Metrika (2018) 81:953–983
https://doi.org/10.1007/s00184-018-0666-z

Approximate maximum likelihood estimation for stochastic differential equations with random effects in the drift and the diffusion

Maud Delattre[1] · Valentine Genon-Catalot[2] · Catherine Larédo[3,4]

Received: 5 October 2017 / Published online: 4 July 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract
Consider N independent stochastic processes $(X_i(t), t \in [0, T])$, $i = 1, \ldots, N$, defined by a stochastic differential equation with random effects where the drift term depends linearly on a random vector Φ_i and the diffusion coefficient depends on another linear random effect Ψ_i. For these effects, we consider a joint parametric distribution. We propose and study two approximate likelihoods for estimating the parameters of this joint distribution based on discrete observations of the processes on a fixed time interval. Consistent and $\sqrt{N}$-asymptotically Gaussian estimators are obtained when both the number of individuals and the number of observations per individual tend to infinity. The estimation methods are investigated on simulated data and show good performances.

Keywords Asymptotic properties · Discrete observations · Estimating equations · Parametric inference · Random effects models · Stochastic differential equations

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00184-018-0666-z) contains supplementary material, which is available to authorized users.

✉ Maud Delattre
maud.delattre@agroparistech.fr

Valentine Genon-Catalot
valentine.genon-catalot@parisdescartes.fr

Catherine Larédo
catherine.laredo@inra.fr

1 UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005 Paris, France

2 UMR CNRS 8145, Laboratoire MAP5, Université Paris Descartes, Sorbonne Paris Cité, 45, rue des Saints-Pères, 75006 Paris, France

3 INRA, MaIAGE, Domaine de Vilvert, 78352 Jouy-en-Josas, France

4 LPMA, Paris Diderot, Sorbonne Paris Cité, Avenue de France, 75013 Paris, France

1 Introduction

Longitudinal data are widely collected in clinical trials, epidemiology, pharmacokinetic pharmacodynamics experiments and agriculture. The interest may focus on population effects among individuals and individual specific behaviour. In mixed effects models, random effects are incorporated to accomodate variability among subjects or inter-individual variability, while the same structural model rules the dynamics of each subject. In stochastic differential equations with mixed effects (SDEMEs), the structural model is a set of stochastic differential equations. The use of SDEMEs is comparatively recent. It has first been motivated by pharmacological applications (see Ditlevsen and De Gaetano 2005; Donnet and Samson 2008; Delattre and Lavielle 2013; Leander et al. 2015; Forman and Picchini 2016 and many others) but also neurobiological ones (as in Dion 2016 for example). The main issue in mixed-effects models is the estimation of the parameters in the distribution of the random effects. This is generally difficult in practice due to the intractable likelihood function. To overcome this problem, many methods based on approximate solutions associated with computationally intensive numerical methods have been proposed (Picchini et al. 2010; Picchini and Ditlevsen 2011; Delattre and Lavielle 2013). For general mixed models as well as for SDEMEs, these methods, due to their iterative settings, lead to large computation times for the estimation. Moreover, when the stuctural model is a stochastic differential equation, there is an additional problem which derives from the intractability of the likelihood associated with discrete observations of one path.

The SDEME framework allows to take account of two sources of randomness in the structural SDE model, random effects in the drift and random effects in the diffusion coefficient, and the temptation to incorporate in this SDE model a joint distribution for these two random effects is quite natural. This modeling has been proposed by several authors: Picchini et al. (2010), Picchini and Ditlevsen (2011), Berglund et al. (2001), Forman and Picchini (2016), Whitaker et al. (2017). But to our knowledge, it has never been studied theoretically. However, it is well known that, for discretely observed diffusion processes, estimation of parameters in the drift coefficient and diffusion coefficient have different properties (see e.g., Kessler et al. 2012). Our aim here is to investigate the statistical properties of such a situation according to n, the number of observations per individual (or path), and the total number of individuals N. The results of Nie and Yang (2005), Nie (2006, 2007) do not provide answers to these questions, and all the numerical methods compulsory to the likelihood intractability, are unable to tackle this problem. Understanding how estimation performs in these kinds of models is an important issue, necessary to clarify what can be expected from all the numerical methods widely used in this domain. Since the work of Ditlevsen and De Gaetano (2005), where the special case of a mixed-effects Brownian motion with drift is treated, the main contributions to our knowledge for a theoretical study of parametric inference in SDEMEs, are from Delattre et al. (2013, 2015, 2017) and Grosse Ruse et al. (2017).

In this paper, we consider discrete observations and simultaneous random effects in the drift and diffusion coefficients by means of a joint parametric distribution. The inclusion of random effects both in the drift and the diffusion coefficient raises new problems which were not addressed in our previous works. We focus on specific dis-

tributions for the random effects. They derive from a Bayesian choice of distributions and lead to explicit approximate likelihoods (this choice indeed corresponds to explicit posterior and marginal distributions for an n-sample of Gaussian distributions with a specific prior distribution on the parameters).

More precisely, we consider N real valued stochastic processes $(X_i(t), t \geq 0)$, $i = 1, \ldots, N$, with dynamics ruled by the following random effects stochastic differential equation (SDE):

$$d\,X_i(t) = \Phi_i' b(X_i(t))dt + \Psi_i \sigma(X_i(t))\,dW_i(t), \quad X_i(0) = x,\ i = 1, \ldots, N, \quad (1)$$

where $(W_1, \ldots, W_N)$ are N independent Wiener processes, $((\Phi_i, \Psi_i), i = 1, \ldots N)$ are N *i.i.d.* $\mathbb{R}^d \times (0, +\infty)$-valued random variables, $((\Phi_i, \Psi_i), i = 1, \ldots, N)$ and $(W_1, \ldots, W_N)$ are independent and x is a known real value. The functions $\sigma(.) : \mathbb{R} \to \mathbb{R}$ and $b(.) = (b_1(.), \ldots, b_d(.))' : \mathbb{R} \to \mathbb{R}^d$ are known. The notation X' for a vector or a matrix X denotes the transpose of X. Each process $(X_i(t))$ represents an individual and the $d+1$-dimensional random vector (Φ_i, Ψ_i) represents the (multivariate) random effect of individual i. In Delattre et al. (2013), the N processes are assumed to be continuously observed throughout a fixed time interval $[0, T]$, $T > 0$ and $\Psi_i = \gamma^{-1/2}$ is non random and known. When Φ_i follows a Gaussian distribution $\mathcal{N}_d(\boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Omega})$, the exact likelihood associated with $(X_i(t), t \in [0, T], i = 1, \ldots, N)$ is explicitely computed and asymptotic properties of the exact maximum likelihood estimators are derived under the asymptotic framework $N \to +\infty$. In Delattre et al. (2015), the case of $b(.) = 0$ and $\Psi_i = \Gamma_i^{-1/2}$ with Γ_i following a Gamma distribution $G(a, \lambda)$ is investigated; and Delattre et al. (2017) is concerned with estimation of mixed effects either in the drift or in the diffusion coefficient from discrete observations. From now on, each process $(X_i(t))$ is discretely observed on a fixed-length time interval $[0, T]$ at n times $t_j = jT/n$ and the random effects (Φ_i, Ψ_i) follow a joint parametric distribution. Our aim is to estimate the unknown parameters from the observations $(X_i(t_j), j = 1, \ldots, n; i = 1, \ldots, N)$. We focus on distributions that give rise to explicit approximations of the likelihood functions so that the construction of estimators is easy and their asymptotic study feasible. To this end, we assume that (Φ_i, Ψ_i) has the following distribution:

$$\Psi_i = \frac{1}{\Gamma_i^{1/2}}, \quad \Gamma_i \sim G(a, \lambda), \text{ and given } \Gamma_i = \gamma, \quad \Phi_i \sim \mathcal{N}_d(\boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Omega}). \quad (2)$$

Contrary to the common Gaussian assumption for random effects, the marginal distribution of Φ_i is not Gaussian: $\Phi_i - \boldsymbol{\mu} = \Gamma_i^{-1/2}\eta_i$, with $\eta_i \sim \mathcal{N}_d(0, \boldsymbol{\Omega})$, is a Student distribution. We propose two distinct approximate likelihood functions which yield asymptotically equivalent estimators as both N and the number n of observations per individual tend to infinity. The first approximate likelihood (Method 1) is natural but proving its existence raises technical difficulties (Proposition 1, Lemma 1). The second one (Method 2) derives from Method 1 and has the advantage of splitting the estimation of parameters in the diffusion coefficient and parameters in the drift term. We obtain consistent and $\sqrt{N}$-asymptotically Gaussian estimators for all parameters

under the condition $N/n \to 0$. For the parameters (λ, a) of random effects in the diffusion coefficient, we obtain that the weaker constraint N/n^2 is enough (Theorems 1, 2). We prove that the estimators obtained with these two approximate likelihoods are asymptotically equivalent (Theorem 3). We compare these results with the estimation for N *i.i.d.* direct observations of the random effects (Φ_i, Ψ_i).

The methods of the present paper and of the previous ones (Delattre et al. 2013, 2015, 2017) are now avalaible in the R package MsdeParEst (Delattre and Dion 2017) for mixed Ornstein–Uhlenbeck and CIR.

The structure of the paper is the following. Some working assumptions and two approximations of the model likelihood are introduced in Sect. 2. Two estimation methods are derived from these approximate likelihoods in Sect. 3, and their respective asymptotic properties are studied. Section 4 provides numerical simulation results for several examples of SDEMEs and illustrates the performances of the proposed methods in practice. Theoretical proofs are gathered in the "Appendix". Some technical proofs are given in an electronic supplementary material. Auxiliary results are given in Sect. 7.

2 Approximate likelihoods

2.1 Framework and assumptions

Let $(X_i(t), t \geq 0)$, $i = 1, \ldots, N$ be N real valued stochastic processes ruled by (1). The processes $(W_1, \ldots, W_N)$ and the r.v.'s (Φ_i, Ψ_i), $i = 1, \ldots, N$ are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and we set $(\mathcal{F}_t = \sigma(\Phi_i, \Psi_i, W_i(s), s \leq t, i = 1, \ldots, N), t \geq 0)$. We assume that:

(H1) The real valued functions $x \to b_j(x)$, $j = 1, \ldots, d$ and $x \to \sigma(x)$ are C^2 on $\mathbb{R}$ with first and second derivatives bounded by L. The function $\sigma^2(.)$ is lower bounded : $\exists\, \sigma_0 \neq 0, \forall x \in \mathbb{R}, \sigma^2(x) \geq \sigma_0^2$.

(H2) There exists a constant K such that, $\forall x \in \mathbb{R}$, $\|b(x)\| + |\sigma(x)| \leq K$.

Assumption (H1) corresponds to the usual linear growth condition and regularity assumptions on $b(.)$, $\sigma(.)$ which ensure the existence and uniqueness of strong solutions of Eq. (1). However, we need here, for the accurate study of discretizations (see Sect. 7.2), that $X_i(t)$ has finite moments. For stochastic differential equations, the property $\{E(X_i^{2p}(t)) < \infty\}$ holds if the initial condition satisfies $\{E(X_i^{2p}(0)) < \infty\}$. This does no longer hold here, even if $\{E(||\Phi_i||^{2p} + \Psi_i^{2p}) < \infty\}$ (see Sect. 7.1). We have either to assume that (Φ_i, Ψ_i) belongs to a bounded set or (H2). To circumvent this last assumption, we could think of applying a localization device (see Jacod and Protter 2012, Chapter 4.4.1). However, while it is straightforward to apply this method for one path, the extension to N paths $X_i(.)$ simultaneously, is complex, especially as $X_i(t)$ has possibly no finite moments (see Sect. 7.1).

The processes $(X_i(t))$ are discretely observed with sampling interval Δ_n on a fixed time interval $[0, T]$ and for sake of clarity we assume a regular sampling on $[0, T]$:

$$\Delta_n = \Delta = \frac{T}{n}, \ X_{i,n} = X_i = (X_i(t_{j,n}), \ \ t_{j,n} = t_j = jT/n, \ \ j = 1, \ldots, n). \quad (3)$$

Let ϑ denote the unknown parameter and Θ the parameter set with

$$\vartheta = (\lambda, a, \boldsymbol{\mu}, \boldsymbol{\Omega}) \quad (4)$$

The canonical space associated with one trajectory on $[0, T]$ is defined by $((\mathbb{R}^d \times (0, +\infty) \times C_T), P_\vartheta)$ where C_T denotes the space of real valued continuous functions on $[0, T]$, P_ϑ denotes the distribution of $(\Phi_i, \Psi_i, (X_i(t), t \in [0, T]))$ and ϑ the unknown parameter. For the N trajectories, the canonical space is $\prod_{i=1}^N((\mathbb{R}^d \times (0, +\infty) \times C_T), \mathbb{P}_\vartheta = \otimes_{i=1}^N P_\vartheta)$. Below, the true value of the parameter is denoted ϑ_0.

