

## Maximum likelihood estimation for stochastic differential equations with random effects

Maud Delattre, Valentine Genon-Catalot, Adeline Samson

► **To cite this version:**

Maud Delattre, Valentine Genon-Catalot, Adeline Samson. Maximum likelihood estimation for stochastic differential equations with random effects. *Scandinavian Journal of Statistics*, Wiley, 2013, 40 (2), pp.322-343. <10.1111/j.1467-9469.2012.00813.x/abstract>. <hal-00650844>

**HAL Id: hal-00650844**

**<https://hal.archives-ouvertes.fr/hal-00650844>**

Submitted on 12 Dec 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Maximum likelihood estimation for stochastic differential equations with random effects

Running headline: Mixed stochastic differential equations

Maud Delattre<sup>1</sup>, Valentine Genon-Catalot<sup>2</sup> and Adeline Samson<sup>2\*</sup>

<sup>1</sup> Laboratoire Mathématiques, Université Paris Sud, France

<sup>2</sup> UMR CNRS 8145, Laboratoire MAP5, Université Paris Descartes, Sorbonne Paris Cité, France

## Abstract

We consider  $N$  independent stochastic processes  $(X_i(t), t \in [0, T_i])$ ,  $i = 1, \dots, N$ , defined by a stochastic differential equation with drift term depending on a random variable  $\phi_i$ . The distribution of the random effect  $\phi_i$  depends on unknown parameters which are to be estimated from the continuous observation of the processes  $X_i$ . We give the expression of the exact likelihood. When the drift term depends linearly on the random effect  $\phi_i$  and  $\phi_i$  has Gaussian distribution, an explicit formula for the likelihood is obtained. We prove that the maximum likelihood estimator is consistent and asymptotically Gaussian, when  $T_i = T$  for all  $i$  and  $N$  tends to infinity. We discuss the case of discrete observations. Estimators are computed on simulated data for several models and show good performances even when the length time interval of observations is not very large.

**Key Words:** Asymptotic normality, consistency, maximum likelihood estimator, mixed-effects models, stochastic differential equations.

## 1 Introduction

Statistical analysis of data collected over time on a series of subjects requires to account for both the intra-individual variability, *i.e.* the variability occurring

---

\*Corresponding author

*Email addresses:* maud.delattre@math.u-psud.fr (Maud Delattre), valentine.genon-catalot@parisdescartes.fr (Valentine Genon-Catalot), adeline.samson@parisdescartes.fr (Adeline Samson)

within the dynamics of each individual over time, and the variability existing between subjects. Modeling of such data goes through mixed-effects models which are very popular in the biomedical field (Davidian and Giltinan, 1995; Pinheiro and Bates, 2000). In mixed-effects stochastic differential equations (SDEs), the model for each individual set of data is given by a SDE, thus modeling the intra-individual variability in the data, and the parameters of each individual SDE are random variables, thus handling the variability between subjects. A major area of application for mixed effects SDEs is in pharmacokinetic/pharmacodynamic modeling, where they have been introduced as an alternative to the classical ODE-based models (Ditlevsen and De Gaetano, 2005; Overgaard *et al.*, 2005; Donnet and Samson, 2008). SDEs with random effects have also been proposed for neuronal data (Picchini *et al.*, 2010).

Maximum likelihood estimation (MLE) of the parameters of the random effects, also called population parameters, is generally not straightforward as the likelihood function can rarely be expressed in a closed-form. Approximations of the likelihood have been proposed, based on linearization (Beal and Sheiner, 1982) or Laplace's approximation (Wolfinger, 1993). Alternative methods have also been developed such as the SAEM algorithm (Kuhn and Lavielle, 2004). Maximum likelihood estimation in SDEs with random effects has been tackled in a few papers. Ditlevsen and De Gaetano (2005) show that in the specific case of a mixed-effects Brownian motion with drift, the likelihood function can be explicitly derived, leading to explicit parameters estimators. For general mixed SDEs, approximations of the likelihood have been proposed (Picchini *et al.*, 2010; Picchini and Ditlevsen, 2011).

For theoretical properties of the MLE in the context of mixed effects models, the main contribution to our knowledge is due to (Nie and Yang, 2005; Nie, 2006, 2007) and covers the asymptotic properties of the MLE for the population parameters under several asymptotic frameworks, depending on whether the number of subjects and/or the number of observations per subject goes to infinity. Nie's results are nevertheless based on a series of technical assumptions, which may be uneasy to check.

In the present work, we focus on mixed-effects SDEs with drift term depending on random effects and diffusion term without random effects. More precisely, we consider  $N$  real valued stochastic processes  $(X_i(t), t \geq 0)$ ,  $i = 1, \dots, N$ , with dynamics ruled by the following SDEs:

$$dX_i(t) = b(X_i(t), \phi_i)dt + \sigma(X_i(t))dW_i(t), \quad X_i(0) = x^i, \quad i = 1, \dots, N, \quad (1)$$

where  $(W_1, \dots, W_N)$  are  $N$  independent Wiener processes,  $\phi_1, \dots, \phi_N$  are  $N$  *i.i.d.*  $\mathbb{R}^d$ -valued random variables,  $(\phi_1, \dots, \phi_N)$  and  $(W_1, \dots, W_N)$  are independent and  $x^i, i = 1, \dots, N$  are known real values. The diffusion coefficient  $\sigma : \mathbb{R} \rightarrow \mathbb{R}$  is a known real-valued function. The drift function  $b(x, \varphi)$  is a known function defined on  $\mathbb{R} \times \mathbb{R}^d$  and real-valued. Each process  $(X_i(t))$  represents an individual and the random vector  $\phi_i$  represents the random effect of individual  $i$ . We assume that the random variables  $\phi_1, \dots, \phi_N$  have a common distribution  $g(\varphi, \theta)d\nu(\varphi)$  on  $\mathbb{R}^d$ , where  $\theta$  is an unknown parameter belonging to a set  $\Theta \subset \mathbb{R}^p$  and, for all  $\theta$ ,  $g(\varphi, \theta)$  is a density w.r.t. a dominating measure  $\nu$  on  $\mathbb{R}^d$ . Below, we denote by  $\theta_0$  the true value of the parameter. The process  $(X_i(t))$  is continuously observed on a time interval  $[0, T_i]$  with  $T_i > 0$  given. Our aim is to estimate the parameters  $\theta$  of the density of the random effects from the observations  $\{X_i(t), 0 \leq t \leq T_i, i = 1, \dots, N\}$ . We introduce assumptions ensuring that the models (1) are well-defined together with the exact likelihood function. Then, we focus on the special case of one-dimensional linear Gaussian random effects, *i.e.*  $b(x, \phi_i) = \phi_i b(x)$ , where  $b$  is a known real function and  $\phi_i$  is Gaussian. It turns out in this case that the likelihood has a simple and explicit expression depending on  $\theta$  and the sufficient statistics:

$$U_i = \int_0^{T_i} \frac{b(X_i(s))}{\sigma^2(X_i(s))} dX_i(s), \quad V_i = \int_0^{T_i} \frac{b^2(X_i(s))}{\sigma^2(X_i(s))} ds, \quad i = 1, \dots, N.$$

For the asymptotic study, the main difficulties are encountered to obtain specific moment properties of the random variables  $(U_i, V_i)$ , and to prove identifiability. We prove the consistency and the asymptotic normality of the exact MLE as  $N$  tends to infinity and give the expression of the Fisher information matrix. The results are extended to Gaussian multidimensional linear random effects. The present likelihood theory is derived from continuous observations of the  $X_i^t$ s. In practice, one rather disposes of discrete observations on the time interval  $[0, T_i]$ . Thus we suggest to discretize the r.v.'s  $U_i, V_i$  in the expression of estimators and we show that under conditions on the discretization step, and thus on the number of observations per subject, the asymptotic properties of the estimates based on continuous observations are preserved. Our simulations are presented within the framework of discretely observed stochastic processes.

The paper is organized as follows. Section 2 introduces the notations and assumptions. In section 3, we make the likelihood function explicit. In sections 4 and 5, we show that the strong consistency and the asymptotic normality of

the MLE when the model includes a Gaussian one-dimensional and a Gaussian multi-dimensional random effect respectively. The impact of discretization on the estimators is detailed in section 6. A simulation study is presented in section 7. Concluding remarks are given in section 8. Proofs are gathered in appendix.

## 2 Model, assumptions and notations

Consider  $N$  real valued stochastic processes  $(X_i(t), t \geq 0)$ ,  $i = 1, \dots, N$ , with dynamics ruled by (1). The processes  $(W_1, \dots, W_N)$  and the r.v.'s  $\phi_1, \dots, \phi_N$  are defined on a common probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . We introduce assumptions ensuring that the processes (1) are well defined and allowing to compute the exact likelihood of our observations. Consider the filtration  $(\mathcal{F}_t, t \geq 0)$  defined by  $\mathcal{F}_t = \sigma(\phi_i, W_i(s), s \leq t, i = 1, \dots, N)$ . 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. Moreover, the random variables  $\phi_i$  are  $\mathcal{F}_0$ -measurable.

(H1) (i) The function  $(x, \varphi) \rightarrow b(x, \varphi)$  is  $C^1$  on  $\mathbb{R} \times \mathbb{R}^d$ , and such that:

$$\exists K > 0, \forall (x, \varphi) \in \mathbb{R} \times \mathbb{R}^d, \quad b^2(x, \varphi) \leq K(1 + x^2 + |\varphi|^2),$$

(ii) The function  $\sigma(\cdot)$  is  $C^1$  on  $\mathbb{R}$  and

$$\forall x \in \mathbb{R}, \quad \sigma^2(x) \leq K(1 + x^2).$$

Under (H1), for all  $\varphi$ , the stochastic differential equation

$$dX_i^\varphi(t) = b(X_i^\varphi(t), \varphi)dt + \sigma(X_i^\varphi(t))dW_i(t), \quad X_i^\varphi(0) = x^i, \quad (2)$$

admits a unique strong solution process  $(X_i^\varphi(t), t \geq 0)$  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  $\sigma$ -field associated with the topology of uniform convergence on compact sets. The distribution of  $X_i^\varphi(\cdot)$  is uniquely defined on this space. Moreover, as  $x^i$  is deterministic, for all integer  $k$ , all  $\varphi$  and all  $t \geq 0$ ,

$$\sup_{s \leq t} \mathbb{E}[X_i^\varphi(s)]^{2k} < +\infty. \quad (3)$$

For the observed processes, we have the following result.

**Proposition 1.** *Under (H1), for  $i = 1, \dots, N$ , equation (1) admits a unique solution process  $(X_i(t), t \geq 0)$ , adapted to the filtration  $(\mathcal{F}_t, t \geq 0)$ . Given*

that  $\phi_i = \varphi$ , the conditional distribution of  $(X_i(t), t \geq 0)$  is identical to the distribution of the process  $(X_i^\varphi(t), t \geq 0)$ . The processes  $(X_i(t), t \geq 0), i = 1, \dots, N$  are independent.

If for  $k \geq 1, \mathbb{E}|\phi_i|^{2k} < \infty$ , then for all  $T > 0, \sup_{t \in [0, T]} \mathbb{E}[X_i(t)]^{2k} < \infty$ .

