Picchini, U. and Ditlevsen, S. (2011). Practical estimation of high dimensional stochastic differential mixed-effects models. *Comput. Statist. Data Anal.* **55**, 1426-1444.

[19] Pinheiro, J.C. and Bates, D.M. (2000). *Mixed-Effects Models in S and S-PLUS.* New York, Springer.

[20] Revuz, D. and Yor, M. (1999). *Continuous martingales and Brownian motion.* Third edition. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 293. Springer-Verlag, Berlin.

[21] Tsybakov, A. B. (2009). *Introduction to nonparametric estimation.* Revised and extended from the 2004 French original. Translated by Vladimir Zaiats. Springer Series in Statistics. Springer, New York.