We introduce the statistics and the assumptions

$$S_{i,n} = \frac{1}{\Delta} \sum_{j=1}^n \frac{\left(X_i(t_j) - X_i(t_{j-1})\right)^2}{\sigma^2(X_i(t_{j-1}))},$$

$$V_{i,n} = \left(\sum_{j=1}^n \Delta \frac{b_k(X_i(t_{j-1}))b_\ell(X_i(t_{j-1}))}{\sigma^2(X_i(t_{j-1}))} \right)_{1 \le k, \ell \le d}, \quad (5)$$

$$U_{i,n} = \left(\sum_{j=1}^n \frac{b_k(X_i(t_{j-1}))(X_i(t_j) - X_i(t_{j-1}))}{\sigma^2(X_i(t_{j-1}))} \right)_{1 \le k \le d}. \quad (6)$$

(H3) The matrix $V_i(T)$ is positive definite a.s., where

$$V_i(T) = \left(\int_0^T \frac{b_k(X_i(s))b_\ell(X_i(s))}{\sigma^2(X_i(s))} ds \right)_{1 \le k, \ell \le d}. \quad (7)$$

(H4) The parameter set Θ satisfies that, for constants $\ell_0, \ell_1, \alpha_0, \alpha_1, m, c_0, c_1$, $0 < \ell_0 \le \lambda \le \ell_1$, $0 < \alpha_0 \le a \le \alpha_1$, $\|\boldsymbol{\mu}\| \le m$, $c_0 \le \lambda_{max}(\boldsymbol{\Omega}) \le c_1$, where $\lambda_{max}(\boldsymbol{\Omega})$ denotes the maximal eigenvalue of $\boldsymbol{\Omega}$.

Assumption (H3) ensures that all the components of Φ_i can be estimated. If the functions (b_k/σ^2) are not linearly independent, the dimension of Φ_i is not well defined and (H3) is not fulfilled. Note that, as n tends to infinity, the matrix $V_{i,n}$ defined in (5) converges a.s. to $V_i(T)$ so that, under (H3), for n large enough, $V_{i,n}$ is positive definite. Assumption (H4) is classically used in a parametric setting. Under (H4), the matrix $\boldsymbol{\Omega}$ may be non invertible which allows including non random effects in the drift term.

2.2 First approximation of the likelihood

Let us now compute our first approximate likelihood $\mathcal{L}_n(X_{i,n}, \vartheta)$ of $X_{i,n}$, with ϑ defined in (4). The exact likelihood of the i-th vector of observations is obtained by

computing first the conditional likelihood given $\Phi_i = \varphi$, $\Psi_i = \psi$, and then integrating the result with respect to the joint distribution of (Φ_i, Ψ_i). The conditional likelihood given fixed values (φ, ψ), *i.e.* the likelihood of $(X_i^{\varphi,\psi}(t_j), j = 1, \ldots, n)$ being not explicit, is approximated by using the Euler scheme likelihood of

$$d\, X_i^{\varphi,\psi}(t) = \varphi' b(X_i^{\varphi,\psi}(t))dt + \psi\sigma(X_i^{\varphi,\psi}(t))\, dW(t), \quad X_i^{\varphi,\psi}(0) = x. \tag{8}$$

Setting $\psi = \gamma^{-1/2}$, the exact likelihood of $(X_i^{\varphi,\psi}(t_j), j = 1, \ldots, n)$ is replaced by the likelihood of $(Y_{i,j}, j = 1, \ldots, n)$:

$$Y_{i,j} - Y_{i,j-1} = \Delta\varphi' b(Y_{i,j-1}) + \sqrt{\Delta}\ \psi\ \sigma(Y_{i,j-1})\ \epsilon_{i,j},$$

with $Y_{i,0} = x$ and $\epsilon_{i,j} = \frac{W_i(t_j) - W_i(t_{j-1})}{\sqrt{\Delta}}$ i.i.d. $\mathcal{N}(0, 1)$. Therefore, this yields the approximate conditional likelihood:

$$L_n(X_{i,n}, \gamma, \varphi) = L_n(X_i, \gamma, \varphi) = \gamma^{n/2} \exp\left[-\frac{\gamma}{2}(S_{i,n} + \varphi' V_{i,n}\varphi - 2\varphi' U_{i,n})\right], \tag{9}$$

where this formula ignores multiplicative functions which do not contain the unknown parameters.

The unconditional approximate likelihood is obtained integrating with respect to the joint distribution $\nu_\vartheta(d\gamma, d\varphi)$ of the random effects $(\Gamma_i = \Psi_i^{-2}, \Phi_i)$.

For this, we first integrate $L_n(X_i, \gamma, \varphi)$ with respect to the Gaussian distribution $\mathcal{N}_d(\boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Omega})$. Then, we integrate the result w.r.t. the distribution of Γ_i. At this point, a difficulty arises. This second integration is only possible on the subset $E_{i,n}(\vartheta)$ defined in (12).

Let I_d denote the identity matrix of $\mathbb{R}^d$ and set, for $i = 1, \ldots, N$, under (H3)

$$R_{i,n} = V_{i,n}^{-1} + \boldsymbol{\Omega} = V_{i,n}^{-1}(I_d + V_{i,n}\boldsymbol{\Omega}) = (I_d + \boldsymbol{\Omega} V_{i,n})V_{i,n}^{-1}, \tag{10}$$

$$T_{i,n}(\boldsymbol{\mu}, \boldsymbol{\Omega}) = (\boldsymbol{\mu} - V_{i,n}^{-1}U_{i,n})' R_{i,n}^{-1}(\boldsymbol{\mu} - V_{i,n}^{-1}U_{i,n}) - U_{i,n}' V_{i,n}^{-1} U_{i,n}, \tag{11}$$

$$E_{i,n}(\vartheta) = \{S_{i,n} + T_{i,n}(\boldsymbol{\mu}, \boldsymbol{\Omega}) > 0\}; \quad \mathbf{E}_{N,n}(\vartheta) = \cap_{i=1}^N E_{i,n}(\vartheta). \tag{12}$$

Proposition 1 *Assume that, for $i = 1, \ldots, N$, (Φ_i, Ψ_i) has distribution (2). Under (H1) and (H3), an explicit approximate likelihood for the observation $(X_{i,n}, i = 1, \ldots, N)$ is on the set $\boldsymbol{E}_{N,n}(\vartheta)$ (see 12),*

$$\mathcal{L}_{N,n}(\vartheta) = \prod_{i=1}^N \mathcal{L}_n(X_{i,n}, \vartheta), \quad \textit{where} \tag{13}$$

$$\mathcal{L}_n(X_{i,n}, \vartheta) = \frac{\lambda^a \Gamma(a + (n/2))}{\Gamma(a)\left(\lambda + \frac{S_{i,n}}{2} + \frac{T_{i,n}(\boldsymbol{\mu},\boldsymbol{\Omega})}{2}\right)^{a+(n/2)}} \frac{1}{(det(I_d + V_{i,n}\boldsymbol{\Omega}))^{1/2}}. \tag{14}$$

Note that this approximate likelihood also ignores multiplicative functions which do not contain the unknown parameters. We must now deal with the set $\mathbf{E}_{N,n}(\vartheta)$. For

each i, elementary properties of quadratic variations yield that, as n tends to infinity, $S_{i,n}/n$ tends to Γ_i^{-1} in probability. On the other hand, the random matrix $V_{i,n}$ tends *a.s.* to the integral $V_i(T)$ and the random vector $U_{i,n}$ tends in probability to the stochastic integral

$$U_i(T) = \left(\int_0^T \frac{b_k(X_i(s))}{\sigma^2(X_i(s))} dX_i(s) \right)_{1 \le k \le d}. \tag{15}$$

Therefore $T_{i,n}(\boldsymbol{\mu}, \boldsymbol{\Omega})/n$ tends to 0 (see 11). This implies that, for all $i = 1, \ldots, N$ and for all (ϑ_0, ϑ), $\mathbb{P}_{\vartheta_0}(E_{i,n}(\vartheta)) \to 1$ as n tends to infinity. However, we need the more precise result

$$\forall \vartheta, \vartheta_0, \quad \mathbb{P}_{\vartheta_0}(\mathbf{E}_{N,n}(\vartheta)) \to 1.$$

Moreover, the set on which the approximate likelihood is considered should not depend on ϑ.

To this end, let us define, for $\alpha > 0$, using the notations of (H4),

$$\begin{aligned} M_{i,n} &= \max\{c_1 + 2, 2m^2\}(1 + \|U_{i,n}\|^2); \\ F_{i,n} &= \{S_{i,n} - M_{i,n} \ge \alpha\sqrt{n}\}; \quad \mathbf{F}_{N,n} = \cap_{i=1}^N F_{i,n}. \end{aligned} \tag{16}$$

Lemma 1 *Assume (H1)–(H4). For all ϑ satisfying (H4) and all i, we have $F_{i,n} \subset E_{i,n}(\vartheta)$. If $a_0 > 4$ and as $N, n \to \infty$, $N = N(n)$ is such that $N/n^2 \to 0$, then,*

$$\forall \vartheta_0, \quad \mathbb{P}_{\vartheta_0}(\boldsymbol{F}_{N,n}) \to 1.$$

For this, we prove that, if $a_0 > 4$,

$$\mathbb{P}_{\vartheta_0}(F_{1,n}^c) \lesssim \frac{1}{n^2}, \tag{17}$$

where $\lesssim$ means lower than or equal to up to a multiplicative known constant which does not depend on n. This explains the constraint $N/n^2 \to 0$.

For this proof, the condition $c_0 \le \lambda_{max}(\boldsymbol{\Omega})$ of (H4) is not required.

The condition $a_0 > 4$ is a moment condition for Ψ_i required by the proof. It is equivalent to the fact that $\mathbb{E}\Psi_i^8 = \mathbb{E}\Gamma_i^{-4} = \lambda_0^4/(a_0-1)(a_0-2)(a_0-3)(a_0-4) < +\infty$. Note that $\mathbb{E}\Psi_i^2 = \mathbb{E}\Gamma_i^{-1} = \lambda_0/(a_0-1)$ so that the random effect takes moderate values which is reasonable.

As a consequence of Lemma 1, for all ϑ, $\mathbb{P}_{\vartheta_0}(\mathbf{E}_{N,n}(\vartheta)) \to 1$ and (13) is well defined on the set $\mathbf{F}_{N,n}$ which is independent of ϑ and has probability tending to 1. The proof of Lemma 1 is surprisingly technical and detailed in an electronic supplementary material.

2.3 Second approximation of the likelihood

Formulae (13)–(14) suggest another approximation of the likelihood which is simpler. We give the heuristics for this approximation. We can write:

$$\begin{aligned}
\mathcal{L}_n(X_{i,n},\vartheta) &= \mathcal{L}_n^{(1)}(X_{i,n},\vartheta)\times\mathcal{L}_n^{(2)}(X_{i,n},\vartheta)\\
\mathcal{L}_n^{(1)}(X_{i,n},\vartheta) &= \mathcal{L}_n^{(1)}(X_{i,n},\lambda,a) = \frac{\lambda^a\Gamma(a+(n/2))}{\Gamma(a)\left(\lambda+\frac{S_{i,n}}{2}\right)^{a+(n/2)}}\\
\mathcal{L}_n^{(2)}(X_{i,n},\vartheta) &= \frac{\left(\lambda+\frac{S_{i,n}}{2}\right)^{a+(n/2)}}{\left(\lambda+\frac{S_{i,n}}{2}+\frac{T_{i,n}(\boldsymbol{\mu},\boldsymbol{\Omega})}{2}\right)^{a+(n/2)}}\frac{1}{(det(I_d+V_{i,n}\boldsymbol{\Omega}))^{1/2}}. \qquad (18)
\end{aligned}$$

The first term $\mathcal{L}_n^{(1)}(X_{i,n},\vartheta)$ only depends on (λ,a) and is equal to the approximate likelihood function obtained in Delattre et al. (2015) for $b\equiv 0$ and $\Gamma_i\sim G(a,\lambda)$. For the second term, we have:

$$\begin{aligned}
\log\mathcal{L}_n^{(2)}(X_{i,n},\vartheta) = &-\left(a+\frac{n}{2}\right)\log\left(1+\frac{1}{n}\frac{T_{i,n}(\boldsymbol{\mu},\boldsymbol{\Omega})}{\left(2\frac{\lambda}{n}+\frac{S_{i,n}}{n}\right)}\right)\\
&-\frac{1}{2}\log\left(det(I_d+V_{i,n}\boldsymbol{\Omega})\right).
\end{aligned}$$

As for n tending to infinity, $T_{i,n}(\boldsymbol{\mu},\boldsymbol{\Omega})$ tends to a fixed limit and $S_{i,n}/n$ tends to Γ_i^{-1}, this yields:

$$\begin{aligned}
\log\mathcal{L}_n^{(2)}(X_{i,n},\vartheta) &\sim -\frac{\left(a+\frac{n}{2}\right)}{n}\frac{T_{i,n}(\boldsymbol{\mu},\boldsymbol{\Omega})}{\left(2\frac{\lambda}{n}+\frac{S_{i,n}}{n}\right)}-\frac{1}{2}\log\left(det(I_d+V_{i,n}\boldsymbol{\Omega})\right)\\
&\sim -\frac{n}{2\,S_{i,n}}T_{i,n}(\boldsymbol{\mu},\boldsymbol{\Omega})-\frac{1}{2}\log\left(det(I_d+V_{i,n}\boldsymbol{\Omega})\right).
\end{aligned}$$

Now, the above expression only depends on $(\boldsymbol{\mu},\boldsymbol{\Omega})$. Morevover, as the term $U_{i,n}'V_{i,n}^{-1}U_{i,n}$ in $T_{i,n}(\boldsymbol{\mu},\boldsymbol{\Omega})$ does not contain parameters, we can forget it and set:

$$\begin{aligned}
\boldsymbol{V}_n(X_{i,n},\vartheta) &= \boldsymbol{V}_n^{(1)}(X_{i,n},\lambda,a)+\boldsymbol{V}_n^{(2)}(X_{i,n},\boldsymbol{\mu},\boldsymbol{\Omega}) \quad \text{with}\\
\boldsymbol{V}_n^{(1)}(X_{i,n},\lambda,a) &= a\log\lambda-\log\Gamma(a)-\left(a+\frac{n}{2}\right)\log\left(\lambda+\frac{S_{i,n}}{2}\right)\\
&\quad+\log\Gamma(a+(n/2)),\\
\boldsymbol{V}_n^{(2)}(X_{i,n},\boldsymbol{\mu},\boldsymbol{\Omega}) &= -\frac{n}{2\,S_{i,n}}(\boldsymbol{\mu}-V_{i,n}^{-1}U_{i,n})'R_{i,n}^{-1}(\boldsymbol{\mu}-V_{i,n}^{-1}U_{i,n})\\
&\quad-\frac{1}{2}\log\left(det(I_d+V_{i,n}\boldsymbol{\Omega})\right), \qquad (19)
\end{aligned}$$

and define the second approximation for the log-likelihood: $\boldsymbol{V}_{N,n}(\vartheta) = \sum_{i=1}^{N} \boldsymbol{V}_n(X_{i,n}, \vartheta)$. Thus, estimators of (λ, a) and $(\boldsymbol{\mu}, \boldsymbol{\Omega})$ can be computed separately. It is noteworthy that this second approximation overcomes the difficulties encountered in Lemma 1.