### 3 Likelihood

We introduce the canonical model associated with the observations. Let  $C_{T_i}$  denote the space of real continuous functions  $(x(t), t \in [0, T_i])$  defined on  $[0, T_i]$ , endowed with the  $\sigma$ -field  $\mathcal{C}_{T_i}$  associated with the topology of uniform convergence on  $[0, T_i]$ . Under (H1), we introduce the distribution  $Q_\varphi^{x_i, T_i}$  on  $(C_{T_i}, \mathcal{C}_{T_i})$  of  $(X_i^\varphi(t), t \in [0, T_i])$  given by (2). On  $\mathbb{R}^d \times C_{T_i}$ , let  $P_\theta^i = g(\varphi, \theta) d\nu(\varphi) \otimes Q_\varphi^{x_i, T_i}$  denote the joint distribution of  $(\phi_i, X_i(\cdot))$  and let  $Q_\theta^i$  denote the marginal distribution of  $(X_i(t), t \in [0, T_i])$  on  $(C_{T_i}, \mathcal{C}_{T_i})$ . From now on, we denote by  $(\phi_i, X_i(\cdot))$  the canonical process of  $\mathbb{R}^d \times C_{T_i}$ . Let us consider the following assumptions.

(H2) For  $i = 1, \dots, N$ , and for all  $\varphi, \varphi'$

$$Q_\varphi^{x_i, T_i} \left( \int_0^{T_i} \frac{b^2(X_i^\varphi(t), \varphi')}{\sigma^2(X_i^\varphi(t))} dt < +\infty \right) = 1.$$

(H3) For  $f = \frac{\partial b}{\partial \varphi_j}, j = 1, \dots, d$ , there exist  $c > 0$  and some  $\gamma \geq 0$  such that

$$\sup_{\varphi \in \mathbb{R}^d} \frac{|f(x, \varphi)|}{\sigma^2(x)} \leq c(1 + |x|^\gamma).$$

**Proposition 2.** Assume (H1)-(H3) and let  $\varphi_0 \in \mathbb{R}^d$ .

- The distributions  $Q_\varphi^{x_i, T_i}$  are absolutely continuous w.r.t.  $Q^i := Q_{\varphi_0}^{x_i, T_i}$  with density:

$$\frac{dQ_\varphi^{x_i, T_i}}{dQ^i}(X_i) = L_{T_i}(X_i, \varphi) = e^{\ell_{T_i}(X_i, \varphi)} \quad \text{with} \quad \ell_{T_i}(X_i, \varphi) =$$

$$\int_0^{T_i} \frac{b(X_i(s), \varphi) - b(X_i(s), \varphi_0)}{\sigma^2(X_i(s))} dX_i(s) - \int_0^{T_i} \frac{b^2(X_i(s), \varphi) - b^2(X_i(s), \varphi_0)}{2\sigma^2(X_i(s))} ds,$$

where  $(X_i = X_i(s), s \leq T_i)$  denotes the canonical process of  $C_{T_i}$  given by  $(X_i(s)(x) = x(s), s \leq T_i)$ .

- The function  $\varphi \rightarrow L_{T_i}(X_i, \varphi)$  admits a continuous version  $Q^i$ -a.s. and  $(X_i, \varphi) \rightarrow L_{T_i}(X_i, \varphi)$  is measurable on  $(C_{T_i} \times \mathbb{R}^d, \mathcal{C}_{T_i} \otimes \mathcal{B}(\mathbb{R}^d))$ .

**Remark 1.** For a given drift function  $b(x, \varphi)$ , it is often possible to check directly that  $\varphi \rightarrow L_{T_i}(X, \varphi)$  is continuous even if (H3) is not fulfilled. Then, the joint measurability follows.

To simplify notations, we assume that there is one value  $\varphi_0$  such that  $b(x, \varphi_0) \equiv 0$ . Thus, we can choose the dominating measure  $Q^i = Q_{\varphi_0}^{x_i, T_i}$  which is the distribution of (2) with nul drift. Formula  $L_{T_i}(X_i, \varphi)$  simplifies into:

$$L_{T_i}(X_i, \varphi) = \exp \left( \int_0^{T_i} \frac{b(X_i(s), \varphi)}{\sigma^2(X_i(s))} dX_i(s) - \frac{1}{2} \int_0^{T_i} \frac{b^2(X_i(s), \varphi)}{\sigma^2(X_i(s))} ds \right). \quad (4)$$

By independence of the individuals,  $P_\theta = \otimes_{i=1}^N P_\theta^i$  is the distribution of  $(\phi_i, X_i(\cdot))$ ,  $i = 1, \dots, N$  on the product space  $\prod_{i=1}^N \mathbb{R}^d \times C_{T_i}$  and  $Q_\theta = \otimes_{i=1}^N Q_\theta^i$  is the distribution of the whole sample  $(X_i(t), t \in [0, T_i], i = 1, \dots, N)$  on  $C = \prod_{i=1}^N C_{T_i}$ . We now compute the density of  $Q_\theta$  w.r.t.  $Q = \otimes_{i=1}^N Q^i$ . We denote by  $E_\theta$  the expectation w.r.t.  $P_\theta$ .

**Proposition 3.** Assume (H1)-(H3).

- The probability measure  $Q_\theta^i$  admits a density w.r.t.  $Q^i$  equal to:

$$\frac{dQ_\theta^i}{dQ^i}(X_i) = \int_{\mathbb{R}^d} L_{T_i}(X_i, \varphi) g(\varphi, \theta) d\nu(\varphi) := \lambda_i(X_i, \theta).$$

- The distribution  $Q_\theta$  on  $C = \prod_{i=1}^N C_{T_i}$  admits a density given by

$$\frac{dQ_\theta}{dQ}(X_1, \dots, X_N) = \prod_{i=1}^N \lambda_i(X_i, \theta).$$

- The exact likelihood of the whole sample  $(X_i(t), t \in [0, T_i], i = 1, \dots, N)$  is

$$\Lambda_N(\theta) = \prod_{i=1}^N \lambda_i(X_i, \theta). \quad (5)$$

On this general expression, if we can check Nie (2006)'s assumptions, then weak consistency and asymptotic normality of the MLE will follow. However, these assumptions, even when the random effects are Gaussian, are uneasy.

## 4 Gaussian one-dimensional linear random effects

In this section, we consider model (1) with drift  $b(x, \varphi) = \varphi b(x)$  where  $\varphi \in \mathbb{R}$ ,  $b(\cdot), \sigma(\cdot)$  are known functions. In this case, we simplify (H1)-(H2) and assume

that  $b, \sigma$  are  $C^1$  and have linear growth, which implies Proposition 1. And, we assume that  $\int_0^{T_i} b^2(X_i(s))/\sigma^2(X_i(s))ds < \infty$ ,  $Q_\varphi^{x^i, T_i}$ -a.s. for all  $\varphi$ . As  $\varphi \rightarrow L_{T_i}(X_i, \varphi)$  is obviously continuous, (H3) is not required. We also assume that, for  $i = 1, \dots, N$ ,  $T_i = T, x^i = x$  so that the observed processes  $(X_i(t), t \in [0, T]), i = 1, \dots, N$  are *i.i.d.*. Let us introduce:

$$U_i = \int_0^T \frac{b(X_i(s))}{\sigma^2(X_i(s))} dX_i(s), \quad V_i = \int_0^T \frac{b^2(X_i(s))}{\sigma^2(X_i(s))} ds, \quad (6)$$

which are well defined under (H1)-(H2). Hence,

$$\lambda_i(X_i, \theta) = \int_{\mathbb{R}} g(\varphi, \theta) \exp(\varphi U_i - \frac{\varphi^2}{2} V_i) d\nu(\varphi). \quad (7)$$

#### 4.1 Exact likelihood

We propose here to model the random effects distribution by a Gaussian distribution  $\mathcal{N}(\mu, \omega^2)$ , and set  $\theta = (\mu, \omega^2) \in \mathbb{R} \times (0, +\infty)$  for the unknown parameters to be estimated. This choice leads to an explicit exact likelihood.

**Proposition 4.** *Assume that  $g(\varphi, \theta)d\nu(\varphi) = \mathcal{N}(\mu, \omega^2)$ . Then,*

$$\lambda_i(X_i, \theta) = \frac{1}{(1 + \omega^2 V_i)^{1/2}} \exp\left[-\frac{V_i}{2(1 + \omega^2 V_i)} \left(\mu - \frac{U_i}{V_i}\right)^2\right] \exp\left(\frac{U_i^2}{2V_i}\right).$$

The conditional distribution, under  $P_\theta^i$ , of  $\phi_i$  given  $X_i$  is the distribution

$$\mathcal{N}\left(\frac{\mu + \omega^2 U_i}{1 + \omega^2 V_i}, \frac{\omega^2}{1 + \omega^2 V_i}\right).$$

Therefore, the logarithm of the likelihood function (5) is explicitly given by

$$\mathcal{L}_N(\theta) = -\frac{1}{2} \sum_{i=1}^N \log(1 + \omega^2 V_i) - \frac{1}{2} \sum_{i=1}^N \frac{V_i}{1 + \omega^2 V_i} \left(\mu - \frac{U_i}{V_i}\right)^2 + \sum_{i=1}^N \frac{U_i^2}{2V_i}. \quad (8)$$

The derivatives of the log-likelihood (8) are

$$\begin{aligned} \frac{\partial}{\partial \mu} \mathcal{L}_N(\theta) &= \sum_{i=1}^N \left( \frac{U_i}{1 + \omega^2 V_i} - \mu \frac{V_i}{1 + \omega^2 V_i} \right), \\ \frac{\partial}{\partial \omega^2} \mathcal{L}_N(\theta) &= \frac{1}{2} \sum_{i=1}^N \left[ \left( \frac{U_i}{1 + \omega^2 V_i} - \mu \frac{V_i}{1 + \omega^2 V_i} \right)^2 - \frac{V_i}{1 + \omega^2 V_i} \right]. \end{aligned}$$



When  $\omega_0^2$  is known, we obtain the explicit estimator for  $\mu_0$ :

$$\hat{\mu}_N = \frac{\sum_{i=1}^N \frac{U_i}{1+\omega_0^2 V_i}}{\sum_{i=1}^N \frac{V_i}{1+\omega_0^2 V_i}}. \quad (9)$$

When both parameters are unknown, the maximum likelihood estimators of  $\theta_0 = (\mu_0, \omega_0^2)$  are given by the system:

$$\begin{aligned} \hat{\mu}_N &= \left( \sum_{i=1}^N \frac{V_i}{1 + \hat{\omega}_N^2 V_i} \right)^{-1} \left( \sum_{i=1}^N \frac{U_i}{1 + \hat{\omega}_N^2 V_i} \right), \\ \sum_{i=1}^N \left( \hat{\mu}_N - \frac{U_i}{V_i} \right)^2 \frac{V_i^2}{(1 + \hat{\omega}_N^2 V_i)^2} &= \sum_{i=1}^N \frac{V_i}{1 + \hat{\omega}_N^2 V_i}. \end{aligned}$$

**Remark 2.** Note that, when the effect  $\phi_i$  is non random and  $\phi_i \equiv \mu_0$ , the estimator of  $\mu_0$  is standardly given by:

$$\tilde{\mu}_N = \frac{\sum_{i=1}^N U_i}{\sum_{i=1}^N V_i}, \quad (10)$$

which corresponds to  $\omega_0^2 = 0$  in  $\hat{\mu}_N$ .

## 4.2 Preliminary moments properties

For studying the maximum likelihood estimators of  $\theta = (\mu, \omega^2)$ , we need investigate properties of the following random variables:

$$\gamma_i(\theta) = \frac{U_i - \mu V_i}{1 + \omega^2 V_i}, \quad I_i(\omega^2) = \frac{V_i}{1 + \omega^2 V_i}. \quad (11)$$

Indeed under  $Q_\theta$ ,  $(\gamma_i(\theta), I_i(\omega^2))_{i=1, \dots, N}$  are *i.i.d.* and the score function is

$$\frac{\partial}{\partial \mu} \mathcal{L}_N(\theta) = \sum_{i=1}^N \gamma_i(\theta), \quad \frac{\partial}{\partial \omega^2} \mathcal{L}_N(\theta) = \frac{1}{2} \sum_{i=1}^N (\gamma_i^2(\theta) - I_i(\omega^2)). \quad (12)$$

Evidently,  $0 < I_i(\omega^2) \leq 1/\omega^2$  is bounded. By the following lemma, which is crucial for the statistical study, we prove that  $\gamma_i(\theta)$  admits a finite Laplace transform, hence moments of any order.

**Lemma 1.** For all  $\theta = (\mu, \omega^2) \in \mathbb{R} \times (0, +\infty)$ , and all  $u \in \mathbb{R}$ ,

$$E_\theta(\exp(u \frac{U_1}{1 + \omega^2 V_1})) < +\infty.$$

We can now compute some useful moments of functions of  $\gamma_1(\theta), I_1(\omega^2)$ .

**Proposition 5.** For all  $\theta \in \mathbb{R} \times (0, +\infty)$ , the following relations hold:

$$E_\theta(\gamma_1(\theta)) = 0, \quad E_\theta(\gamma_1^2(\theta)) = E_\theta(I_1(\omega^2)), \quad E_\theta(\gamma_1^3(\theta)) = 3E_\theta(\gamma_1(\theta) I_1(\omega^2)),$$

$$E_\theta (\gamma_1^2(\theta) - I_1(\omega^2))^2 = 4E_\theta(\gamma_1^2(\theta)I_1(\omega^2)) - 2E_\theta(I_1^2(\omega^2)).$$

### 4.3 Convergence in distribution of the normalized score function

Based on lemma 1 and Proposition 5, we can state:

**Proposition 6.** For all  $\theta$ , under  $Q_\theta$ , as  $N$  tends to infinity, the random vector

$$\frac{1}{\sqrt{N}} \begin{pmatrix} \frac{\partial}{\partial \mu} \mathcal{L}_N(\theta) \\ \frac{\partial}{\partial \omega^2} \mathcal{L}_N(\theta) \end{pmatrix} = \frac{1}{\sqrt{N}} \begin{pmatrix} \sum_{i=1}^N \gamma_i(\theta) \\ \frac{1}{2} \sum_{i=1}^N (\gamma_i^2(\theta) - I_i(\omega^2)) \end{pmatrix}$$

converges in distribution to  $\mathcal{N}_2(0, \mathcal{I}(\theta))$  and the matrix

$$-\frac{1}{N} \begin{pmatrix} \frac{\partial^2}{\partial \mu^2} \mathcal{L}_N(\theta) & \frac{\partial^2}{\partial \mu \partial \omega^2} \mathcal{L}_N(\theta) \\ \frac{\partial^2}{\partial \mu \partial \omega^2} \mathcal{L}_N(\theta) & \frac{\partial^2}{\partial \omega^2 \partial \omega^2} \mathcal{L}_N(\theta) \end{pmatrix}$$

converges in probability to  $\mathcal{I}(\theta)$  where

$$\mathcal{I}(\theta) = \begin{pmatrix} E_\theta(I_1(\omega^2)) & E_\theta(\gamma_1(\theta)I_1(\omega^2)) \\ E_\theta(\gamma_1(\theta)I_1(\omega^2)) & E_\theta(\gamma_1^2(\theta)I_1(\omega^2)) - \frac{1}{2}E_\theta(I_1^2(\omega^2)) \end{pmatrix} \quad (13)$$

is the covariance matrix of the vector

$$\begin{pmatrix} \gamma_1(\theta) \\ \frac{1}{2}(\gamma_1^2(\theta) - I_1(\omega^2)) \end{pmatrix}.$$

The following corollary holds immediately.

**Corollary 1.** When  $\omega^2 = \omega_0^2$  is known, the explicit estimator  $\hat{\mu}_N$  (9) is consistent and  $\sqrt{N}(\hat{\mu}_N - \mu_0)$  converges in distribution under  $Q_{\mu_0}$  to  $\mathcal{N}(0, 1/E_{\mu_0}(V_1/(1 + \omega_0^2 V_1)))$  where  $1/E_{\mu_0}(V_1/(1 + \omega_0^2 V_1)) \geq \omega_0^2$ .

If  $\phi_1, \dots, \phi_N$  were observed, the MLE of  $\mu_0$  would be  $\bar{\phi} = \frac{1}{N}(\phi_1 + \dots + \phi_N)$  which satisfies  $\sqrt{N}(\bar{\phi} - \mu_0) \sim \mathcal{N}(0, \omega_0^2)$ . As  $\phi_1, \dots, \phi_N$  are not observed, we obtain that the MLE  $\hat{\mu}_N$  has a larger asymptotic variance.

#### 4.4 Consistency and asymptotic normality

When both parameters  $\mu_0, \omega_0^2$  are unknown, we need to introduce additional assumptions to prove identifiability and consistency. Recall that  $Q$  is the distribution on  $C$  such that the canonical processes  $(X_i(s), s \leq T, i = 1, \dots, N)$  are *i.i.d.* and  $X_i$  satisfies the SDE with nul drift:

$$dX_i(t) = \sigma(X_i(t))dW_i(t), \quad X_i(0) = x.$$

We assume that

- (H4) The function  $b(\cdot)/\sigma(\cdot)$  is not constant. Under  $Q$ , the random variable  $(U_1, V_1)$  admits a density  $f(u, v)$  w.r.t. the Lebesgue measure on  $\mathbb{R} \times (0, +\infty)$  which is jointly continuous and positive on an open ball of  $\mathbb{R} \times (0, +\infty)$ .
- (H5) The parameter set  $\Theta$  is a compact subset of  $\mathbb{R} \times (0, +\infty)$ .
- (H6) The true value  $\theta_0$  belongs to  $\mathring{\Theta}$ .
- (H7) The matrix  $\mathcal{I}(\theta_0)$  is invertible (see (13)).

Under smoothness assumptions on functions  $b, \sigma$ , assumption (H4) will be fulfilled by application of Malliavin calculus tools<sup>†</sup>. The case where  $b(\cdot)/\sigma(\cdot)$  is constant is rather simple and is treated separately in Section 7. Assumptions (H5)-(H7) are classical. We first state an identifiability result.

**Proposition 7.** *Set  $K(Q_{\theta_0}^1, Q_{\theta}^1)$  the Kullback information of  $Q_{\theta_0}^1$  w.r.t.  $Q_{\theta}^1$ .*

- (i) *Under (H1)-(H2) and (H4),  $Q_{\theta}^1 = Q_{\theta_0}^1$  implies that  $\theta = \theta_0$ . Hence,  $\theta \rightarrow K(Q_{\theta_0}^1, Q_{\theta}^1)$  admits a unique minimum at  $\theta = \theta_0$ .*
- (ii) *Under (H1)-(H2), the function  $\theta \rightarrow K(Q_{\theta_0}^1, Q_{\theta}^1)$  is continuous on  $\mathbb{R} \times (0, +, \infty)$ .*

We are now able to prove the consistency and asymptotic normality of  $\hat{\theta}_N$ .

**Proposition 8.** *1. Assume (H1)-(H2) and (H4)-(H5). Let  $\hat{\theta}_N$  be a maximum likelihood estimator defined as any solution of  $\mathcal{L}_N(\hat{\theta}_N) = \sup_{\theta \in \Theta} \mathcal{L}_N(\theta)$ . Under  $Q_{\theta_0}$ ,  $\hat{\theta}_N$  converges in probability to  $\theta_0$ .*

<sup>†</sup>If  $\sigma$  and  $f = b/\sigma$  are  $C^\infty$ , if their derivatives of any order greater than 1 are bounded, if  $\sigma$  is bounded below, and if the Lebesgue measure of the set of values  $x$  such that  $f(x)f'(x) = 0$  is zero, then assumption (H4) holds.

2. Assume (H1)-(H2) and (H4)-(H7). The maximum likelihood estimator satisfies, as  $N$  tends to infinity,

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \rightarrow_{\mathcal{D}} \mathcal{N}_2(0, \mathcal{I}^{-1}(\theta_0)).$$

Note that the consistency obtained here is a strong consistency in the sense that any solution of the likelihood equation is consistent.

## 5 Gaussian multidimensional linear random effects

In this section, we extend the previous results to multidimensional linear random effects. Let  $\phi_i = (\phi_i^1, \dots, \phi_i^d)'$  be a  $d$ -dimensional random vector and  $b(x) = (b^1(x), \dots, b^d(x))'$  be a function  $\mathbb{R} \rightarrow \mathbb{R}^d$ . Consider the SDE

$$dX_i(t) = \phi_i' b(X_i(t))dt + \sigma(X_i(t)) dW_i(t), \quad X_i(0) = x. \quad (14)$$

We assume that  $b^1(x), \dots, b^d(x)$  are such that  $b(x, \varphi) = \sum_{j=1}^d \varphi^j b^j(x)$  satisfies (H1)-(H2) and that  $(\phi_i, i = 1, \dots, N)$  are i.i.d. Gaussian vectors, with expectation vector  $\mu$  and covariance matrix  $\Omega \in \mathcal{S}_d(\mathbb{R})$  where  $\mathcal{S}_d(\mathbb{R})$  is the set of positive definite symmetric matrices. The parameter to be estimated is  $\theta = (\mu, \Omega) \in \mathbb{R}^d \times \mathcal{S}_d(\mathbb{R})$ . To compute the likelihood, we introduce the random vectors

$$U_i = \int_0^T \frac{b(X_i(s))}{\sigma^2(X_i(s))} dX_i(s),$$

and the  $d \times d$  random matrices

$$V_i = \int_0^T \frac{b(X_i(s))b'(X_i(s))}{\sigma^2(X_i(s))} ds.$$

The following assumption is now required.

(H8) For  $i = 1, \dots, N$  the matrix  $V_i$  is positive definite  $Q^i$ -a.s. and  $Q_\theta^i$ -a.s. for all  $\theta$ .

If the functions  $(b^j/\sigma^2)$  are not linearly independent, (H8) is not true. Thus, (H8) can be interpreted as ensuring a well-defined dimension of the vector  $\phi_i$ . We deduce the invertibility of matrices involved in the likelihood computation.

**Lemma 2.** Under (H8), the matrices  $V_i + \Omega^{-1}$ ,  $I_d + V_i\Omega$ ,  $I_d + \Omega V_i$  are invertible  $Q^i$ -a.s. and  $Q_\theta^i$ -a.s. for all  $\theta$ .