3 Asymptotic properties of estimators

In this section, we study the asymptotic behaviour of the estimators based on the two approximate likelihood functions of the previous section. To serve as a baseline, the estimation when an *i.i.d.* sample (Φ_i, Γ_i), $i = 1, \ldots, N$ is observed is presented in Sect. 7.4.

3.1 Estimation based on the first approximation of the likelihood

We study the estimation of $\vartheta = (\lambda, a, \boldsymbol{\mu}, \boldsymbol{\Omega})$ using the approximate likelihood $\mathcal{L}_{N,n}(\vartheta)$ given in Proposition 1 on the set $\mathbf{F_{N,n}}$ studied in Lemma 1 (see 13, 14, 16). Let

$$Z_{i,n}(a, \lambda, \boldsymbol{\mu}, \boldsymbol{\Omega}) = Z_i = \frac{\lambda + (S_{i,n}/2) + (T_{i,n}(\boldsymbol{\mu}, \boldsymbol{\Omega})/2)}{a + n/2}. \tag{20}$$

By (H4), using notations (16), we have $\lambda + S_{i,n}/2 + T_{i,n}(\boldsymbol{\mu}, \boldsymbol{\Omega})/2 \geq (S_{i,n} - M_{i,n})/2$. We refer to the supplementary material (proof of Lemma 1), where we give a proof of the nontrivial property that, if N/n^2 tends to 0, with probability tending to 1, $T_{i,n}(\boldsymbol{\mu}, \boldsymbol{\Omega}) \geq -M_{i,n}$ for all $i = 1, \ldots, N$.

More precisely, on the set $F_{i,n}$, $Z_{i,n} > \alpha/(\sqrt{n} + 2(\alpha_1/\sqrt{n})) > 0$ where α_1 is defined in (H4). Instead of considering $\log \mathcal{L}_n(X_i, \vartheta)$ on $F_{i,n}$, to define a contrast, we replace $\log Z_{i,n}$ by $1_{F_{i,n}} \log Z_{i,n}$ and set $\mathbf{U}_{N,n}(\vartheta) = \sum_{i=1}^{N} \mathbf{U}_n(X_i, \vartheta)$, with

$$\begin{aligned}\mathbf{U}_n(X_{i,n}, \vartheta) = {} & a \log \lambda - \log \Gamma(a) + \log \Gamma(a + (n/2)) - (a + (n/2)) \log(a + (n/2)) \\ & - \frac{1}{2} \log det(I_d + V_{i,n}\boldsymbol{\Omega}) - (a + (n/2)) 1_{F_{i,n}} \log Z_{i,n}.\end{aligned}$$

Define the pseudo-score function and the associated estimators $\tilde{\vartheta}_{N,n}$:

$$\mathcal{G}_{N,n}(\vartheta) = \nabla_\vartheta \mathbf{U}_{N,n}(\vartheta), \quad \mathcal{G}_{N,n}(\tilde{\vartheta}_{N,n}) = 0. \tag{21}$$

To investigate their asymptotic behaviour, we need to prove that $Z_{i,n}$ defined in (20) is close to Γ_i^{-1} and that $1_{F_{i,n}} Z_{i,n}^{-1}$ is close to Γ_i. For this, we introduce the random variable

$$S_{i,n}^{(1)} = S_i^{(1)} = \Psi_i^2 \sum_{j=1}^{n} \frac{(W_i(t_j) - W_i(t_{j-1}))^2}{\Delta} = \frac{1}{\Gamma_i} C_{i,n}^{(1)} \tag{22}$$

which corresponds to $S_{i,n}$ when $b(.) = 0, \sigma(.) = 1$. Then, we split $Z_{i,n} - \Gamma_i^{-1}$ into $Z_{i,n} - \frac{S_{i,n}^{(1)}}{n} + \frac{S_{i,n}^{(1)}}{n} - \Gamma_i^{-1}$ and study successively the two terms. The second term has

explicit distribution as $C_{i,n}^{(1)}$ has $\chi^2(n)$ distribution and is independent of Γ_i. The first term is treated below. We proceed analogously for $1_{F_{i,n}} Z_{i,n}^{-1} - \Gamma_i$ introducing $n/S_{i,n}^{(1)}$.

Lemma 2 *Assume (H1)–(H4). For all $p \geq 1$, we have*

$$\mathbb{E}_\vartheta\left(\left|Z_{i,n} - \frac{S_{i,n}^{(1)}}{n}\right|^p |\Psi_i = \psi, \Phi_i = \varphi\right) \lesssim \frac{1}{n^p}(1 + \varphi^{2p} + \psi^{2p} + \psi^{4p} + \varphi^{2p}\psi^{2p}), \tag{23}$$

$$\mathbb{E}_\vartheta\left(\left[\left|Z_{i,n}^{-1} - \frac{n}{S_{i,n}^{(1)}}\right|^p 1_{F_{i,n}}\right] |\Psi_i = \psi, \Phi_i = \varphi\right) \lesssim \frac{1}{n^p}(1 + (1 + \varphi^{2p})(\psi^{-2p} + \psi^{-4p}) + \varphi^{4p} + \psi^{4p} + \varphi^{4p}\psi^{-4p}). \tag{24}$$

Using (5), (6), (10), we set

$$\begin{aligned} A_{i,n} &= (I_d + V_{i,n}\boldsymbol{\Omega})^{-1}(U_{i,n} - V_{i,n}\boldsymbol{\mu}) \\ &= B_{i,n}(V_{i,n}^{-1}U_{i,n} - \boldsymbol{\mu}) \quad \text{where} \quad B_{i,n} = B_{i,n}(\boldsymbol{\Omega}) = R_{i,n}^{-1}. \end{aligned} \tag{25}$$

Let $\psi(u) = \frac{\Gamma'(u)}{\Gamma(u)}$ and F^c denote the complementary set of F. We obtain,using (20), (25),

$$\begin{aligned} \frac{\partial \mathbf{U}_{N,n}}{\partial \lambda}(\vartheta) &= \sum_{i=1}^N \left(\frac{a}{\lambda} - Z_{i,n}^{-1}\, 1_{F_{i,n}}\right), \\ \frac{\partial \mathbf{U}_{N,n}}{\partial a}(\vartheta) &= N\left(\psi(a + n/2) - \log(a + n/2) - \psi(a) + \log\lambda\right) \\ &\quad - \sum_{i=1}^N \left(1_{F_{i,n}} \log Z_{i,n} - 1_{F_{i,n}^c}\right), \\ \nabla_{\boldsymbol{\mu}} \mathbf{U}_{N,n}(\vartheta) &= -\frac{1}{2}\sum_{i=1}^N 1_{F_{i,n}} Z_{i,n}^{-1} \nabla_{\boldsymbol{\mu}} T_{i,n}(\boldsymbol{\mu}, \boldsymbol{\Omega}) = \sum_{i=1}^N 1_{F_{i,n}} Z_{i,n}^{-1} A_{i,n}, \\ \nabla_{\boldsymbol{\Omega}} \mathbf{U}_{N,n}(\vartheta) &= -\frac{1}{2}\sum_{i=1}^N \left(B_{i,n} + 1_{F_{i,n}} Z_{i,n}^{-1}\, \nabla_{\boldsymbol{\Omega}} T_{i,n}(\boldsymbol{\mu}, \boldsymbol{\Omega})\right) \\ &= \frac{1}{2}\sum_{i=1}^N \left(1_{F_{i,n}} Z_{i,n}^{-1}\, A_{i,n}A_{i,n}' - B_{i,n}\right). \end{aligned}$$

Using (7), (15), we set

$$B_i(T; \boldsymbol{\Omega}) = (V_i(T)^{-1} + \boldsymbol{\Omega})^{-1}, \quad A_i(T; \boldsymbol{\mu}, \boldsymbol{\Omega}) = B_i(T; \boldsymbol{\Omega})(V_i(T)^{-1}U_i(T) - \boldsymbol{\mu}) \tag{26}$$

and let $J(\vartheta)$ denote the covariance matrix of

$$\begin{pmatrix} \Gamma_1 A_1(T;\boldsymbol{\mu},\boldsymbol{\Omega}) \\ \frac{1}{2}\Gamma_1 A_1(T;\boldsymbol{\mu},\boldsymbol{\Omega})A_1'(T;\boldsymbol{\mu},\boldsymbol{\Omega}) - B_1(T;\boldsymbol{\Omega}) \end{pmatrix} \tag{27}$$

The Fisher information matrix associated with the direct observation $(\Gamma_1,\ldots,\Gamma_N)$ (see Sect. 7.4) is

$$I_0(\lambda,a) = \begin{pmatrix} \frac{a}{\lambda^2} & -\frac{1}{\lambda} \\ -\frac{1}{\lambda} & \psi'(a) \end{pmatrix} \tag{28}$$

We now state:

Theorem 1 *Assume (H1)–(H4), $a > 4$, and that n and $N = N(n)$ tend to infinity. Then, for all ϑ,*

- *If $N/n^2 \to 0$, $N^{-1/2}\left(\frac{\partial \boldsymbol{U}_{N,n}}{\partial\lambda}(\vartheta), \frac{\partial \boldsymbol{U}_{N,n}}{\partial a}(\vartheta)\right)'$ converges in distribution under $\mathbb{P}_\vartheta$ to the Gaussian law $\mathcal{N}_2(0, I_0(\lambda,a))$.*
- *If $N/n \to 0$, $N^{-1/2}\nabla_\vartheta \boldsymbol{U}_{N,n}(\vartheta)$ converges in distribution under $\mathbb{P}_\vartheta$ to $\mathcal{N}_q(0,\mathcal{J}(\vartheta))$ ($q = 2+d+d(d+1)/2$) where*

$$\mathcal{J}(\vartheta) = \left(\begin{array}{c|c} I_0(\lambda,a) & \boldsymbol{0} \\ \hline \boldsymbol{0} & J(\vartheta) \end{array}\right)$$

  with $J(\vartheta)$, $I_0(\lambda,a)$ defined in (27), (28).
- *Define the opposite of the Hessian of $\boldsymbol{U}_{N,n}(\vartheta)$ by $\mathcal{J}_{N,n}(\vartheta) = -\nabla^2_\vartheta \boldsymbol{U}_{N,n}(\vartheta)$. Then, if $a > 6$, as n and $N = N(n)$ tend to infinity, $\frac{1}{N}\mathcal{J}_{N,n}(\vartheta)$ converges in $\mathbb{P}_\vartheta$-probability to $\mathcal{J}(\vartheta)$.*

Remark 1 The constraint $a > 6$ might seem too stringent. Noting that $\mathbb{E}(\Gamma_i^{-1}) = \frac{\lambda}{a-1}$, it is indeed an assumption on the quadratic variations of $(X_i(t))$ and $a > 6$ corresponds to moderate quadratic variations.