Then we can compute the likelihood.

**Proposition 9.** Under (H8), set  $R_i^{-1} = (I_d + V_i\Omega)^{-1}V_i$ , we have

$$\lambda_i(X_i, \theta) = \frac{1}{\sqrt{\det(I_d + V_i\Omega)}} \exp \left[ -\frac{1}{2}(\mu - V_i^{-1}U_i)' R_i^{-1}(\mu - V_i^{-1}U_i) \right] \exp \left( \frac{1}{2}U_i' V_i^{-1}U_i \right).$$

The conditional distribution of  $\phi_i$  given  $X_i$  is the Gaussian distribution

$$\mathcal{N}_d \left( (I_d + \Omega V_i)^{-1}\mu + (\Omega^{-1} + V_i)^{-1}V_i, (I_d + \Omega V_i)^{-1}\Omega \right)$$

The likelihood is  $\Lambda_N(\theta) = \prod_{i=1}^N \lambda_i(X_i, \theta)$ .

The score function (respectively a  $d$ -vector and a  $d \times d$  matrix) is given by:

$$\frac{\partial}{\partial \mu} \mathcal{L}_N(\theta) = \sum_{i=1}^N \gamma_i(\theta), \quad \frac{\partial}{\partial \Omega} \mathcal{L}_N(\theta) = \frac{1}{2} \sum_{i=1}^N (\gamma_i(\theta)\gamma_i'(\theta) - I_i(\Omega))$$

where  $\theta = (\mu, \Omega)$  and :

$$\gamma_i(\theta) = (I_d + \Omega V_i)^{-1}(U_i - V_i\mu), \quad I_i(\Omega) = (I_d + \Omega V_i)^{-1}V_i. \quad (15)$$

When  $\Omega_0$  is known, the estimator for  $\mu_0$  is explicit.

Lemma 1 and Propositions 5 and 6 can be readily extended to the multidimensional case.

**Proposition 10.** 1. For all  $\theta = (\mu, \Omega) \in \mathbb{R} \times (0, +\infty)$ , and all  $u \in \mathbb{R}$ ,

$$E_\theta(\exp(u'(I_d + \Omega V_i)^{-1}U_1)) < +\infty.$$

2. For all  $\theta \in \mathbb{R} \times (0, +\infty)$ , the following relations hold:

$$E_\theta(\gamma_1(\theta)\gamma_1'(\theta)) = E_\theta(I_1(\Omega)), \quad E_\theta(\gamma_1(\theta)\gamma_1'(\theta)\gamma_1(\theta)) = 3E_\theta(I_1(\Omega)\gamma_1(\theta)),$$

$$E_\theta(\gamma_1(\theta)) = 0, \quad E_\theta(\gamma_1(\theta)\gamma_1'(\theta) - I_1(\Omega))^2 = 4E_\theta(I_1(\Omega)\gamma_1(\theta)\gamma_1'(\theta)) - 2E_\theta(I_1^2(\Omega)).$$

3. For all  $\theta$ , under  $Q_\theta$ , as  $N$  tends to infinity, the random vector

$$\frac{1}{\sqrt{N}} \begin{pmatrix} \frac{\partial}{\partial \mu} \mathcal{L}_N(\theta) \\ \frac{\partial}{\partial \Omega} \mathcal{L}_N(\theta) \end{pmatrix} = \frac{1}{\sqrt{N}} \begin{pmatrix} \sum_{i=1}^N \gamma_i(\theta) \\ \frac{1}{2} \sum_{i=1}^N (\gamma_i(\theta)\gamma_i'(\theta) - I_i(\Omega)) \end{pmatrix}$$

converges in distribution to  $\mathcal{N}(0, \mathcal{I}(\theta))$  where  $\mathcal{I}(\theta)$  is the covariance matrix

of the vector

$$\begin{pmatrix} \gamma_1(\theta) \\ \frac{1}{2}(\gamma_1(\theta)\gamma_1'(\theta) - I_1(\Omega)) \end{pmatrix}$$

which is also the limit of the observed Fisher information matrix.

The study of  $\widehat{\theta}_N = (\widehat{\mu}_N, \widehat{\Omega}_N)$  can be done as above.

## 6 Discrete data

In this section, we briefly discuss the case of discrete data. Let us assume that we observe synchronously the processes  $X_i(t)$  at times  $t_k^n = t_k = k\frac{T}{n}$ ,  $k = 0, 1, \dots, n$ . To build estimators  $\widehat{\theta}_N^{(n)}$  based on these data, we simply replace the r.v.'s  $U_i, V_i$ ,  $i = 1, \dots, N$  by their discretized versions:

$$U_i^n = \sum_{k=0}^{n-1} \frac{b(X_i(t_k))}{\sigma^2(X_i(t_k))} (X_i(t_{k+1}) - X_i(t_k)), \quad (16)$$

$$V_i^n = \sum_{k=0}^{n-1} \frac{b^2(X_i(t_k))}{\sigma^2(X_i(t_k))} (t_{k+1} - t_k). \quad (17)$$

Looking at the expressions of (8) and its derivatives w.r.t.  $\theta$ , it is enough to study the differences  $U_i - U_i^n$ ,  $V_i - V_i^n$ . We can prove

**Lemma 3.** *Assume that  $b/\sigma$  is bounded and Lipschitz,  $\sigma(\cdot) \geq \epsilon > 0$ ,  $b$  and  $\sigma$  Lipschitz, then for all  $p \geq 1$  and all  $i = 1, \dots, N$ , there exists a constant  $C$  such that*

$$\mathbb{E}_{\theta_0} (|V_i - V_i^n|^p + |U_i - U_i^n|^p) \leq \frac{C}{n^{p/2}}.$$

We deduce

**Proposition 11.** *If  $n \rightarrow +\infty$ , then  $\widehat{\theta}_N - \widehat{\theta}_N^{(n)} = o_{P_{\theta_0}}(1)$ . If  $n = n(N) \rightarrow +\infty$  in such a way that  $\frac{n}{N} \rightarrow +\infty$ , then  $\sqrt{N}(\widehat{\theta}_N - \widehat{\theta}_N^{(n)}) = o_{P_{\theta_0}}(1)$ .*

## 7 Simulation study

Several models are simulated. For each SDE model, 100 datasets are generated with  $N$  subjects on the same time interval  $[0, T]$  and three experimental designs:  $(N = 20, T = 5)$ ,  $(N = 50, T = 5)$  and  $(N = 50, T = 10)$ . The empirical mean and variance of the MLE are computed from the 100 datasets. When possible,  $N^{-1}\mathcal{I}_N^{-1}(\theta_0)$  is also computed and compared with the empirical variance.

### 7.1 When $b(x) = c\sigma(x)$

Let us consider the case where  $b(x) = c\sigma(x)$ , with  $c \neq 0$  known. Then we have

$$V_i = c^2T, \quad U_i = c \int_0^T \frac{dX_i(s)}{\sigma(X_i(s))}. \quad (18)$$

The estimators of  $\mu_0, \omega_0^2$  are simple and explicit:

$$\hat{\mu}_N = \frac{1}{c^2TN} \sum_{i=1}^N U_i = \frac{1}{c^2T} \bar{U}_N, \quad \hat{\omega}_N^2 = \frac{1}{(c^2T)^2} \left( \frac{1}{N} \sum_{i=1}^N (U_i - \bar{U}_N)^2 - c^2T \right).$$

Using that  $U_i = c^2T\phi_i + cW_i(T)$ , an elementary study shows that  $\hat{\mu}_N$  and  $\hat{\omega}_N^2$  are strongly consistent, that  $\sqrt{N}(\hat{\mu}_N - \mu_0)$  has distribution  $\mathcal{N}(0, \omega_0^2 + 1/(c^2T))$  and that  $\sqrt{N}(\hat{\omega}_N^2 - \omega_0^2)$  converges in distribution to  $\mathcal{N}(0, 2(\omega_0^2 + 1/(c^2T))^2)$ . The asymptotic variances of  $\hat{\mu}_N$  and  $\hat{\omega}_N^2$  are increased in comparison with the case of non random effects.

We stress the fact that, whatever the drift function  $b(\cdot)$ , when  $b(\cdot)/\sigma(\cdot)$  is constant, the estimators have the same distribution.

*Example 1.* Consider a mixed-effects Brownian motion with drift

$$dX_i(t) = \phi_i dt + \sigma dW_i(t), \quad X_i(0) = 0,$$

with  $\phi_i \sim \mathcal{N}(\mu, \omega^2)$ . This model is considered by Ditlevsen and De Gaetano (2005) and the estimators are the same. We use two sets of population parameters:  $(\mu = -1, \omega^2 = 1)$  and  $(\mu = 5, \omega^2 = 1)$ . All simulations are performed with a discretization step-size  $\delta = 0.001$  on  $[0, T]$  and  $\sigma = 1$  ( $\sigma$  known). Results, presented in Table 1, are satisfactory overall. Increasing  $N$  improves the accuracy of both estimates  $\hat{\mu}_N$  and  $\hat{\omega}_N^2$ . For  $T = 5$ , both estimates are less biased when the number of subjects is 50 instead of 20. The variance of the estimates is also decreased with larger values of  $N$ . In general the empirical variances coincide with the values of the asymptotic variances. Apparently, increasing  $T$  does not have any significant impact on the properties of the estimates. Additional simulations with other values for  $\mu_0$  and  $\omega_0^2$  have basically shown that the properties of the estimates degrade when  $\omega_0^2$  takes bigger values.

**[Table 1 about here.]**

We have also considered simulations of the mixed-effects geometric Brownian motion where  $b(x) = x$ ,  $\sigma(x) = \sigma x$  and the model where  $b(x) = \sqrt{1+x^2}$ ,

$\sigma(x) = \sigma\sqrt{1+x^2}$ . On our simulated data, the true parameter values were correctly estimated.

## 7.2 General case

*Example 2.* Consider an Ornstein-Uhlenbeck process with one random effect

$$dX_i(t) = \phi_i X_i(t)dt + \sigma dW_i(t), \quad X_i(0) = 0,$$

with  $\phi_i \sim \mathcal{N}(\mu, \omega^2)$ . We separate three situations: *i)*  $\omega^2$  is known ( $\theta = \mu$ ), *ii)*  $\mu$  is known ( $\theta = \omega^2$ ), *iii)* both parameters are unknown ( $\theta = (\mu, \omega^2)$ ). When  $\omega_0^2$  is known, we use the explicit expression of  $\hat{\mu}_N$ . Otherwise, numerical optimization procedures are required for maximizing the log-likelihood with respect to  $\mu$  and  $\omega^2$ . In situations *i)* and *ii)*, the asymptotic variance of the estimate has an explicit expression and is computed. Each individual diffusion is simulated with a discretization step-size  $\delta = 0.001$  on  $[0, T]$  and  $\sigma = 1$ . Several sets of parameter values are used:  $(\mu = -5, \omega^2 = 1)$  and  $(\mu = 10, \omega^2 = 1)$ . Table 2 displays the results of the three inferences *i)*, *ii)* and *iii)*. Results highlight the accuracy of the estimates of both parameters  $\mu$  and  $\omega^2$  whatever the design and the parameter values. In the three considered inference situations, increasing  $N$  leads to smaller bias of the parameter estimates and smaller variances. Moreover, we notice great similarities between  $N^{-1}\mathcal{I}_N^{-1}(\theta_0)$  and the empirical variance of  $\hat{\theta}_N$ , especially when  $N$  is large. Finally, we don't observe any significant impact of  $T$  neither on the bias nor on the variance of the parameter estimates.