Remark 2 In the case of univariate random effect $\Phi_i \sim \mathcal{N}(\mu,\omega^2)$, $J(\vartheta)$ writes,

$$J(\vartheta) = \begin{pmatrix} \mathbb{E}_\vartheta \Gamma_1 B_1(T;\omega^2) & \mathbb{E}_\vartheta \Gamma_1 A_1(T;\mu,\omega^2)B_1(T;\omega^2) \\ \mathbb{E}_\vartheta \Gamma_1 A_1(T;\mu,\omega^2)B_1(T;\omega^2) & \mathbb{E}_\vartheta\left(\Gamma_1 A_1^2(T;\mu,\omega^2)B_1(T;\omega^2) - \frac{1}{2}B_1^2(T;\omega^2)\right) \end{pmatrix}. \tag{29}$$

For the proof, we use that $A_{i,n}$ (resp. $B_{i,n}$) converges to $A_i(T;\boldsymbol{\mu},\boldsymbol{\Omega})$ (resp. $B_i(T;\boldsymbol{\Omega})$) as n tends to infinity and that $Z_{i,n}^{-1}1_{F_{i,n}}$ is close to Γ_i for large n. Note that $I_0(\lambda,a)$ is invertible for all $(\lambda,a) \in (0,+\infty)^2$ (see Sect. 7.4). We conclude:

Theorem 2 *Assume (H1)–(H4), $a_0 > 6$, that n and $N = N(n)$ tend to infinity with $N/n \to 0$ and that $J(\vartheta_0)$ is invertible. Then, with probability tending to 1, a solution $\tilde{\vartheta}_{N,n}$ to (21) exists which is consistent and such that $\sqrt{N}(\tilde{\vartheta}_{N,n} - \vartheta_0)$ converges in distribution to $\mathcal{N}_q(0,\mathcal{J}^{-1}(\vartheta_0))$ under $\mathbb{P}_{\vartheta_0}$ for all ϑ_0.*

For the first two components, the constraint $N/n^2 \to 0$ is enough.

The proof of Theorem 2 is deduced standardly from the previous theorem and omitted.

The estimators of (λ, a) are asymptotically equivalent to the exact maximum likelihood estimators of the same parameters based on the observation of $(\Gamma_i, i = 1, \ldots, N)$ under the constraint $N/n^2 \to 0$. There is no loss of information for parameters coming from the random effects in the diffusion coefficient. For the parameters $(\boldsymbol{\mu}, \boldsymbol{\Omega})$, which come from the random effects in the drift, the constraint is $N/n \to 0$ and there is a loss of information (see 49 in Sect. 7.4) w.r.t. the direct observation of $((\Phi_i, \Gamma_i), i = 1, \ldots, N)$. For instance, consider a univariate random effect $\Phi \sim \mathcal{N}(\mu, \omega^2)$. Then, if ω^2 is known, one can see that

$$\mathbb{E}_\vartheta \Gamma_1 B_1(T; \omega^2) \leq \mathbb{E}_\vartheta \Gamma_1 \omega^{-2} = \frac{a}{\lambda \omega^2}.$$

If for all i, $\Gamma_i = \gamma$ is deterministic, there is no loss of information for the estimation of (μ, ω^2), with respect to the continuous observation of the processes $(X_i(t), t \in [0, T])$ (see Delattre et al. 2013).

3.2 Estimation based on the second approximation of the likelihood

Now, we consider the second approximation of the loglikelihood (19). We set

$$\begin{aligned} \xi_{i,n} &= \frac{\lambda + (S_{i,n}/2)}{a + n/2}, \qquad \text{so that} \\ \mathbf{V}_n^{(1)}(X_{i,n}, \lambda, a) &= a \log \lambda - \log \Gamma(a) + \log \Gamma(a + (n/2)) \\ &\quad - (a + (n/2)) \log(a + (n/2)) - (a + (n/2)) \log \xi_{i,n}. \end{aligned} \tag{30}$$

We do not need to truncate $\xi_{i,n}$ as it is bounded from below. For the second term $\mathbf{V}_n^{(2)}(X_{i,n}, \boldsymbol{\mu}, \boldsymbol{\Omega})$, we need a truncation to deal with $n/S_{i,n}$ and make a slight modification. Let, for k a given constant,

$$\begin{aligned} \mathbf{W}_n(X_{i,n}, \boldsymbol{\mu}, \boldsymbol{\Omega}) &= -\frac{n}{2\, S_{i,n}} 1_{S_{i,n} \geq k\sqrt{n}} (\boldsymbol{\mu} - V_{i,n}^{-1} U_{i,n})' R_{i,n}^{-1} (\boldsymbol{\mu} - V_{i,n}^{-1} U_{i,n}) \\ &\quad - \frac{1}{2} \log\left(det(I_d + V_{i,n} \boldsymbol{\Omega})\right), \end{aligned} \tag{31}$$

and $\mathbf{W}_{N,n}(\boldsymbol{\mu}, \boldsymbol{\Omega}) = \sum_{i=1}^N \mathbf{W}_n(X_{i,n}, \boldsymbol{\mu}, \boldsymbol{\Omega})$.

We define the estimators $\vartheta_{N,n}^*$ by

$$\mathcal{H}_{N,n}(\vartheta) = (\nabla_{\lambda,a} \mathbf{V}_{N,n}^{(1)}(\lambda, a), \nabla_{\boldsymbol{\mu},\boldsymbol{\Omega}} \mathbf{W}_{N,n}(\boldsymbol{\mu}, \boldsymbol{\Omega})), \quad \mathcal{H}_{N,n}(\vartheta_{N,n}^*) = 0. \tag{32}$$

We have the following result.

Theorem 3 *Assume (H1)–(H4), $a_0 > 6$, that n and $N = N(n)$ tend to infinity with $N/n \to 0$ and that $J(\vartheta_0)$ is invertible. Then, with probability tending to 1, a solution $\vartheta_{N,n}^*$ to (32) exists which is consistent and such that $\sqrt{N}(\vartheta_{N,n}^* - \vartheta_0)$ converges in*

distribution to $\mathcal{N}_q(0, \mathcal{J}^{-1}(\vartheta_0))$ *under* $\mathbb{P}_{\vartheta_0}$ *for all* ϑ_0 *(see* (29), (28) *and the statement of Theorem* 2*). For the first two components, the constraint* $N/n^2 \to 0$ *is enough.*

The estimators $\vartheta^*_{N,n}$ *and* $\tilde{\vartheta}_{N,n}$ *are asymptotically equivalent.*

This equivalence result is quite important: it could be expected that Method 1, being a natural approximation of the likelihood, would lead to better asymptotic results. This is not the case. Implementing Method 2 to compute the estimators is numerically easier because it splits the estimation of random effect parameters in the diffusion coefficient and those in the drift coefficient. On simulated data, both methods are comparable even for reasonably small sample sizes.

4 Simulation study

We illustrate and assess the properties of the parameter estimators. Model parameters are estimated by the two approximations of the likelihood. We found little difference between results of both methods, and so we only report results for Method 2.

Several models are simulated:

Model 1 Mixed effects Brownian motion:

$$\begin{aligned} dX_i(t) =& \Phi_i dt + \Gamma_i^{-1/2} dW_i(t),\ X_i(0) = 0 \\ \Gamma_i \underset{i.i.d}{\sim} & G(a, \lambda)\ ,\ \Phi_i|\Gamma_i = \gamma \sim \mathcal{N}(\mu, \gamma^{-1}\omega^2) \end{aligned}$$

Model 2 A model satisfying (H1)–(H2) with constant $\sigma(\cdot)$:

$$\begin{aligned} dX_i(t) =& \frac{\Phi_i X_i(t)}{1 + X_i^2(t)} dt + \Gamma_i^{-1/2} dW_i(t),\ X_i(0) = 0 \\ \Gamma_i \underset{i.i.d}{\sim} & G(a, \lambda),\ \Phi_i|\Gamma_i = \gamma \sim \mathcal{N}(\mu, \gamma^{-1}\omega^2) \end{aligned}$$

Model 3 A model satisfying (H1)–(H2) with non constant $\sigma(\cdot)$:

$$dX_i(t) = \left(\frac{\Phi_{i,1} X_i(t)}{1 + X_i^2(t)} + \Phi_{i,2}\right) dt + \left(1 + \frac{X_i(t)}{1 + X_i^2(t)}\right) \Gamma_i^{-1/2} dW_i(t),\ X_i(0) = 0.1,$$

$\Gamma_i \sim G(a, \lambda)$, $\Phi_i = (\Phi_{i,1}, \Phi_{i,2})'|\Gamma_i = \gamma \sim \mathcal{N}_2(\boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Omega})$ with $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ and $\boldsymbol{\Omega} = \mathrm{diag}(\omega_1^2, \omega_2^2)$

Model 4 Mixed Ornstein–Uhlenbeck process:

$$dX_i(t) = (\Phi_{i,1} X_i(t) + \Phi_{i,2}) dt + \Gamma_i^{-1/2} dW_i(t),\ X_i(0) = 0,$$

$\Gamma_i \sim G(a, \lambda)$, $\Phi_i = (\Phi_{i,1}, \Phi_{i,2})'|\Gamma_i = \gamma \sim \mathcal{N}_2(\boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Omega})$ with $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ and $\boldsymbol{\Omega} = \mathrm{diag}(\omega_1^2, \omega_2^2)$.

Table 1 Empirical mean and (empirical s.d.) of estimators for different values of N and n, truncation $k = 0.1$, $m = a/\lambda$, $t = \psi(a) - \log\lambda$

Model 1: $dX_i(t) = \Phi_i dt + \Gamma_i^{-1/2} dW_i(t)$, $X_i(0) = 0$
$\Gamma_i \sim G(a, \lambda)$, $\Phi_i|\Gamma_i = \gamma \sim \mathcal{N}(\mu, \gamma^{-1}\omega^2)$

Parameter	True value	$N = 50$		$N = 100$	
		$\Delta = 0.025$ ($n = 200$)	$\Delta = 0.005$ ($n = 1000$)	$\Delta = 0.025$ ($n = 200$)	$\Delta = 0.005$ ($n = 1000$)
μ	-0.50	-0.49 (0.06)	-0.51 (0.06)	-0.48 (0.05)	-0.50 (0.05)
ω^2	0.50	0.45 (0.11)	0.49 (0.12)	0.44 (0.09)	0.48 (0.10)
m	4.00	3.85 (0.22)	3.96 (0.23)	3.87 (0.14)	3.97 (0.14)
t	1.32	1.24 (0.06)	1.30 (0.06)	1.24 (0.04)	1.30 (0.04)

Table 2 Empirical mean and (empirical s.d.) of estimators for different values of N and n, truncation $k = 0.1$, $m = a/\lambda$, $t = \psi(a) - \log\lambda$

Model 2: $dX_i(t) = \Phi_i X_i(t)/(1 + X_i^2(t))dt + \Gamma_i^{-1/2} dW_i(t)$, $X_i(0) = 0$
$\Gamma_i \sim G(a, \lambda)$, $\Phi_i|\Gamma_i = \gamma \sim \mathcal{N}(\mu, \gamma^{-1}\omega^2)$

Parameter	True value	$N = 50$		$N = 100$	
		$\Delta = 0.025$ ($n = 200$)	$\Delta = 0.005$ ($n = 1000$)	$\Delta = 0.025$ ($n = 200$)	$\Delta = 0.005$ ($n = 1000$)
μ	-2.00	-1.97 (0.16)	-2.03 (0.16)	-1.93 (0.11)	-1.99 (0.12)
ω^2	0.50	0.53 (0.64)	0.61 (0.64)	0.41 (0.39)	0.49 (0.42)
m	4.00	4.09 (0.22)	4.00 (0.21)	4.10 (0.14)	4.02 (0.13)
t	1.32	1.30 (0.05)	1.31 (0.05)	1.30 (0.03)	1.32 (0.03)

Note that Model 4 does not fulfill assumptions (H1)–(H2) but it is widely used in practice and the estimation results show that the estimation methods still perform well.

For each SDE model, 100 data sets are generated with N subjects on the same time interval $[0, T]$, $T = 5$. Each data set is simulated as follows. First, the random effects are drawn, then, the diffusion sample paths are simulated with a very small discretization step-size $\delta = 0.001$. Exact simulation is performed for Models 1 and 4 whereas a Euler discretization scheme is used for Models 2 and 3. The time interval between consecutive observations is taken equal to $\Delta = 0.025$ or 0.005 with a resulting number of observations $n = 200, 1000$ and fixed time interval $[0, 5]$. The model parameters are then estimated by using the second approximation of the likelihood from each simulated dataset. The empirical mean and standard deviation of the estimates are computed from the 100 datasets (Tables 1, 2, 3 and 4). For Gamma distributions, the parameters $m = a/\lambda$ and $t = \psi(a) - \log(\lambda)$ have unbiased empirical estimators and are easier to estimate. This is why we present the estimates of m and t rather than the estimates of a and λ which are highly biased even for direct observations. The estimation procedure requires a truncation (see 31), we use $k = 0.1$.

Table 3 Empirical mean and (empirical s.d.) of estimators for different values of N and n, truncation $k = 0.1$, $m = a/\lambda$, $t = \psi(a) - \log\lambda$

Model 3: $dX_i(t) = \left(\frac{\Phi_{i,1}X_i(t)}{1+X_i^2(t)} + \Phi_{i,2}\right)dt + \left(1 + \frac{X_i(t)}{1+X_i^2(t)}\right)\Gamma_i^{-1/2}dW_i(t)$, $X_i(0) = 0$
$\Gamma_i \sim G(a, \lambda)$, $\Phi_i = (\Phi_{i,1}, \Phi_{i,2})'|\Gamma_i = \gamma \sim \mathcal{N}_2(\boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Omega})$ with $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ and $\boldsymbol{\Omega} = \text{diag}(\omega_1^2, \omega_2^2)$

Parameter	True value	$N = 50$		$N = 100$	
		$\Delta = 0.025$ ($n = 200$)	$\Delta = 0.005$ ($n = 1000$)	$\Delta = 0.025$ ($n = 200$)	$\Delta = 0.005$ ($n = 1000$)
μ_1	−2.00	−1.95 (0.18)	−1.99 (0.19)	−1.92 (0.14)	−1.97 (0.14)
μ_2	−1.00	−0.96 (0.08)	−0.99 (0.09)	−0.96 (0.06)	−0.98 (0.06)
ω_1^2	0.25	0.37 (0.45)	0.38 (0.43)	0.26 (0.32)	0.28 (0.32)
ω_2^2	0.25	0.21 (0.09)	0.23 (0.08)	0.22 (0.08)	0.23 (0.08)
m	4.00	3.93 (0.18)	3.97 (0.17)	3.97 (0.14)	4.00 (0.15)
t	1.32	1.26 (0.05)	1.30 (0.04)	1.27 (0.04)	1.31 (0.04)

Table 4 Empirical mean and (empirical s.d.) of estimators for different values of N and n, truncation $k = 0.1$, $m = a/\lambda$, $t = \psi(a) - \log\lambda$

Model 4: $dX_i(t) = (\Phi_{i,1}X_i(t) + \Phi_{i,2})dt + \Gamma_i^{-1/2}dW_i(t)$, $X_i(0) = 0$
$\Gamma_i \sim G(a, \lambda)$, $\Phi_i = (\Phi_{i,1}, \Phi_{i,2})'|\Gamma_i = \gamma \sim \mathcal{N}_2(\boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Omega})$ with $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ and $\boldsymbol{\Omega} = \text{diag}(\omega_1^2, \omega_2^2)$

Parameter	True value	$N = 50$		$N = 100$	
		$\Delta = 0.025$ ($n = 200$)	$\Delta = 0.005$ ($n = 1000$)	$\Delta = 0.025$ ($n = 200$)	$\Delta = 0.005$ ($n = 1000$)
μ_1	−0.50	−0.48 (0.07)	−0.49 (0.07)	−0.49 (0.04)	−0.50 (0.04)
μ_2	1.00	0.97 (0.11)	0.99 (0.11)	0.97 (0.06)	0.99 (0.06)
ω_1^2	0.10	0.09 (0.06)	0.09 (0.06)	0.09 (0.05)	0.10 (0.05)
ω_2^2	0.50	0.49 (0.20)	0.52 (0.23)	0.48 (0.13)	0.50 (0.13)
m	4.00	3.96 (0.22)	4.00 (0.21)	3.95 (0.12)	4.00 (0.12)
t	1.32	1.27 (0.06)	1.31 (0.05)	1.26 (0.03)	1.31 (0.03)

We observe from Tables 1, 2, 3 and 4 that the estimation method has satisfactory performances overall. The parameters are estimated with very little bias whatever the model and the values of n and N. The bias becomes smaller as both N and n increase. The standard deviations of the estimates are small, they become smaller when N increases and they do not depend on n. These results are in accordance with the theory since the asymptotic distribution of the estimates is obtained when $N/n \to 0$ or $N/n^2 \to 0$ respectively, and the rate of convergence is $\sqrt{N}$. Although for $n = 200$, we don't have N/n very small in our simulation design, the results in this case are quite satisfactory which encourages the use of this estimation method not only for high but also moderate values of numbers of observations per path n. In all the examples below, the Gamma distribution parameters are $a = 8$, $\lambda = 2$. The associated theoretical s.d. for m and t for direct observations are respectively 0.2, 0.05 for $N = 50$ and 0.14, 0.04

Table 5 Empirical mean and standard deviation from 100 datasets; theoretical s.d. of estimators for different values of N, $a = 8$, $\lambda = 2$, $m = a/\lambda$, $t = \psi(a) - \log\lambda$

Direct observations of N i.i.d. (Φ_i, Γ_i) s.t. $\Gamma_i \sim G(a, \lambda)$, $\Phi_i|\Gamma_i = \gamma \sim \mathcal{N}_2(\boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Omega})$ with $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ and $\boldsymbol{\Omega} = \mathrm{diag}(\omega_1^2, \omega_2^2)$

Parameter	True value	$N = 50$			$N = 100$		
		Mean	s.d.	Theor. s.d.	Mean	s.d.	Theor. s.d.
μ_1	−0.50	−0.50	0.02	0.02	−0.50	0.01	0.02
μ_2	1.00	1.00	0.05	0.05	1.00	0.03	0.04
ω_1^2	0.10	0.10	0.02	0.02	0.10	0.01	0.01
ω_2^2	0.50	0.51	0.11	0.10	0.50	0.07	0.07
m	4.00	4.01	0.21	0.20	4.01	0.12	0.14
t	1.32	1.33	0.05	0.05	1.33	0.03	0.04

for $N = 100$ (see 7.4 and Table 5). This agrees with the results obtained in Examples 1–4 for (a, λ) .

5 Concluding comments

In this paper, we have addressed the new problem of estimating parameters when having discrete observations and simultaneous random effects in the drift and in the diffusion coefficients by means of a joint parametric distribution. We have considered N paths and n observations per path. For linear random effects and a specific joint distribution for these random effects, we have proved that the model parameters in the drift and in the diffusion can be estimated consistently and with a rate $\sqrt{N}$ under the condition $N/n \to 0$. For the parameters in the diffusion coefficient, the constraint is weaker ($N/n^2 \to 0$). We have proposed two methods for the estimation leading to asymptotically equivalent properties. The second one is now implemented in a R package (MsdeParEst). Including random effects both in the drift and in the diffusion coefficient of SDE has been proposed in several data applications but to our knowledge, it had not been studied from a theoretical point of view.

Our results are obtained for a regular sampling but could be easily extended to non regular ones. They are derived in an asymptotic framework. An important issue would be to study non asymptotic properties. It would also be interesting to consider more general SDEMEs where fixed and random effects are no longer linear and different distributions for the random effects.

6 Appendix

6.1 Proof of Proposition 1

For the likelihood of the i-th vector of observations, we first compute the conditional likelihood given $\Phi_i = \varphi$, $\Psi_i = \psi$, and then integrate the result with respect to the joint

distribution of (Φ_i, Ψ_i). We replace the exact likelihood given fixed values (φ, ψ), *i.e.* the likelihood of $(X_i^{\varphi,\psi}(t_j), j = 1, \ldots, n)$ by the Euler scheme likelihood of (8). Setting $\psi = \gamma^{-1/2}$, it is given up to constants by:

$$L_n(X_{i,n}, \gamma, \varphi) = L_n(X_i, \gamma, \varphi) = \gamma^{n/2} \exp\left[-\frac{\gamma}{2}(S_{i,n} + \varphi' V_{i,n}\varphi - 2\varphi' U_{i,n})\right]. \quad (33)$$

The unconditional approximate likelihood is obtained integrating with respect to the joint distribution $\nu_\vartheta(d\gamma, d\varphi)$ of the random effects $(\Gamma_i = \Psi_i^{-2}, \Phi_i)$. For this, we first integrate $L_n(X_i, \gamma, \varphi)$ with respect to the Gaussian distribution $\mathcal{N}_d(\boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Omega})$. Then, we integrate the result w.r.t. the distribution of Γ_i. This second integration is only possible on the subset $E_{i,n}(\vartheta)$ defined in (12).

Assume first that $\boldsymbol{\Omega}$ is invertible. Integrating (33) with respect to the distribution $\mathcal{N}(\boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Omega})$ yields the expression:

$$\begin{aligned}
\Lambda_n(X_{i,n}, \gamma, \boldsymbol{\mu}, \boldsymbol{\Omega}) &= \gamma^{n/2} \exp\left(-\frac{\gamma}{2} S_{i,n}\right) \frac{\gamma^{d/2}}{(2\pi)^{d/2}(det(\boldsymbol{\Omega}))^{1/2}} \\
&\quad \times \int_{\mathbb{R}^d} \exp\left(\gamma\left(\varphi' U_{i,n} - \frac{1}{2}\varphi' V_{i,n}\varphi\right)\right) \\
&\quad \times \exp\left(-\frac{\gamma}{2}(\varphi - \boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\varphi - \boldsymbol{\mu})\right) d\varphi \\
&= \gamma^{n/2} \exp\left(-\frac{\gamma}{2} S_{i,n}\right) \left(\frac{det(\boldsymbol{\Sigma}_{i,n})}{det(\boldsymbol{\Omega})}\right)^{1/2} \exp\left(-\frac{\gamma}{2} T_{i,n}(\boldsymbol{\mu}, \boldsymbol{\Omega})\right)
\end{aligned}$$

where

$$T_{i,n}(\boldsymbol{\mu}, \boldsymbol{\Omega}) = \boldsymbol{\mu}'\boldsymbol{\Omega}^{-1}\boldsymbol{\mu} - \mathbf{m}_{i,n}'\boldsymbol{\Sigma}_{i,n}^{-1}\mathbf{m}_{i,n}, \quad (34)$$

and

$$\boldsymbol{\Sigma}_{i,n} = \boldsymbol{\Omega}(I_d + V_{i,n}\boldsymbol{\Omega})^{-1}, \qquad \mathbf{m}_{i,n} = \boldsymbol{\Sigma}_{i,n}(U_{i,n} + \boldsymbol{\Omega}^{-1}\boldsymbol{\mu}).$$

Computations using matrices equalities and (5), (6), (10) yield that $T_{i,n}(\boldsymbol{\mu}, \boldsymbol{\Omega})$ is equal to the expression given in (11), *i.e.*:

$$T_{i,n}(\boldsymbol{\mu}, \boldsymbol{\Omega}) = (\boldsymbol{\mu} - V_{i,n}^{-1}U_{i,n})' R_{i,n}^{-1}(\boldsymbol{\mu} - V_{i,n}^{-1}U_{i,n}) - U_{i,n}' V_{i,n}^{-1} U_{i,n}.$$

Noting that $\frac{det(\boldsymbol{\Sigma}_{i,n})}{det(\boldsymbol{\Omega})} = (det(I_d + V_{i,n}\boldsymbol{\Omega}))^{-1}$, we get

$$\Lambda_n(X_{i,n}, \gamma, \boldsymbol{\mu}, \boldsymbol{\Omega}) = \gamma^{n/2}(det(I_d + V_{i,n}\boldsymbol{\Omega}))^{-1/2} \exp\left[-\frac{\gamma}{2}\left(S_{i,n} + T_{i,n}(\boldsymbol{\mu}, \boldsymbol{\Omega})\right)\right]. \tag{35}$$

Then, we multiply $\Lambda_n(X_{i,n}, \gamma, \boldsymbol{\mu}, \boldsymbol{\Omega})$ by the Gamma density $(\lambda^a/\Gamma(a))\gamma^{a-1}$ $\exp(-\lambda\gamma)$, and on the set $E_{i,n}(\vartheta)$ (see 12), we can integrate w.r.t. to γ on $(0, +\infty)$. This gives $\mathcal{L}_n(X_{i,n}, \vartheta)$.

At this point, we observe that the formula (14) and the set $E_{i,n}(\vartheta)$ are still well defined for non invertible $\boldsymbol{\Omega}$. Consequently, we can consider $\mathcal{L}_n(X_{i,n}, \vartheta)$ as an approximate likelihood for non invertible $\boldsymbol{\Omega}$. □

6.2 Proof of Theorem 1

For simplicity of notations, we consider the case $d = 1$ (univariate random effect in the drift, *i.e.* $\mu = \boldsymbol{\mu}, \omega^2 = \boldsymbol{\Omega}$). The case $d > 1$ does not present additional difficulties.

Proof of Lemma 2 We omit the index n but keep the index i for the i-th sample path. We have (see 20):

$$Z_i - \frac{S_i^{(1)}}{n} = \frac{n}{2a+n}\left(\frac{S_i}{n} - \frac{S_i^{(1)}}{n}\right) - \frac{2a}{2a+n}\frac{S_i^{(1)}}{n} + \frac{2\lambda + T_i(\mu, \omega^2)}{2a+n}.$$

Therefore, using definition (22) and the notations introduced in (H4) (bounds on the parameter set),

$$\left|Z_i - \frac{S_i^{(1)}}{n}\right| \leq \left|\frac{S_i}{n} - \frac{S_i^{(1)}}{n}\right| + \frac{2\alpha_1}{n}\Psi_i^2\frac{C_i^{(1)}}{n} + \frac{2\ell_1}{n} + \frac{|T_i(\mu, \omega^2)|}{n} \tag{36}$$

Note that, using (11):

$$\begin{aligned} T_i(\mu, \omega^2) = &\frac{V_i}{(1+\omega^2 V_i)}\left(\mu - \frac{U_i}{V_i}\right)^2 - \frac{U_i^2}{V_i} = \frac{\mu^2 V_i}{(1+\omega^2 V_i)} \\ &- \frac{\omega^2 U_i^2}{(1+\omega^2 V_i)} - 2\mu\frac{U_i}{(1+\omega^2 V_i)}. \end{aligned} \tag{37}$$

Thus, using (H4) and the fact that $x + x^2 \leq 2(1 + x^2)$ for all $x \geq 0$,

$$\begin{aligned} |T_i(\mu, \omega^2)| &\leq \frac{\mu^2}{\omega^2} + \omega^2 U_i^2 + 2|\mu|U_i| \leq \frac{m^2}{c_0} + c_1 U_i^2 + 2m|U_i| \\ &\leq \frac{m^2}{c_0} + 2\max\{c_1, 2m\}(1 + U_i^2) \lesssim 1 + U_i^2. \end{aligned}$$

By the Hölder inequality, we obtain using (6), (36) and (37):

$$\left|Z_i - \frac{S_i^{(1)}}{n}\right|^p \lesssim \left|\frac{S_i}{n} - \frac{S_i^{(1)}}{n}\right|^p + n^{-p}\left[1 + \Psi_i^{2p}\left(\frac{C_i^{(1)}}{n}\right)^p + U_i^{2p}\right].$$

Now, we apply Lemmas (7) and (5) of Sect. 7.2 and (48) of Sect. 7.3 and this yields inequality (23) of Lemma 2.