**[Table 2 about here.]**

*Example 3.* Consider an Ornstein-Uhlenbeck process with two random effects

$$dX_i(t) = (-\phi_i^1 X_i(t) + \phi_i^2)dt + \sigma dW_i(t), \quad X_i(0) = 0,$$

with  $\phi_i = (\phi_i^1, \phi_i^2)' \sim \mathcal{N}_2(\mu, \Omega)$ ,  $\mu = (\mu_1, \mu_2)'$  and a diagonal matrix  $\Omega$  with components  $(\omega_1^2, \omega_2^2)$ . For this model, assumption (H8) is satisfied. Indeed

$$V_i = \frac{1}{\sigma^2} \begin{pmatrix} \int_0^T X_i^2(s)ds & \int_0^T X_i(s)ds \\ \int_0^T X_i(s)ds & T \end{pmatrix}.$$

Using the equality case in the Cauchy-Schwarz inequality, we see that  $\det(V_i) = 0$  if and only if  $X_i(t) \equiv cste$  on  $[0, T]$  which is impossible. The estimation of



$\theta = (\mu_1, \mu_2, \omega_1^2, \omega_2^2)$  is obtained by optimizing numerically the log-likelihood. Each individual diffusion is simulated with a discretization step-size  $\delta = 0.001$  on  $[0, T]$  and  $\sigma = 1$ . Several sets of parameter values are used:  $(\mu_1 = 0.1, \mu_2 = 1, \omega_1^2 = 0.01, \omega_2^2 = 1)$  and  $(\mu_1 = 0.1, \mu_2 = 1, \omega_1^2 = 0.001, \omega_2^2 = 1)$ . Table 3 displays the results of estimation. Parameters are well estimated although  $\mu_2$  has a larger bias when  $\omega_2^2$  is greater. Biases decrease when  $N$  increases. Again, the influence of  $T$  is small.

[Table 3 about here.]

*Example 4.* Consider the process with single random effect

$$dX_i(t) = \phi_i X_i(t) dt + \sigma \sqrt{1 + X_i(t)^2} dW_i(t), \quad X_i(0) = 0,$$

with  $\phi_i \sim \mathcal{N}(\mu, \omega^2)$ . We obtain the estimate for  $\theta = (\mu, \omega^2)$  by numerical optimization of the log-likelihood with respect to  $\mu$  and  $\omega^2$ . We simulate  $N$  individual diffusions with a discretization step-size  $\delta = 0.001$  on  $[0, T]$ , and several sets of parameter values are used:  $(\mu = -1, \omega^2 = 1)$  and  $(\mu = 5, \omega^2 = 1)$ . Table 4 displays the results of estimation. The results are satisfactory overall, although parameter  $\omega^2$  is estimated with larger bias than parameter  $\mu$ . Bias and empirical variances of the estimates decrease when  $N$  becomes larger.  $T = 10$  leads to better results (smaller bias) for the estimation of  $\omega^2$  than  $T = 5$ .

[Table 4 about here.]

## 8 Concluding remarks

In this paper, we have studied maximum likelihood estimation for *i.i.d.* observations of stochastic differential equations including a random effect in the drift term. When the drift term depends linearly on the random effect, we prove that the likelihood is given by a closed-form formula and that the exact MLE is strongly consistent and asymptotically Gaussian as the number of observed processes tends to infinity.

For the clarity of exposure, we have considered only one-dimensional SDEs, but the theory can be done in the same way for multidimensional SDEs. For a drift term depending linearly on the random effects, the likelihood is still explicit although its formulae may be much more cumbersome.

One could also include non random effects in the drift without much changes.

**Acknowledgements.** We thank Arnaud Gloter for his helpful contribution on assumption (H4).

The corresponding author for negotiations concerning the manuscript is  
Adeline Samson (adeline.samson@parisdescartes.fr)  
Laboratoire MAP5  
Universite Paris Descartes  
45 rue des St Peres  
75006 Paris  
France

## References

- Beal, S. and Sheiner, L. (1982). Estimating population kinetics. *Critical Reviews in Biomedical Engineering* **8**, 195–222.
- Davidian, M. and Giltinan, D. (1995). *Nonlinear models to repeated measurement data*. Chapman and Hall.
- Ditlevsen, S. and De Gaetano, A. (2005). Stochastic vs. deterministic uptake of dodecanedioic acid by isolated rat livers. *Bull. Math. Biol.* **67**, 547–561.
- Donnet, S. and Samson, A. (2008). Parametric inference for mixed models defined by stochastic differential equations. *ESAIM P&S* **12**, 196–218.
- Karatsas, I. and Shreve, S. (1997). *Brownian Motion and Stochastic Calculus*. Springer-Verlag.
- Kuhn, E. and Lavielle, M. (2004). Coupling an approximation version of em with an mcmc procedure. *ESAIM Probability and Statistics* **8**, 115–131.
- Lipster, R. and Shiryaev, A. (2001). *Statistics of random processes I : general theory*. Springer.
- Nie, L. (2006). Strong consistency of the maximum likelihood estimator in generalized linear and nonlinear mixed-effects models. *Metrika* **63**, 123–143.
- Nie, L. (2007). Convergence rate of the mle in generalized linear and nonlinear mixed-effects models: Theory and applications. *Journal of Statistical Planning and Inference* **137**, 1787–1804.

- Nie, L. and Yang, M. (2005). Strong consistency of the mle in nonlinear mixed-effects models with large cluster size. *Sankhya: The Indian Journal of Statistics* **67**, 736–763.
- 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.
- Picchini, U., De Gaetano, A. and Ditlevsen, S. (2010). Stochastic differential mixed-effects models. *Scand. J. Statist.* **37**, 67–90.
- Picchini, U. and Ditlevsen, S. (2011). Practicle estimation of high dimensional stochastic differential mixed-effects models. *Computational Statistics & Data Analysis* **55**, 1426–1444.
- Pinheiro, J. and Bates, D. (2000). *Mixed-effect models in S and Splus*. Springer-Verlag.
- Revuz, D. and Yor, M. (1999). *Continuous martingales and Brownian motion*, vol. 293. Springer-Verlag, Berlin, third edn.
- van der Vaart, A. W. (1998). *Asymptotic statistics*, vol. 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge. ISBN 0-521-49603-9; 0-521-78450-6.
- Wolfinger, R. (1993). Laplace’s approximation for nonlinear mixed models. *Biometrika* **80**, 791–795. ISSN 0006-3444.

## 9 Appendix: proofs

*Proof of Proposition 1.* Consider the two-dimensional SDE:

$$\begin{aligned} dX_i(t) &= b(X_i(t), \phi_i(t))dt + \sigma(X_i(t))dW_i(t), & X_i(0) &= x^i, \\ d\phi_i(t) &= 0, & \phi_i(0) &= \phi_i. \end{aligned}$$

Under (H1), the above system admits a unique strong solution and there exists a functional  $F$  such that  $X_i(\cdot) = F(x^i, \phi_i, W_i(\cdot))$  where  $F : \mathbb{R} \times \mathbb{R}^d \times C(\mathbb{R}^+, \mathbb{R}) \rightarrow C(\mathbb{R}^+, \mathbb{R})$  is measurable (see *e.g.* Karatsas and Shreve, 1997, p.310).

Moreover,  $X_i^\varphi(\cdot) = F(x^i, \varphi, W_i(\cdot))$ . By the Markov property of the joint process  $((X_i(t), \phi_i(t) \equiv \phi_i), t \geq 0)$ , the conditional distribution of  $X_i(\cdot)$  given  $\phi_i = \varphi$  is identical to the distribution of  $X_i^\varphi(\cdot)$ . As  $(\phi_i, W_i(\cdot))$  are independent,

the processes  $(X_i(\cdot))$  are independent. As  $(x^i, \phi_i)$  is the initial condition, the moment result follows.  $\square$

*Proof of Proposition 2.* Under (H1)-(H2), the first part is classical (see *e.g.* Lipster and Shiryaev, 2001).

To prove the continuity in  $\varphi$ , two kinds of terms are to be studied. The first is, for a given  $X_i$ , the ordinary integral

$$\varphi \rightarrow \int_0^{T_i} \frac{b^2(X_i(s), \varphi)}{\sigma^2(X_i(s))} ds. \quad (19)$$

Using (H1)-(H3) and the continuity theorem for ordinary integrals, we obtain easily the continuity of (19).

Second, the stochastic integral

$$\varphi \rightarrow \int_0^{T_i} \frac{b(X_i(s), \varphi)}{\sigma^2(X_i(s))} dX_i(s) := I_1(\varphi) + I_2(\varphi), \quad (20)$$

with

$$I_1(\varphi) = \int_0^{T_i} \frac{b(X_i(s), \varphi)}{\sigma^2(X_i(s))} b(X_i(s), \varphi_0) ds,$$

$$I_2(\varphi) = \int_0^{T_i} \frac{b(X_i(s), \varphi)}{\sigma^2(X_i(s))} (dX_i(s) - b(X_i(s), \varphi_0) ds).$$

The function  $\varphi \rightarrow I_1(\varphi)$  is studied using (H1)-(H3) and the continuity theorem for ordinary integrals again. For  $I_2(\varphi)$ , using the Burkholder-Davis-Gundy inequality, we get, using (H1)-(H3):

$$\begin{aligned} \mathbb{E}_{Q^i} (I_2(\varphi) - I_2(\varphi'))^{2k} &\leq C_k \mathbb{E}_{Q^i} \left[ \int_0^{T_i} \frac{(b(X_i(s), \varphi) - b(X_i(s), \varphi'))^2}{\sigma^2(X_i(s))} ds \right]^k \\ &\leq C_k |\varphi - \varphi'|^{2k} \mathbb{E}_{Q^i} \left[ \int_0^{T_i} c^2 K (1 + |X_i(s)|^{2\gamma}) (1 + |X_i(s)|)^2 ds \right]^k \\ &\leq C(k, \gamma) |\varphi - \varphi'|^{2k} \int_0^{T_i} (1 + \mathbb{E}_{Q^i} (|X_i(s)|^{2(\gamma+1)})) ds. \end{aligned}$$

Using (3) and choosing  $2k > d$ , the Kolmogorov criterion (see *e.g.* Revuz and Yor, 1999) yields that  $\varphi \rightarrow I_2(\varphi)$  admits a continuous version  $Q^i$ -a.s..