For the second inequality, we write:

$$\begin{aligned} Z_i^{-1} - \frac{n}{S_i^{(1)}} &= Z_i^{-1}\frac{n}{S_i^{(1)}}\left(\frac{S_i^{(1)}}{n} - Z_i\right) \\ &= \left(Z_i^{-1} - \frac{n}{S_i^{(1)}} + \frac{n}{S_i^{(1)}}\right)\frac{n}{S_i^{(1)}}\left(\frac{S_i^{(1)}}{n} - Z_i\right) \\ &= \left(\frac{n}{S_i^{(1)}}\right)^2 Z_i^{-1}\left(\frac{S_i^{(1)}}{n} - Z_i\right)^2 + \left(\frac{n}{S_i^{(1)}}\right)^2\left(\frac{S_i^{(1)}}{n} - Z_i\right). \end{aligned}$$

On F_i, $Z_i \geq c/\sqrt{n}$, so:

$$\left|Z_i^{-1} - \frac{n}{S_i^{(1)}}\right| 1_{F_i} \leq \left(\frac{n}{C_i^{(1)}}\right)^2 \Psi_i^{-4}\left(\frac{\sqrt{n}}{c}\left(\frac{S_i^{(1)}}{n} - Z_i\right)^2 + \left|\frac{S_i^{(1)}}{n} - Z_i\right|\right).$$

Consequently, by the Hölder inequality,

$$\left|Z_i^{-1} - \frac{n}{S_i^{(1)}}\right|^p 1_{F_i} \lesssim \left(\frac{n}{C_i^{(1)}}\right)^{2p} \Psi_i^{-4p}\left(n^{p/2}\left(\frac{S_i^{(1)}}{n} - Z_i\right)^{2p} + \left|\frac{S_i^{(1)}}{n} - Z_i\right|^p\right). \tag{38}$$

Now, we take conditional expectation w.r.t. $\Psi_i = \psi$, $\Phi_i = \varphi$ and apply the Cauchy–Schwarz inequality. We use that, for n large enough, (see (48) of Sect. 7.3):

$$\mathbb{E}_\vartheta\left(\left(\frac{n}{C_i^{(1)}}\right)^{4p} | \Psi_i = \psi, \Phi_i = \varphi\right) = \mathbb{E}\left(\frac{n}{C_i^{(1)}}\right)^{4p} = O(1).$$

And, we apply inequality (23) to get (24). □

Now, we start proving Theorem 1. Omitting the index n in $A_{i,n}$, $B_{i,n}$, we have [see (21), (25), (26)]

$$N^{-1/2}\frac{\partial \mathbf{U}_{N,n}}{\partial \lambda}(\vartheta) = N^{-1/2}\sum_{i=1}^{N}\left(\frac{a}{\lambda} - \Gamma_i\right) + R_1,$$

$$N^{-1/2}\frac{\partial \mathbf{U}_{N,n}}{\partial a}(\vartheta) = N^{-1/2}\sum_{i=1}^{N}(-\psi(a) + \log\lambda + \log\Gamma_i) + R_2 + R_2',$$

$$N^{-1/2}\frac{\partial \mathbf{U}_{N,n}}{\partial \mu}(\vartheta) = N^{-1/2}\sum_{i=1}^{N}\Gamma_i A_i + R_3 = N^{-1/2}\sum_{i=1}^{N}\Gamma_i A_i(T;\mu,\omega^2) + R_3' + R_3$$

$$\begin{aligned} N^{-1/2}\frac{\partial \mathbf{U}_{N,n}}{\partial \omega^2}(\vartheta) &= N^{-1/2}\frac{1}{2}\sum_{i=1}^{N}\left(\Gamma_i A_i^2 - B_i\right) + R_4 \\ &= N^{-1/2}\frac{1}{2}\sum_{i=1}^{N}(\Gamma_i A_i^2(T;\mu,\omega^2) - B_i(T,\omega^2)) + R_4' + R_4. \end{aligned}$$

The remainder terms are:

$$R_1 = N^{-1/2}\sum_{i=1}^{N}(\Gamma_i - Z_i^{-1}1_{F_i}), \quad R_2 = N^{-1/2}\sum_{i=1}^{N}(\log Z_i^{-1}\, 1_{F_i} - \log\Gamma_i),$$

$$R_2' = N^{1/2}(\psi(a + (n/2)) - \log(a + (n/2))) + N^{-1/2}\sum_{i=1}^{N}1_{F_i^c},$$

$$R_3 = N^{-1/2}\sum_{i=1}^{N}A_i(Z_i^{-1}1_{F_i} - \Gamma_i), \quad R_4 = N^{-1/2}\frac{1}{2}\sum_{i=1}^{N}A_i^2(Z_i^{-1}1_{F_i} - \Gamma_i),$$

$$R_3' = N^{-1/2}\sum_{i=1}^{N}\Gamma_i(A_i - A_i(T;\mu,\omega^2)),$$

$$R_4' = N^{-1/2}\frac{1}{2}\sum_{i=1}^{N}\left((B_i - B_i(T,\omega^2)) - \Gamma_i(A_i^2 - A_i^2(T;\mu,\omega^2))\right).$$

The most difficult remainder terms are R_1, R_2, R_3, R_4. They are treated in Lemma 3 below. For the term R_2', we use that $(\psi(a+(n/2)) - \log(a+(n/2))) = O(n^{-1})$ (see 47) and that, for $a > 4$, $\mathbb{P}_\theta(F_i^c) \lesssim n^{-2}$ (see 17) to get that $R_2' = O_P(\sqrt{N}/n)$.

Using Lemma 5 in Sect. 7.2, it is easy to check that R_3' and R_4' are $O_P(\sqrt{N/n})$.

Therefore, there remains to find the limiting distribution of:

$$\begin{pmatrix} N^{-1/2}\sum_{i=1}^{N}\left(\frac{a}{\lambda} - \Gamma_i\right), \\ N^{-1/2}\sum_{i=1}^{N}(-\psi(a) + \log\lambda + \log\Gamma_i), \\ N^{-1/2}\sum_{i=1}^{N}\Gamma_i A_i(T;\mu,\omega^2), \\ N^{-1/2}\frac{1}{2}\sum_{i=1}^{N}(B_i(T,\omega^2) - \Gamma_i A_i^2(T;\mu,\omega^2)) \end{pmatrix}.$$

The first two components are exactly the score function corresponding to the exact observation of $(\Gamma_i, i = 1, \ldots, N)$ (see Sect. 7.4). Hence, the first part of Theorem 1 is proved.

The whole vector is ruled by the standard central limit theorem. To compute the limiting distribution, we use results from Delattre et al. (2013) which deals with the case of $\Gamma_i = \gamma$ fixed and $\Phi_i \sim \mathcal{N}(\mu, \gamma^{-1}\omega^2)$. It is proved in this paper that

$$\mathbb{E}_\vartheta(A_i(T;\mu,\omega^2)|\Gamma_i = \gamma) = 0, \quad \mathbb{E}_\vartheta(B_i(T,\omega^2) - \Gamma_i A_i^2(T;\mu,\omega^2)|\Gamma_i = \gamma) = 0.$$

This result is stated for $\gamma = 1$ in Proposition 5, p. 328 of Delattre et al. (2013) with (unfortunately) different notations: $A_i(T;\mu,\omega^2)$ is denoted $\gamma_i(\theta)$ and $B_i(T,\omega^2)$ is denoted $I_i(\omega^2)$ (formula (11) of this paper). It can be extended for any value of γ using formula (35) and the regularity properties of the statistical model. Hence, the third and fourth component are centered and the covariances between the first two components and the last two ones are null. Moreover, it is also proved (in the same Proposition 5) that

$$\begin{aligned}
&\gamma\mathbb{E}_\vartheta(A_i^2(T;\mu,\omega^2)|\Gamma_i = \gamma) = \mathbb{E}_\vartheta(B_i(T;\omega^2)|\Gamma_i = \gamma),\\
&\mathbb{E}_\vartheta\left(\frac{1}{4}\left(\gamma A_i^2(T;\mu,\omega^2) - B_i(T;\omega^2)\right)^2|\Gamma_i = \gamma\right)\\
&\quad = \frac{1}{2}\mathbb{E}_\vartheta\left(2\gamma A_i^2(T;\mu,\omega^2)B_i(T;\omega^2) - B_i^2(T;\omega^2)|\Gamma_i = \gamma\right)\\
&\mathbb{E}_\vartheta\left(\frac{1}{2}(\gamma A_i^2(T;\mu,\omega^2) - B_i(T;\omega^2))A_i(T;\mu,\omega^2)|\Gamma_i = \gamma\right)\\
&\quad = \mathbb{E}_\vartheta\left(A_i(T;\mu,\omega^2)B_i(T,\omega^2)|\Gamma_i = \gamma\right).
\end{aligned}$$

Hence, the covariance matrix of the last two components is equal to $J(\vartheta)$ defined in (29).

The proof of the last item (second order derivatives) relies on the same tools with more cumbersome computations but no additional difficulty. It is detailed in the electronic supplementary material. Note that this part only requires that N, n both tend to infinity without further constraint. So the proof is complete. □

Lemma 3 *Recall* (20). *Then, for* $a > 4$, *(see* (16) *for the definition of* $F_{i,n}$*),* R_1, R_2 *are* $O_P(\max\{1/\sqrt{n}, \frac{\sqrt{N}}{n}\})$, R_3, R_4 *are* $O_P(\sqrt{\frac{N}{n}})$.

Proof The proof goes in several steps. We have introduced

$$S_i^{(1)} = \Psi_i^2 \frac{1}{\Delta}\sum_{j=1}^{n}\left(W_i(t_j) - W_i(t_{j-1})\right)^2 = \Gamma_i^{-1}C_i^{(1)}.$$

We know the exact distribution of $S_i^{(1)}$: $C_i^{(1)}$ is independent of Γ_i and has distribution $\chi^2(n) = G(n/2, 1/2)$. By exact computations, using Gamma distributions (see Sect. 7), we obtain:

$$\mathbb{E}_\vartheta\left(\frac{n}{C_i^{(1)}}-1\right)=\frac{2}{n-2},\quad \mathbb{E}_\vartheta\left(\frac{n}{C_i^{(1)}}-1\right)^2 = \frac{2n+8}{(n-2)(n-4)}=O(n^{-1}),$$

$$\mathbb{E}_\vartheta(\log C_i^{(1)}/2)-\log(n/2)=\psi(n/2)-\log(n/2)=O(n^{-1}),$$
$$\mathrm{Var}_\vartheta(\log C_i^{(1)}/2)=\psi'(n/2)=O(n^{-1}).$$

Thus,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N\left(\frac{n}{S_i^{(1)}}-\Gamma_i\right)=O_P\left(\max\left\{\frac{1}{\sqrt{n}},\frac{\sqrt{N}}{n}\right\}\right),\ \frac{1}{\sqrt{N}}\sum_{i=1}^N\left(\log\frac{n}{S_i^{(1)}}-\log\Gamma_i\right)$$
$$=O_P\left(\max\left\{\frac{1}{\sqrt{n}},\frac{\sqrt{N}}{n}\right\}\right) \tag{39}$$

Let us detail the first computation. Note that $(\Gamma_i, i=1,\dots,N)$ and $(C_i^{(1)}, i=1,\dots,N)$ are independent. Thus,

$$\begin{aligned}\mathbb{E}_\vartheta\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N\left(\frac{n}{S_i^{(1)}}-\Gamma_i\right)\right)^2 &= \frac{1}{N}\mathbb{E}_\vartheta\left(\sum_{i=1}^N\Gamma_i\left(\frac{n}{C_i^{(1)}}-1\right)\right)^2\\ &= \frac{a^2}{\lambda(\lambda+1)}\mathbb{E}_\vartheta\left(\frac{n}{C_1^{(1)}}-1\right)^2\\ &\quad+(N-1)\frac{a^2}{\lambda^2}\left(\mathbb{E}_\vartheta\left(\frac{n}{C_1^{(1)}}-1\right)\right)^2\\ &= O(n^{-1})+O(N/n^2)=O(\max\{n^{-1},N/n^2\}).\end{aligned}$$

The second computation is similar.