As  $X_i \rightarrow L_{T_i}(X_i, \varphi)$  is measurable for all  $\varphi$  and  $\varphi \rightarrow L_{T_i}(X_i, \varphi)$  is continuous for all  $X_i$ , the joint measurability can be proved as follows. For  $m \in \mathbb{N}$  and  $k = (k_1, \dots, k_d) \in \mathbb{Z}^d$ , set  $B_{k,m} = \prod_{i=1}^d [k_i/2^m, (k_i + 1)/2^m[$ . These sets are disjoint and for all  $m$ ,  $\mathbb{R}^d = \cup_{k \in \mathbb{Z}^d} B_{k,m}$ . Let  $\varphi_{k,m} = (k_i/2^m, i = 1, \dots, d)$  and

set:

$$L_m(X_i, \varphi) = \sum_{k \in \mathbb{Z}^d} L_{T_i}(X_i, \varphi_{k,m}) 1_{B_{k,m}}(\varphi).$$

Thus,  $L_m(X_i, \varphi)$  is jointly measurable. As  $L_{T_i}(X_i, \varphi)$  is continuous w.r.t.  $\varphi$ ,  $L_m(X_i, \varphi) \rightarrow_{m \rightarrow +\infty} L_{T_i}(X_i, \varphi)$ . Hence, the result.  $\square$

*Proof of Proposition 3.* For  $H$  a positive measurable function on  $C_{T_i}$ , we have:

$$E_{Q_\theta^i}(H(X_i)) = E_{P_\theta^i}(H(X_i)) = E_{P_\theta^i}[E_{P_\theta^i}(H(X_i)|\phi_i)].$$

By Propositions 1 and 2, as  $L_{T_i}(X_i, \varphi)$  is the density of  $Q_\varphi^{x_i, T_i}$  w.r.t.  $Q^i$ , we get:

$$E_{P_\theta^i}(H(X_i)|\phi_i = \varphi) = E_{Q_\varphi^{x_i, T_i}}(H(X_i)) = \mathbb{E}_{Q^i}(H(X_i)L_{T_i}(X_i, \varphi)).$$

Using the joint measurability of  $L_{T_i}(X_i, \varphi)$  w.r.t.  $(X_i, \varphi)$ , the Fubini theorem yields:

$$\begin{aligned} \mathbb{E}_{Q_\theta^i}(H(X_i)) &= \int_{\mathbb{R}^d} g(\varphi, \theta) d\nu(\varphi) \mathbb{E}_{Q^i}(H(X_i)L_{T_i}(X_i, \varphi)) \\ &= \mathbb{E}_{Q^i} H(X_i) \int_{\mathbb{R}^d} g(\varphi, \theta) L_{T_i}(X_i, \varphi) d\nu(\varphi). \end{aligned}$$

Thus, the density of  $Q_\theta^i$  w.r.t.  $Q^i$  is computed as:

$$\frac{dQ_\theta^i}{dQ^i}(X_i) = \int_{\mathbb{R}^d} L_{T_i}(X_i, \varphi) g(\varphi, \theta) d\nu(\varphi) := \lambda_i(X_i, \theta).$$

The formula for the exact likelihood follows.  $\square$

*Proof of Proposition 4.* We need compute first the joint density of  $(\phi_i, X_i)$  w.r.t.  $d\varphi \otimes dQ^i$ :

$$\exp\left(\varphi U_i - \frac{\varphi^2}{2} V_i\right) \times \frac{1}{\omega \sqrt{2\pi}} \exp\left[-\frac{1}{2\omega^2}(\varphi - \mu)^2\right].$$

Developping the exponent yields:

$$E_i = -\frac{1}{2} [\varphi^2(V_i + \omega^{-2}) - 2\varphi(U_i + \omega^{-2}\mu)] - \frac{1}{2}\omega^{-2}\mu^2. \quad (21)$$

Let us set:

$$m_i = \frac{U_i + \omega^{-2}\mu}{V_i + \omega^{-2}} = \frac{\mu + \omega^2 U_i}{1 + \omega^2 V_i}, \quad \sigma_i^2 = (V_i + \omega^{-2})^{-1} = \frac{\omega^2}{1 + \omega^2 V_i}. \quad (22)$$

Thus, the conditional distribution of  $\phi_i$  given  $X_i$  is the Gaussian law  $\mathcal{N}(m_i, \sigma_i^2)$ .

After some elementary algebra, we get:

$$E_i = -\frac{1}{2\sigma_i^2}(\varphi - m_i)^2 - \frac{1}{2}V_i(1 + \omega^2 V_i)^{-1}(\mu - V_i^{-1}U_i)^2 + \frac{1}{2}V_i^{-1}U_i^2.$$

Thus, the result.  $\square$

*Proof of Lemma 1.* For the proof, we set  $\gamma_1(\theta) = \gamma_1$  and  $I_1(\omega^2) = I_1$  (see

(11)). Let  $l(X_1, \theta) = \log \lambda_1(X_1, \theta)$  and set  $\theta(u) = (\mu + u, \omega^2)$ . Developping  $(U_1 - (\mu + u)V_1)^2$ , we get:

$$l(X_1, \theta(u)) = l(X_1, \theta) + u\gamma_1 - \frac{u^2}{2}I_1.$$

Here,  $\frac{\partial}{\partial \mu}l(X_1, \theta) = \gamma_1$  and  $\frac{\partial^2}{\partial \mu^2}l(X_1, \theta) = -I_1$ . Taking exponential yields:

$$\lambda_1(X_1, \theta) \exp(u\gamma_1) = \lambda_1(X_1, \theta(u)) \exp\left(\frac{u^2}{2}I_1\right).$$

Integrating w.r.t. the dominating measure  $Q^1$ , we obtain, as  $I_1 \leq 1/\omega^2$ ,

$$E_\theta \exp(u\gamma_1) = E_{\theta(u)} \exp\left(\frac{u^2}{2}I_1\right) \leq \exp\left(\frac{u^2}{2\omega^2}\right) < +\infty.$$

Now, as  $u\mu \leq (u + \mu)^2/4$ ,

$$E_\theta \left( \exp\left(u \frac{U_1}{1 + \omega^2 V_1}\right) \right) \leq E_\theta \exp(u\gamma_1) \exp\left(\frac{(u + \mu)^2}{4\omega^2}\right) < +\infty.$$

□

*Proof of Proposition 5.* We set  $\gamma_1(\theta) = \gamma_1$  and  $I_1(\omega^2) = I_1$ . Let  $\theta = (\mu, \omega^2)$  and  $\tau = (0, \omega^2)$  and set

$$p_1(\theta) = \frac{\lambda_1(X_1, \theta)}{\lambda_1(X_1, \tau)} = \frac{dQ_\theta^1}{dQ_\tau^1} = \exp\left(\mu \frac{U_1}{1 + \omega^2 V_1} - \frac{\mu^2}{2} \frac{V_1}{1 + \omega^2 V_1}\right),$$

so that  $\int_{C_\tau} p_1(\theta) dQ_\tau^1 = 1$ . Provided that we can interchange derivation w.r.t.  $\mu$  and integration w.r.t.  $Q_\tau^1$ , we have for  $j \geq 1$ ,

$$\int_{C_\tau} \frac{\partial^j p_1}{\partial \mu^j}(\theta) dQ_\tau^1 = 0. \quad (23)$$

Before justifying the interchange of derivation and integration, let us compute the successive derivatives of  $p_1(\theta)$ . We have:

$$\frac{\partial p_1}{\partial \mu}(\theta) = \gamma_1 p_1(\theta), \quad \frac{\partial^2 p_1}{\partial \mu^2}(\theta) = (\gamma_1^2 - I_1) p_1(\theta), \quad \frac{\partial^3 p_1}{\partial \mu^3}(\theta) = (\gamma_1^3 - 3\gamma_1 I_1) p_1(\theta),$$

$$\frac{\partial^4 p_1}{\partial \mu^4}(\theta) = (\gamma_1^4 - 6\gamma_1^2 I_1 + 3I_1^2) p_1(\theta).$$

Therefore, (23) for  $j = 1, 2, 3, 4$  imply the announced moments relations.

It remains to justify the interchange of derivation and integration. Let us fix  $\bar{\mu}$  and  $\varepsilon > 0$ . For  $\mu \in [\bar{\mu} - \varepsilon, \bar{\mu} + \varepsilon]$ , we have the bound

$$\left| \frac{\partial p_1}{\partial \mu}(\theta) \right| \leq \left( \left| \frac{U_1}{1 + \omega^2 V_1} \right| + \frac{C}{\omega^2} \right) \left( \exp\left((\bar{\mu} - \varepsilon) \frac{U_1}{1 + \omega^2 V_1}\right) + \exp\left((\bar{\mu} + \varepsilon) \frac{U_1}{1 + \omega^2 V_1}\right) \right),$$

where  $C = |\bar{\mu} + \varepsilon| + |\bar{\mu} - \varepsilon|$ . The upper bound is integrable w.r.t.  $Q_\tau^1$  by Lemma 1 and independent of  $\mu$ . Therefore, the interchange is justified. We proceed analogously for the other derivatives. □

*Proof of Proposition 6.* We set  $\gamma_1(\theta) = \gamma_1$  and  $I_1(\omega^2) = I_1$ . Let us compute the second derivatives of the loglikelihood. Using that

$$\partial\gamma_i/\partial\mu = -I_i, \quad \partial\gamma_i/\partial\omega^2 = -\gamma_i I_i, \quad \partial I_i/\partial\omega^2 = -I_i^2(\omega^2),$$

we obtain

$$\frac{\partial^2}{\partial\mu^2}\mathcal{L}_N(\theta) = -\sum_{i=1}^N I_i(\omega^2), \quad \frac{\partial^2}{\partial\mu\partial\omega^2}\mathcal{L}_N(\theta) = -\sum_{i=1}^N \gamma_i(\theta)I_i(\omega^2), \quad (24)$$

$$\frac{\partial^2}{\partial\omega^2\partial\omega^2}\mathcal{L}_N(\theta) = -\frac{1}{2}\sum_{i=1}^N (2\gamma_i^2(\theta)I_i(\omega^2) - I_i^2(\omega^2)). \quad (25)$$

We use the simple law of large numbers, the standard central limit theorem and Proposition 5 to conclude.  $\square$

*Proof of Proposition 7.* First note that  $\lambda_1(X_1, \theta) = \lambda_1(U_1, V_1, \theta)$  depends on  $X_1$  only through  $(U_1, V_1)$ . Hence, under  $Q_\theta^1$ , by (H4), the couple  $(U_1, V_1)$  admits a density w.r.t. the Lebesgue measure equal to  $f_\theta(u, v) = \lambda_1(u, v, \theta)f(u, v)$ . Assuming that  $Q_\theta^1 = Q_{\theta_0}^1$  implies that  $f_\theta(u, v) = f_{\theta_0}(u, v)$  a.e., by the continuity of the functions  $f_\theta(u, v), f_{\theta_0}(u, v)$ , the equality holds everywhere. As  $f(u, v)$  is positive on an open ball  $B$  of  $\mathbb{R} \times (0, +\infty)$ , we deduce that, on the ball  $B$ , the following equality holds:

$$\left(\frac{1 + \omega_0^2 v}{1 + \omega^2 v}\right)^{1/2} = \exp\left[-\frac{v}{2(1 + \omega_0^2 v)}\left(\mu_0 - \frac{u}{v}\right)^2 + \frac{v}{2(1 + \omega^2 v)}\left(\mu - \frac{u}{v}\right)^2\right].$$

The left-hand side is a function of  $v$  only while the right-hand side is a function of  $(u, v)$ . This is only possible if  $\omega = \omega_0$  and  $\mu = \mu_0$ . As  $K(Q_{\theta_0}^1, Q_\theta^1) \geq 0$  and  $= 0$  if and only if  $Q_{\theta_0}^1 = Q_\theta^1$ , we get (i).