Then, we have to study

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N\left(Z_i^{-1}1_{F_i}-\frac{n}{S_i^{(1)}}\right),\quad \frac{1}{\sqrt{N}}\sum_{i=1}^N\left(\log(Z_i^{-1})1_{F_i}-\log\frac{n}{S_i^{(1)}}\right). \tag{40}$$

We write:

$$\left(Z_i^{-1}1_{F_i}-\frac{n}{S_i^{(1)}}\right)=\left(Z_i^{-1}-\frac{n}{S_i^{(1)}}\right)1_{F_i}-\frac{n}{S_i^{(1)}}1_{F_i^c}.$$

For $a > 4$, $\mathbb{P}_\vartheta(F_i^c) \lesssim n^{-2}$ (see 17). We have, applying the Cauchy–Schwarz inequality:

$$\mathbb{E}_\vartheta\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \frac{n}{S_i^{(1)}} 1_{F_i^c}\right) \le \frac{1}{\sqrt{N}}\sum_{i=1}^N \left(\mathbb{E}_\vartheta \Gamma_i^2\, \mathbb{E}_\vartheta (\frac{n}{C_i^{(1)}})^2\, \mathbb{P}_\vartheta(F_i^c)\right)^{1/2} \lesssim \frac{\sqrt{N}}{n}.$$

Next, apply Lemma 2,

$$\begin{aligned}&\mathbb{E}_\vartheta\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N (Z_i^{-1} - \frac{n}{S_i^{(1)}}) 1_{F_i}\right| \\ &\lesssim \frac{\sqrt{N}}{n}\mathbb{E}_\vartheta\left(1 + (1+\Phi_i^2)(\Psi_i^{-2} + \Psi_i^{-4}) + \Phi_i^4 + \Psi_i^4 + \Phi_i^4\Psi_i^{-4}\right).\end{aligned}$$

As noted above, $\Psi_i = \Gamma_i^{-1/2}$, $\mathbb{E}_\vartheta(\Psi_i^{-q}) < +\infty$ for all $q \ge 0$. We can write $\Phi_i = \mu + \omega\Psi_i\varepsilon_i$ with ε_i a standard Gaussian variable independent of Ψ_i. We have

$$\begin{aligned}\mathbb{E}_\vartheta \Phi_i^2\Psi_i^{-2} &= \mathbb{E}_\vartheta(\mu\Psi_i^{-1} + \omega\varepsilon_i)^2 < +\infty,\\ \mathbb{E}_\vartheta \Phi_i^2\Psi_i^{-4} &= \mathbb{E}_\vartheta(\mu\Psi_i^{-2} + \omega\Psi_i^{-1}\varepsilon_i)^2 < +\infty, \quad \text{and for } a > 2,\\ \mathbb{E}_\vartheta \Phi_i^4 &\lesssim 1 + \mathbb{E}_\vartheta \Psi_i^4 = 1 + \mathbb{E}_\vartheta \Gamma_i^{-2} < +\infty.\end{aligned}$$

Thus, for $a > 4$,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N \left(Z_i^{-1} 1_{F_i} - \frac{n}{S_i^{(1)}}\right) = O_P\left(\frac{\sqrt{N}}{n}\right). \tag{41}$$

Joining (39) (left)–(40) (left) and (41), we obtain that, for $a > 4$, $R_1 = O_P(\max\{n^{-1/2}, \sqrt{N}/n\})$.

Analogously,

$$\left(\log(Z_i^{-1}) 1_{F_i} - \log\frac{n}{S_i^{(1)}}\right) = \left(\log(Z_i^{-1}) - \log\frac{n}{S_i^{(1)}}\right) 1_{F_i} - 1_{F_i^c}\log\frac{n}{S_i^{(1)}}.$$

As above, we can prove

$$\left|\mathbb{E}_\vartheta\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \log\frac{n}{S_i^{(1)}} 1_{F_i^c}\right)\right| = O\left(\frac{\sqrt{N}}{n}\right), \qquad \text{for} \quad a > 4. \tag{42}$$

And:

$$\begin{aligned}\log(Z_i^{-1}) - \log\frac{n}{S_i^{(1)}} &= \log\frac{S_i^{(1)}}{n} - \log Z_i \\ &= \left(\frac{S_i^{(1)}}{n} - Z_i\right)\int_0^1 ds\left(\frac{1}{s\frac{S_i^{(1)}}{n} + (1-s)Z_i} - \frac{1}{\frac{S_i^{(1)}}{n}} + \frac{1}{\frac{S_i^{(1)}}{n}}\right) \\ &= \frac{1}{\frac{S_i^{(1)}}{n}}\left(\frac{S_i^{(1)}}{n} - Z_i\right) \\ &\quad + \frac{1}{\frac{S_i^{(1)}}{n}}\left(\frac{S_i^{(1)}}{n} - Z_i\right)^2\int_0^1 ds\frac{(1-s)}{s\frac{S_i^{(1)}}{n} + (1-s)Z_i}.\end{aligned}$$

On F_i,

$$\int_0^1 ds\frac{(1-s)}{s\frac{S_i^{(1)}}{n} + (1-s)Z_i} \le Z_i^{-1} \le \frac{\sqrt{n}}{c}.$$

Therefore,

$$\left|\log(Z_i^{-1}) - \log\frac{n}{S_i^{(1)}}\right| \le \frac{1}{\frac{C_i^{(1)}}{n}}\Psi_i^{-2}\left(\left|\frac{S_i^{(1)}}{n} - Z_i\right| + \frac{\sqrt{n}}{c}(\frac{S_i^{(1)}}{n} - Z_i)^2\right). \tag{43}$$

Now, we take conditional expectation w.r.t. $\Phi_i = \varphi, \Psi_i = \psi$, and apply first the Cauchy–Schwarz inequality and then Lemma 2. This yields:

$$\begin{aligned}&\mathbb{E}_\vartheta\left(\left|\log(Z_i^{-1}) - \log\frac{n}{S_i^{(1)}}\right| |\Phi_i = \varphi, \Psi_i = \psi\right) \\ &\lesssim \frac{1}{n}(1 + \psi^{-2}(1+\varphi^2) + \psi^2(1+\varphi^4) + \varphi^2 + \varphi^4 + \psi^6).\end{aligned} \tag{44}$$

We have to check that the expectation above is finite. The worst term is $\mathbb{E}_\vartheta\Psi_i^6 = \mathbb{E}_\vartheta\Gamma_i^{-3}$ which requires the constraint $a > 3$. Thus, for $a > 3$, we have:

$$\mathbb{E}_\vartheta\frac{1}{\sqrt{N}}\left|\sum_{i=1}^N(\log(Z_i^{-1}) - \log\frac{n}{S_i^{(1)}})1_{F_i}\right| = O\left(\frac{\sqrt{N}}{n}\right).$$

Therefore, we have proved that R_1, R_2 are $O_P(\max\{\frac{1}{\sqrt{n}}, \frac{\sqrt{N}}{n}\})$.

For R_3, R_4, we proceed analogously but we have to deal with the terms A_i, A_i^2. We write again:

$$Z_i^{-1}1_{F_i} - \Gamma_i = \left(Z_i^{-1} - \frac{n}{S_i^{(1)}}\right)1_{F_i} - \frac{n}{S_i^{(1)}}1_{F_i^c} + \frac{n}{S_i^{(1)}} - \Gamma_i.$$

Using Lemma 6 and the Cauchy–Schwarz inequality, we obtain:

$$\mathbb{E}_\vartheta\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N A_i\left(\frac{n}{S_i^{(1)}} - \Gamma_i\right)\right| \lesssim \sqrt{\frac{N}{n}}\mathbb{E}_\vartheta\left(\Gamma_i|\Phi_i| + \Gamma_i^{1/2}\right) = O\left(\sqrt{\frac{N}{n}}\right),$$

$$\mathbb{E}_\vartheta\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N A_i\frac{n}{S_i^{(1)}}1_{F_i^c}\right| \lesssim \sqrt{\frac{N}{n}}\mathbb{E}_\vartheta\left(\Gamma_i|\Phi_i| + \Gamma_i^{1/2}\right) = O\left(\sqrt{\frac{N}{n}}\right).$$

Applying now Lemma 2, we obtain:

$$\begin{aligned}&\mathbb{E}_\vartheta\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N A_i\left(Z_i^{-1} - \frac{n}{S_i^{(1)}}\right)1_{F_i}\right|\\ &\lesssim \frac{\sqrt{N}}{n}\mathbb{E}_\vartheta(|\Phi_i|(1+\Psi_i+\Psi_i^2) + \Phi_i^2\Psi_i + \Psi_i + \Psi_i^2 + \Psi_i^3).\end{aligned}$$

This requires the constraint $\mathbb{E}_\vartheta\Psi_i^3\mathbb{E}_\vartheta\Gamma_i^{-3/2} < +\infty$, *i.e.* $a > 3/2$. Note that $\mathbb{E}_\vartheta(\Psi_i^{-4p})$ is finite for any p. This is why these moments are omitted.

We proceed analogously for R_4 and find that $R_4 = O_P(\sqrt{N/n})$ for $a > 2$. The different constraint on a derives from the presence of A_i^2 in R_4 which requires higher moments for Γ_i^{-1}. □

6.3 Proof of Theorem 3

We only give a sketch of the proof and assume $d = 1$ for simplicity. We compute $\mathcal{H}_{N,n}(\vartheta)$ (see 32) and set $G_{i,n} = \{S_{i,n} \geq k\sqrt{n}\}$. We have:

$$\frac{\partial \mathbf{V}_{N,n}^{(1)}}{\partial\lambda}(\lambda, a) = \sum_{i=1}^N\left(\frac{a}{\lambda} - \xi_{i,n}^{-1}\right),$$

$$\frac{\partial \mathbf{V}_{N,n}^{(1)}}{\partial a}(\lambda, a) = N\left(\psi(a+n/2) - \log(a+n/2) - \psi(a) + \log\lambda\right) - \sum_{i=1}^N\log\xi_{i,n},$$

$$\frac{\partial \mathbf{W}_{N,n}}{\partial\mu}(\mu, \omega^2) = \sum_{i=1}^N 1_{G_{i,n}}\frac{n}{S_{i,n}}A_{i,n},$$

$$\frac{\partial \mathbf{W}_{N,n}}{\partial\omega^2}(\mu, \omega^2) = \frac{1}{2}\sum_{i=1}^N\left(1_{G_{i,n}}\frac{n}{S_{i,n}}A_{i,n}^2 - B_{i,n}\right).$$

We can prove that, under (H1)–(H2), if $a_0 > 2$, $\mathbb{P}_{\vartheta_0}(G_{i,n}^c) \lesssim n^{-2}$ under analogous and simpler tools as in Lemma 1. The result of Lemma 2 holds with $\xi_{i,n}$ instead of $Z_{i,n}$ and without $1_{F_{i,n}}$. This allows to prove that:

$$N^{-1/2}\frac{\partial \mathbf{V}_{N,n}^{(1)}}{\partial \lambda}(\lambda, a) = N^{-1/2}\sum_{i=1}^{N}\left(\frac{a}{\lambda} - \Gamma_i\right) + r_1,$$
$$N^{-1/2}\frac{\partial \mathbf{V}_{N,n}^{(1)}}{\partial a}(\lambda, a) = N^{-1/2}\sum_{i=1}^{N}(\log \Gamma_i - \psi(a) + \log \lambda) + r_2$$

where r_1 and r_2 are $O_P(\sqrt{N}/n)$.

The result of Lemma 2 holds with $S_{i,n}/n$ instead of $Z_{i,n}$ and $G_{i,n}$ instead of $F_{i,n}$ (and the proof is much simpler). This implies that:

$$N^{-1/2}\frac{\partial \mathbf{W}_{N,n}}{\partial \mu}(\mu, \omega^2) = N^{-1/2}\sum_{i=1}^{N}\Gamma_i A_i(T; \mu, \omega^2) + r_3,$$
$$N^{-1/2}\frac{\partial \mathbf{U}_{N,n}}{\partial \omega^2}(\mu, \omega^2) = N^{-1/2}\frac{1}{2}\sum_{i=1}^{N}(\Gamma_i A_i^2(T; \mu, \omega^2) - B_i(T, \omega^2)) + r_4,$$

and we can prove that r_3 and r_4 are $O_P(N/n)$.

7 Auxiliary results

In Sect. 7.1, we explain why Assumption (H2) is required. Results on discretizations are recalled in Sect. 7.2. Sections 7.3 and 7.4 contain results on Gamma distributions and estimation from direct observations of the random effects.

7.1 Preliminary results for SDEs with random effects

SDEMEs have specific features that differ from usual SDEs. The discussion below justifies our setup and assumptions. Consider $(X(t), t \geq 0)$ a stochastic process ruled by:

$$dX(t) = b(X(t), \Phi)dt + \sigma(X(t), \Psi)dW(t), \quad X(0) = x \tag{45}$$

where the Wiener process $(W(t))$ and the r.v.'s (Φ, Ψ) are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and independent. We set $(\mathcal{F}_t = \sigma(\Phi, \Psi, W(s), s \leq t), t \geq 0)$. To understand the properties of $(X(t))$, let us introduce the system of stochastic differential equations:

$$dX(t) = b(X(t), \Phi(t))dt + \sigma(X(t), \Psi(t))dW(t), \quad X(0) = x$$
$$d\Phi(t) = 0, \quad \Phi(0) = \Phi$$
$$d\Psi(t) = 0, \quad \Psi(0) = \Psi.$$

Existence and uniqueness of strong solutions is therefore ensured by the classical assumptions:

(A1) The real valued functions $(x, \varphi) \in \mathbb{R} \times \mathbb{R}^d \to b(x, \varphi)$ and $(x, \psi) \in \mathbb{R} \times \mathbb{R}^{d'} \to \sigma(x, \psi)$ are C^1.
(A2) There exists a constant K such that, $\forall x, \varphi, \psi \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^{d'}$, ($\|.\|$ is the Euclidian norm):

$$|b(x, \varphi)| \leq K(1 + |x| + \|\varphi\|), \quad |\sigma(x, \psi)| \leq K(1 + |x| + \|\psi\|)$$

If the following additional assumption holds:
(A3) The r.v.'s Φ, Ψ satisfy $\mathbb{E}(\|\Phi\|^{2p} + \|\Psi\|^{2p}) < +\infty$

then, under (A1)–(A2), for all T, $\sup_{t \leq T} \mathbb{E}X^{2p}(t) < +\infty$.