Let  $\mathcal{L}_1(\theta) = \log \lambda_1(X_1, \theta)$  (see (4)-(8)). We have

$$K(Q_{\theta_0}^1, Q_\theta^1) = E_{\theta_0}(\mathcal{L}_1(\theta_0) - \mathcal{L}_1(\theta)).$$

Rearranging terms, we get:

$$\begin{aligned} \mathcal{L}_1(\theta_0) - \mathcal{L}_1(\theta) &= \frac{1}{2} \log \left( \frac{1 + \omega^2 V_1}{1 + \omega_0^2 V_1} \right) + \frac{1}{2} \frac{(\omega_0^2 - \omega^2) U_1^2}{(1 + \omega^2 V_1)(1 + \omega_0^2 V_1)} \\ &\quad + \frac{\mu^2 V_1}{2(1 + \omega^2 V_1)} - \frac{\mu U_1}{1 + \omega^2 V_1} - \left( \frac{\mu_0^2 V_1}{2(1 + \omega_0^2 V_1)} - \frac{\mu_0 U_1}{1 + \omega_0^2 V_1} \right). \end{aligned}$$

Let us prove that this r.v. has finite expectation under  $E_{\theta_0}$ . We have the upper bound:

$$0 < \frac{1 + \omega^2 V_1}{1 + \omega_0^2 V_1} < 1 + \frac{\omega^2}{\omega_0^2}.$$

Introducing the function  $h(x) = x - 1 - \log x$ , which is defined on  $(0, +\infty)$  and

non-negative, we get the lower bound:

$$\log\left(\frac{1+\omega^2V_1}{1+\omega_0^2V_1}\right) = h\left(\frac{1+\omega_0^2V_1}{1+\omega^2V_1}\right) + (\omega^2 - \omega_0^2)\frac{V_1}{1+\omega^2V_1} \geq (\omega^2 - \omega_0^2)\frac{V_1}{1+\omega^2V_1}.$$

Thus,

$$\left|\log\left(\frac{1+\omega^2V_1}{1+\omega_0^2V_1}\right)\right| \leq \log\left(1 + \frac{\omega^2}{\omega_0^2}\right) + \frac{|\omega^2 - \omega_0^2|}{\omega^2}. \quad (26)$$

For the second term, we write:

$$0 < \frac{U_1^2}{(1+\omega^2V_1)(1+\omega_0^2V_1)} = \left(\frac{U_1}{1+\omega_0^2V_1}\right)^2 \frac{1+\omega_0^2V_1}{1+\omega^2V_1} \leq \left(\frac{U_1}{1+\omega_0^2V_1}\right)^2 \left(1 + \frac{\omega_0^2}{\omega^2}\right) \quad (27)$$

which has finite  $E_{\theta_0}$ -expectation by Lemma 1. For the last terms, we only need to check that  $E_{\theta_0}|U_1/(1+\omega^2V_1)| < +\infty$ . For this, we remark that:

$$\frac{U_1}{1+\omega^2V_1} = \frac{U_1}{1+\omega_0^2V_1} \left(1 + (\omega_0^2 - \omega^2)\frac{V_1}{1+\omega^2V_1}\right).$$

Therefore,

$$\left|\frac{U_1}{1+\omega^2V_1}\right| \leq \left|\frac{U_1}{1+\omega_0^2V_1}\right| \left(1 + \frac{|\omega_0^2 - \omega^2|}{\omega^2}\right). \quad (28)$$

By Lemma 1, the right-hand side has finite  $E_{\theta_0}$ -expectation. The function  $\theta \rightarrow \mathcal{L}_1(\theta_0) - \mathcal{L}_1(\theta)$  is continuous. For all  $\theta = (\mu, \omega^2) \in [\underline{\mu}, \bar{\mu}] \times [\underline{\omega}^2, \bar{\omega}^2] \subset \mathbb{R} \times (0, +\infty)$ , using inequalities (26)-(27)-(28), we can easily obtain an upper bound for  $|\mathcal{L}_1(\theta_0) - \mathcal{L}_1(\theta)|$  which has finite  $E_{\theta_0}$ -expectation and is uniform on the interval  $[\underline{\mu}, \bar{\mu}] \times [\underline{\omega}^2, \bar{\omega}^2]$ . The continuity of the Kullback information follows. This gives (ii).  $\square$

**Remark 3.** *It is worth noting that, although we have an explicit expression of  $K(Q_{\theta_0}^1, Q_{\theta}^1)$ , we cannot prove directly, using this expression, that  $K(Q_{\theta_0}^1, Q_{\theta}^1) = 0$  implies  $\theta = \theta_0$ .*

*Proof of Proposition 8.* We first prove 1. As  $(1/N)(\mathcal{L}_N(\theta_0) - \mathcal{L}_N(\theta))$  converges to  $K(Q_{\theta_0}^1, Q_{\theta}^1)$  in  $Q_{\theta_0}$ -probability, the loglikelihood  $-(1/N)\mathcal{L}_N(\theta)$  is a contrast process with contrast function  $\theta \rightarrow K(Q_{\theta_0}^1, Q_{\theta}^1)$ . Following the usual standard proof of consistency of minimum contrasts estimators (see *e.g.* van der Vaart, 1998), it remains to study the continuity modulus of  $-(1/N)\mathcal{L}_N(\theta)$  defined by:

$$w_N(\eta) = \sup_{\|\theta - \theta'\| \leq \eta, \theta, \theta' \in \Theta} |\mathcal{L}_N(\theta) - \mathcal{L}_N(\theta')|/N.$$

We simply use  $w_N(\eta) \leq \eta \sup_{\theta \in \Theta} \|\nabla \mathcal{L}_N(\theta)/N\|$  and bound the score function (12). By (H5), we have  $\Theta \subset [\underline{\mu}, \bar{\mu}] \times [\underline{\omega}^2, \bar{\omega}^2]$  with  $\underline{\mu} < \bar{\mu}, 0 < \underline{\omega}^2 < \bar{\omega}^2$ . We have:



$$\gamma_i(\theta) = \frac{U_i}{1 + \omega_0^2 V_i} \left( 1 + \frac{(\omega_0^2 - \omega^2) V_i}{1 + \omega^2 V_i} \right) - \mu \frac{V_i}{1 + \omega^2 V_i}.$$

Thus,

$$\sup_{\theta \in \Theta} |\gamma_i(\theta)| \leq \left| \frac{U_i}{1 + \omega_0^2 V_i} \right| \left( 2 + \frac{\omega_0^2}{\underline{\omega}^2} \right) + \frac{|\bar{\mu}|}{\underline{\omega}^2}. \quad (29)$$

Therefore, there is a constant  $C$  such that

$$E_{\theta_0} w_N(\eta) \leq C \eta E_{\theta_0} \left( \left| \frac{U_1}{1 + \omega_0^2 V_1} \right| + \left( \frac{U_1}{1 + \omega_0^2 V_1} \right)^2 \right).$$

This leads to the consistency of  $\hat{\theta}_N$ .

For the second point, the proof follows the standard scheme. By the consistency and (H5),  $Q_{\theta_0}(\hat{\theta}_N \in \mathring{\Theta}) \rightarrow 1$ . For the proof, let  $\hat{\theta}_{N,i}, \theta_{0,i}$  be the components of  $\hat{\theta}_N, \theta_0$ . Set  $U_N(\theta) = -(1/N)\mathcal{L}_N(\theta)$  and denote by  $U'_{N,i}, U''_{N,ij}$  the derivatives of  $U_N$  w.r.t.  $\theta_i$  or  $\theta_i \theta_j$ . The Taylor formula writes:

$$0 = U'_{N,i}(\hat{\theta}_N) = U'_{N,i}(\theta_0) + \sum_{j=1,2} (\hat{\theta}_{N,j} - \theta_{0,j}) (U''_{N,ij}(\theta_0) + R_N),$$

with

$$R_N = \int_0^1 \left( U''_{N,ij}(\theta_0 + s(\hat{\theta}_N - \theta_0)) - U''_{N,ij}(\theta_0) \right) ds.$$

Using Propositions 8 and 6, it remains to check that  $R_N$  tends in  $Q_{\theta_0}$ -probability to 0. For this, we compute the third derivatives of  $U_N$  using (24)-(25):

$$\begin{aligned} \frac{1}{N} \frac{\partial^3}{\partial \mu^3} \mathcal{L}_N(\theta) &= 0, & \frac{1}{N} \frac{\partial^3}{\partial \omega^2 \partial \omega^2 \partial \omega^2} \mathcal{L}_N(\theta) &= \frac{1}{N} \sum_{i=1}^N [3\gamma_i^2(\theta) I_i^2(\omega^2) - I_i^3(\omega^2)], \\ \frac{1}{N} \frac{\partial^3}{\partial^2 \mu \partial \omega^2} \mathcal{L}_N(\theta) &= \frac{1}{N} \sum_{i=1}^N I_i^2(\omega^2), & \frac{1}{N} \frac{\partial^3}{\partial \mu \partial \omega^2 \partial \omega^2} \mathcal{L}_N(\theta) &= \frac{2}{N} \sum_{i=1}^N \gamma_i(\theta) I_i^2(\omega^2). \end{aligned}$$

Using (29), we obtain, for  $C$  a constant depending on  $\Theta$

$$|R_N| \leq C |\hat{\theta}_N - \theta_0| \frac{1}{N} \sum_{i=1}^N \left( 1 + \left( \left| \frac{U_i}{1 + \omega_0^2 V_i} \right| + \left( \frac{U_i}{1 + \omega_0^2 V_i} \right)^2 \right) \right).$$

As we have proved the consistency,  $R_N$  tends to 0. Hence the result.  $\square$

*Proof of Lemma 2.* The matrix  $V_i + \Omega^{-1}$  is symmetric and satisfies, for all  $x \in \mathbb{R}^d$ ,  $x \neq 0$ ,  $x'(V_i + \Omega^{-1})x = x'V_i x + x'\Omega^{-1}x > 0$  as  $V_i$  and  $\Omega^{-1}$  are positive definite. Thus  $V_i + \Omega^{-1}$  is positive definite hence invertible. Noting that  $I_d + V_i \Omega = (\Omega^{-1} + V_i)\Omega$  and  $I_d + \Omega V_i = \Omega(\Omega^{-1} + V_i)$  yields that both matrices are invertible.  $\square$

*Proof of Proposition 9.* We compute the joint density of  $(\phi_i, X_i)$  w.r.t.  $d\varphi \otimes dQ^i$ :

$$\exp\left(\varphi'U_i - \frac{1}{2}\varphi'V_i\varphi\right) \times \exp\left[-\frac{1}{2}(\varphi - \mu)'\Omega^{-1}(\varphi - \mu)\right].$$

Let

$$E_i = -\frac{1}{2}[\varphi'(V_i + \Omega^{-1})\varphi - 2\varphi'(U_i + \Omega^{-1}\mu)] - \frac{1}{2}\mu'\Omega^{-1}\mu, \quad (30)$$

and set:

$$m_i = \Sigma_i(U_i + \Omega^{-1}\mu), \quad \Sigma_i^2 = (V_i + \Omega^{-1})^{-1}. \quad (31)$$

Thus, the conditional distribution of  $\phi_i$  given  $X_i$  is the Gaussian law  $\mathcal{N}(m_i, \Sigma_i^2)$ .

Hence, we have

$$E_i = -\frac{1}{2}(\varphi - m_i)'\Sigma_i^{-1}(\varphi - m_i) - \frac{1}{2}(\mu - V_i^{-1}U_i)'R_i^{-1}(\mu - V_i^{-1}U_i) + \frac{1}{2}U_i'V_i^{-1}U_i.$$

Thus, the result.  $\square$

*Proof of Proposition 10.* For the first point, we proceed as in Lemma 1. Let  $\gamma_1(\theta) = \gamma_1$ ,  $I_1(\Omega) = I_1$ ,  $l(X_1, \theta) = \log \lambda_1(X_1, \theta)$  and  $\theta(u) = (\mu + u, \Omega)$ . Developing  $((\mu + u) - V_1^{-1}U_1)'R_1^{-1}((\mu + u) - V_1^{-1}U_1)$  yields:

$$l(X_1, \theta(u)) = l(X_1, \theta) + u'\gamma_1 - \frac{1}{2}u'I_1u.$$

Thus

$$\lambda_1(X_1, \theta) \exp(u'\gamma_1) = \lambda_1(X_1, \theta(u)) \exp\left(\frac{1}{2}u'I_1u\right).$$

Remark that  $u'I_1u \leq u'\Omega^{-1}u$  as  $I_1 = \Omega^{-1} - (\Omega V_1 \Omega + \Omega)^{-1}$ . Integrating w.r.t.  $Q^1$ , we obtain,

$$E_\theta \exp(u'\gamma_1) = E_{\theta(u)} \exp\left(\frac{1}{2}u'I_1u\right) \leq \exp\left(\frac{1}{2}u'\Omega^{-1}u\right) < +\infty.$$

Now, we can write, as  $u'I_1\mu = \frac{1}{4}((u + \mu)'I_1(u + \mu) - (u - \mu)'I_1(u - \mu)) \leq \frac{1}{4}(u + \mu)'I_1(u + \mu)$ :

$$E_\theta(\exp(u'(I_d + \Omega V_i)^{-1}U_1)) \leq E_\theta \exp(u'\gamma_1) \exp\left(\frac{1}{4}(u + \mu)'\Omega^{-1}(u + \mu)\right) < +\infty.$$

This gives 1. The proof of the second point and third points is analogous to Proposition 5 and Proposition 6.  $\square$

*Proof of Lemma 3.* We only treat the case  $p \geq 2$ . Assumptions imply that  $b^2/\sigma^2$  is Lipschitz say with constant  $L$ . Using the Hölder inequality twice, we get

$$\mathbb{E}_{\theta_0} |V_i - V_i^{(n)}|^p \leq L^p T^{p-1} \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} \mathbb{E}_{\theta_0} |X_i(s) - X_i(t_k)|^p ds.$$

Now, for  $0 \leq t \leq t+h \leq T$ , we have

$$X_i(t+h) - X_i(t) = \int_t^{t+h} \phi_i b(X_i(s)) ds + \int_t^{t+h} \sigma(X_i(s)) dW_i(s).$$

Using  $\phi_i b(X_i(s)) \leq \phi_i^2/2 + b^2(X_i(s))/2$  and the Hölder inequality, we get

$$|X_i(t+h) - X_i(t)|^p \leq C \left( \phi_i^{2p} h^p + \int_t^{t+h} b^{2p}(X_i(s)) ds h^{p-1} + \left| \int_t^{t+h} \sigma(X_i(s)) dW_i(s) \right|^p \right).$$

We use that  $b$  and  $\sigma$  have linear growth, the Burkholder-Davis-Gundy inequality and  $M_k = \sup_{s \in [0, T]} \mathbb{E}_{\theta_0} |X_i(s)|^k < \infty$  for all  $k$ , then

$$\mathbb{E}_{\theta_0} |X_i(t+h) - X_i(t)|^p \leq C \left( h^p \mathbb{E}_{\theta_0} \phi_i^{2p} + h^p (1 + M_{2p}) + h^{p/2} (1 + M_p) \right) \leq Ch^{p/2}.$$

Thus we obtain

$$\mathbb{E}_{\theta_0} |V_i - V_i^{(n)}|^p \leq \frac{C}{n^{p/2}}.$$

For the difference  $U_i - U_i^n$ , we have

$$U_i - U_i^n = \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} \left( \frac{b}{\sigma^2}(X_i(s)) - \frac{b}{\sigma^2}(X_i(t_k)) \right) dX_i(s) = A_1 + A_2$$

where  $A_1$  is a term analogous to the one already studied above and

$$A_2 = \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} \left( \frac{b}{\sigma^2}(X_i(s)) - \frac{b}{\sigma^2}(X_i(t_k)) \right) \sigma(X_i(s)) dW_i(s).$$

We introduce the process

$$H_s^{(n)} = \sum_{k=0}^{n-1} 1_{]t_k, t_{k+1}[}(s) \left( \frac{b}{\sigma^2}(X_i(s)) - \frac{b}{\sigma^2}(X_i(t_k)) \right) \sigma(X_i(s))$$

so that  $A_2 = \int_0^T H_s^{(n)} dW_i(s)$ . We treat  $\mathbb{E}_{\theta_0} |A_2|^p$  using the Burkholder-Davis-Gundy inequality and similar tools as above.  $\square$

## Tables

True value	$N = 20, T = 5$		$N = 50, T = 5$		$N = 50, T = 10$	
	Mean (Var)	$\frac{1}{N}\mathcal{I}_N^{-1}$	Mean (Var)	$\frac{1}{N}\mathcal{I}_N^{-1}$	Mean (Var)	$\frac{1}{N}\mathcal{I}_N^{-1}$
$\mu = -1$	-0.97 (0.06)	0.06	-0.99 (0.02)	0.02	-0.99 (0.03)	0.02
$\omega^2 = 1$	0.91 (0.16)	0.14	0.99 (0.06)	0.06	0.97 (0.06)	0.05
$\mu = 5$	5.01 (0.06)	0.06	5.00 (0.03)	0.02	5.01 (0.02)	0.02
$\omega^2 = 1$	0.99 (0.16)	0.14	0.97 (0.05)	0.06	0.97 (0.05)	0.05

Table 1: Example 1: Mixed Brownian motion with drift. Empirical mean and variance (in brackets) of  $\hat{\mu}_N$  and  $\hat{\omega}_N^2$  computed from 100 datasets for three designs and two sets of parameters  $(\mu, \omega^2)$ . The exact asymptotic variance  $\text{diag}(N^{-1}\mathcal{I}_N^{-1}(\theta_0))$  is also computed.

True value		$N = 20, T = 5$		$N = 50, T = 5$		$N = 50, T = 10$	
		Mean (Var)	$\frac{1}{N}\mathcal{I}_N^{-1}$	Mean (Var)	$\frac{1}{N}\mathcal{I}_N^{-1}$	Mean (Var)	$\frac{1}{N}\mathcal{I}_N^{-1}$
$\mu$ estimated, $\omega^2$ fixed							
$\mu = -5, \omega^2 = 1$	$\hat{\mu}_N$	-5.03 (0.14)	0.11	-5.01 (0.06)	0.05	-4.99 (0.04)	0.02
$\mu = 10, \omega^2 = 1$	$\hat{\mu}_N$	9.90 (0.05)	0.05	9.91 (0.01)	0.02	9.90 (0.02)	0.02
$\mu$ fixed, $\omega^2$ estimated							
$\mu = -5, \omega^2 = 1$	$\hat{\omega}_N^2$	0.96 (0.31)	0.62	0.96 (0.14)	1.49	0.93 (0.16)	0.27
$\mu = 10, \omega^2 = 1$	$\hat{\omega}_N^2$	0.99 (0.08)	0.20	1.00 (0.04)	0.04	0.97(0.03)	0.05
$\mu$ and $\omega^2$ estimated							
$\mu = -5, \omega^2 = 1$	$\hat{\mu}_N$	-4.95 (0.22)	-	-4.99 (0.07)	-	-4.96 (0.03)	-
	$\hat{\omega}_N^2$	0.74 (1.00)	-	0.91 (0.34)	-	0.99 (0.21)	-
$\mu = 10, \omega^2 = 1$	$\hat{\mu}_N$	9.85 (0.05)	-	9.85 (0.02)	-	9.84 (0.01)	-
	$\hat{\omega}_N^2$	0.94 (0.08)	-	0.95 (0.04)	-	0.94 (0.05)	-

Table 2: Example 2: Mixed Ornstein-Uhlenbeck model with one random effect. Estimation of  $\mu$  when  $\omega_0^2$  is known, of  $\omega^2$  when  $\mu_0$  is known and simultaneous estimation of  $\mu$  and  $\omega^2$ , for different values of  $(N, T)$  and different parameter values. Empirical mean, variance (in brackets) and estimated value of the asymptotic variance  $\text{diag}(N^{-1}\mathcal{I}_N^{-1}(\theta_0))$  are computed from 100 repeated simulated datasets.

True parameter values	$N = 20, T = 5$ Mean (Var)	$N = 50, T = 5$ Mean (Var)	$N = 50, T = 10$ Mean (Var)
$\mu_1 = 0.1$	0.095 (0.082)	0.102 (0.059)	0.102 (0.028)
$\mu_2 = 1$	1.007 (0.275)	1.020 (0.193)	0.984 (0.175)
$\omega_1^2 = 0.01$	0.010 (0.014)	0.010 (0.009)	0.010 (0.005)
$\omega_2^2 = 1$	0.919 (0.526)	0.956 (0.342)	1.029 (0.278)
$\mu_1 = 0.1$	0.123 (0.073)	0.104 (0.047)	0.106 (0.019)
$\mu_2 = 1$	1.085 (0.287)	1.010 (0.166)	1.022 (0.171)
$\omega_1^2 = 0.001$	0.002 (0.004)	0.002 (0.004)	0.001 (0.002)
$\omega_2^2 = 1$	1.095 (0.488)	1.005 (0.353)	1.024 (0.253)

Table 3: Example 3: Mixed Ornstein-Uhlenbeck model with two random effects. Empirical mean, variance (in brackets) are computed from 100 repeated simulated datasets for different values of  $(N, T)$  and different parameter values.

True value	Est	$N = 20, T = 5$ Mean (Var)	$N = 50, T = 5$ Mean (Var)	$N = 50, T = 10$ Mean (Var)
$\mu = -1$	$\hat{\mu}_N$	-1.03 (0.12)	-0.98 (0.04)	-1.02 (0.03)
$\omega^2 = 1$	$\hat{\omega}_N^2$	0.95 (0.42)	0.94 (0.11)	0.98 (0.09)
$\mu = 5$	$\hat{\mu}_N$	4.99 (0.05)	5.00 (0.02)	5.04 (0.03)
$\omega^2 = 1$	$\hat{\omega}_N^2$	0.96 (0.17)	0.97 (0.04)	1.00 (0.04)

Table 4: Example 4. Empirical mean and variance (in brackets) of the estimates  $\hat{\mu}_N$  and  $\hat{\omega}_N^2$  computed from 100 datasets for different values of  $(N, T)$  and different parameter values.