To deal with discrete observations, moment properties are required to control error terms. Here, we consider a simple model for the drift term, $b(x, \varphi) = \varphi' b(x)$, and for the diffusion coefficient, $\sigma(x, \psi) = \psi\sigma(x)$, with $\varphi \in \mathbb{R}^d$, $\psi \in \mathbb{R}$. The common assumptions on b and σ should be that b, σ are $C^1(\mathbb{R})$ and have linear growth w.r.t. x. Then, for all fixed $(\varphi, \psi) \in \mathbb{R}^d \times (0, +\infty)$, the stochastic differential equation

$$d\, X^{\varphi,\psi}(t) = \varphi' b(X^{\varphi,\psi}(t))dt + \psi\sigma(X^{\varphi,\psi}(t))\, dW(t), \quad X^{\varphi,\psi}(0) = x \tag{46}$$

admits a unique strong solution process $(X^{\varphi,\psi}(t), t \geq 0)$ adapted to the filtration $(\mathcal{F}_t)$. Moreover, the stochastic differential equation with random effects (1) admits a unique strong solution adapted to $(\mathcal{F}_t)$ such that the joint process $(\Phi, \Psi, X(t), t \geq 0)$ is strong Markov and the conditional distribution of $(X(t))$ given $\Phi = \varphi$, $\Psi = \psi$ is identical to the distribution of (46). We need more than this property. Indeed, moment properties for $X(t)$ are required. Let us stress that, with $b(x, \varphi) = \varphi' b(x)$, $\sigma(x, \psi) = \psi\sigma(x)$, (A2) does not hold even if b, σ have linear growth. In particular, (A3) does not ensure that $X(t)$ has moments of order $2p$. Let us illustrate this point on an example.

Example Consider the mixed effect Ornstein–Uhlenbeck process $dX(t) = \Phi X(t)dt + \Psi dW(t)$, $X(0) = x$. Then, $X(t) = x\exp(\Phi t) + \Psi\exp(\Phi t)\int_0^t \exp(-\Phi s)\; dW(s)$. The first order moment of $X(t)$ is finite iff $\mathbb{E}\exp(\Phi t) < +\infty$ and $\mathbb{E}\left[|\Psi|((\exp(2\Phi t) -1)/2\Phi)^{1/2}\right] < +\infty$ which is much stronger than the existence of moments for Φ, Ψ. When (Φ, Ψ) has distribution (2), $\mathbb{E}\exp(\Phi t) = \exp(\mu t)\mathbb{E}(\exp(\omega^2 t^2/\Gamma)) = +\infty$.

In a more general setup, the existence of moments is standardly proved by using the Gronwall lemma. In this case, it would lead to an even stronger condition such as $\mathbb{E}(\exp(\Phi^2 t) + \exp(\Psi^2 t)) < +\infty$.

Hence, stronger assumptions than usual are necessary. Indeed, (A2) holds for model (1) either if φ and ψ belong to a bounded set, or if b and σ are uniformly bounded.

We could think of using a localisation device to get bounded b and σ. The problem here is to deal with N sample paths $(X_i(t))$, $i = 1, \ldots, N$ which have possibly no moments. The localisation device is here complex.

7.2 Approximation results for discretizations

The following lemmas are proved in Delattre et al. (2017). In the first two lemmas, we set $X_1(t) = X(t)$, $\Phi_1 = \Phi$, $\Psi_1 = \Psi$.

Lemma 4 *Under (H1)–(H2), for $s \le t$ and $t - s \le 1$, $p \ge 1$,*

$$\mathbb{E}_\vartheta(|X(t) - X(s)|^p | \Phi = \varphi, \Psi = \psi) \lesssim K^p (t-s)^{p/2} (|\varphi|^p + \psi^p).$$

For $t \to H(t, X.)$ a predictable process, let $V(H;T) = \int_0^T H(s, X.)ds$ and $U(H;T) = \int_0^T H(s, X.)dX(s)$. The following results can be standardly proved.

Lemma 5 *Assume (H1)–(H2) and $p \ge 1$. If H is bounded, $\mathbb{E}_\vartheta(|U(H;T)|^p | \Phi = \varphi, \Psi = \psi) \lesssim |\varphi|^p + \psi^p$.*

Consider $f : \mathbb{R} \to \mathbb{R}$ and set $H(s, X.) = f(X(s))$, $H_n(s, X.) = \sum_{j=1}^n f(X((j-1)\Delta))1_{((j-1)\Delta, j\Delta]}(s)$. If f is Lipschitz,

$$\mathbb{E}_\vartheta(|V(H;T) - V(H_n;T)|^p | \Phi = \varphi, \Psi = \psi) \lesssim \Delta^{p/2}(|\varphi|^p + \psi^p).$$

If f is C^2 with f', f'' bounded

$$\mathbb{E}_\vartheta(|U(H;T) - U(H_n;T)|^p | \Phi{=}\varphi, \Psi{=}\psi) \lesssim \Delta^{p/2}(\varphi^{2p} + |\varphi|^p \psi^p + \psi^{2p} + \psi^{3p}).$$

Lemma 6 *Recall notations (25)–(26). Under (H1)–(H2),*

$$\begin{aligned}
\mathbb{E}_\vartheta(|B_{i,n} - B_i(T;\omega^2)|^p | \Phi_i = \varphi, \Psi_i = \psi) &\lesssim \Delta^{p/2}(|\varphi|^p + \psi^p) \\
\mathbb{E}_\vartheta(|A_{i,n} - A_i(T;\mu,\omega^2)|^p | \Phi_i = \varphi, \Psi_i = \psi) &\lesssim \Delta^{p/2}(\varphi^{2p} + \psi^{3p}) \\
\mathbb{E}_\vartheta(|A_i(T;\mu,\omega^2)|^p | \Phi_i = \varphi, \Psi_i = \psi)) &\lesssim |\varphi|^p + \psi^p.
\end{aligned}$$

Let

$$S_{i,n}^{(1)} = \frac{1}{\Gamma_i} \sum_{j=1}^n \frac{(W_i(t_j) - W_i(t_{j-1}))^2}{\Delta}.$$

Lemma 7 *Then, for all $p \ge 1$,*

$$\mathbb{E}_\vartheta\left(\left| \frac{S_{i,n}}{n} - \frac{S_{i,n}^{(1)}}{n} \right|^p | \Phi_i = \varphi, \Psi_i = \psi \right) \lesssim \Delta^p (\psi^{2p}\varphi^{2p} + \psi^{4p} + \varphi^{2p}).$$

7.3 Properties of the Gamma distribution

The digamma function $\psi(a) = \Gamma'(a)/\Gamma(a)$ admits the following integral representation: $\psi(z) = -\gamma + \int_0^1 (1 - t^{z-1})/(1-t)dt$ (where $-\gamma = \psi(1) = \Gamma'(1)$). For all

positive a, we have $\psi'(a) = -\int_0^1 \frac{\log t}{1-t}\, t^{a-1}dt$. Consequently, using an integration by parts, $-a\psi'(a) = -1 - \int_0^1 t^a g(t)dt$, where $g(t) = (\log t/(1-t))'$. A simple study yields that $t^a g(t)$ is integrable on $(0, 1)$ and positive except at $t = 1$. Thus, $1 - a\psi'(a) \neq 0$. The following asymptotic expansion as a tends to infinity holds:

$$\psi(a) = \log a - \frac{1}{2a} + O\left(\frac{1}{a^2}\right), \quad \psi'(a) = \frac{1}{a} + O\left(\frac{1}{a^2}\right). \tag{47}$$

If X has distribution $G(a, \lambda)$, then λX has distribution $G(a, 1)$. For all integer k, $\mathbb{E}(\lambda X)^k = \frac{\Gamma(a+k)}{\Gamma(a)}$. For $a > k$, $\mathbb{E}(\lambda X)^{-k} = \frac{\Gamma(a-k)}{\Gamma(a)}$. Moreover, $\mathbb{E}\log(\lambda X) = \psi(a)$, $\mathrm{Var}\,[\log(\lambda X)] = \psi'(a)$.

In particular, if $X = \sum_{j=1}^n \varepsilon_i^2$ where the ε_i's are *i.i.d.* $\mathcal{N}(0, 1)$, then $X \sim \chi^2(n) = G(n/2, 1/2)$. Therefore, $\mathbb{E}X^{-p} < +\infty$ for $n > 2p$ and as $n \to +\infty$,

$$\mathbb{E}\left(\frac{X}{n}\right)^p = O(1), \qquad \mathbb{E}\left(\frac{n}{X}\right)^p = O(1). \tag{48}$$

7.4 Direct observation of the random effects

Assume that a sample (Φ_i, Γ_i), $i = 1, \ldots, N$, is observed and that $d = 1$ for simplicity. The Gamma distribution with parameters (a, λ) $(a > 0, \lambda > 0)$ $G(a, \lambda)$, has density $\gamma_{a,\lambda}(x) = (\lambda^a/\Gamma(a))x^{a-1}e^{-\lambda x}\mathbb{1}_{(0,+\infty)}(x)$, where $\Gamma(a)$ is the Gamma function. We set $\psi(a) = \Gamma'(a)/\Gamma(a)$. The log-likelihood $\ell_N(\vartheta)$ of the N-sample (Φ_i, Γ_i), $i = 1, \ldots, N$ has score function $\mathcal{S}_N(\vartheta) = \left(\frac{\partial}{\partial\lambda}\ell_N(\vartheta) \;\; \frac{\partial}{\partial a}\ell_N(\vartheta) \;\; \frac{\partial}{\partial\mu}\ell_N(\vartheta) \;\; \frac{\partial}{\partial\omega^2}\ell_N(\vartheta)\right)'$ given by

$$\frac{\partial}{\partial\lambda}\ell_N(\vartheta) = \sum_{i=1}^N \left(\frac{a}{\lambda} - \Gamma_i\right), \quad \frac{\partial}{\partial a}\ell_N(\vartheta) = \sum_{i=1}^N \left(-\psi(a) + \log\lambda + \log\Gamma_i\right),$$
$$\frac{\partial}{\partial\mu}\ell_N(\vartheta) = \omega^{-2}\sum_{i=1}^N \Gamma_i(\Phi_i - \mu), \quad \frac{\partial}{\partial\omega^2}\ell_N(\vartheta) = \frac{1}{2\omega^4}\sum_{i=1}^N \left(\Gamma_i(\Phi_i - \mu)^2 - \omega^2\right).$$

By standard properties, we have, under $\mathbb{P}_\vartheta$, $N^{-1/2}\mathcal{S}_N(\vartheta) \to_{\mathcal{D}} \mathcal{N}_4(0, \mathcal{J}_0(\vartheta))$, where

$$\mathcal{J}_0(\vartheta) = \left(\begin{array}{c|c} I_0(\lambda, a) & \mathbf{0} \\ \hline \mathbf{0} & J_0(\lambda, a, \mu, \omega^2) \end{array}\right),$$
$$I_0(\lambda, a) = \begin{pmatrix} \frac{a}{\lambda^2} & -\frac{1}{\lambda} \\ -\frac{1}{\lambda} & \psi'(a) \end{pmatrix}, \quad J_0(\lambda, a, \mu, \omega^2) = \begin{pmatrix} \frac{a}{\lambda\omega^2} & 0 \\ 0 & \frac{1}{2\omega^4} \end{pmatrix}. \tag{49}$$

Using properties of the di-gamma function ($a\psi'(a) - 1 \neq 0$), $I_0(a, \lambda)$ is invertible for all $(a, \lambda) \in (0, +\infty)^2$. The maximum likelihood estimator based on the observation of $(\Phi_i, \Gamma_i, i = 1, \ldots, N)$, denoted $\vartheta_N = \vartheta_N(\Phi_i, \Gamma_i, i = 1, \ldots, N)$, is consistent and satisfies $\sqrt{N}(\vartheta_N - \vartheta) \to_{\mathcal{D}} \mathcal{N}_4(0, \mathcal{J}_0^{-1}(\vartheta))$ under $\mathbb{P}_\vartheta$ as N tends to infinity.

In the simulations presented in Sect. 4, we took $a = 8$ and observed that estimations of a are biased with a large standard deviation. This can be seen on $I_0^{-1}(\lambda, a)$:

$$I_0^{-1}(\lambda, a) = \frac{1}{a\psi'(a) - 1} \begin{pmatrix} \lambda^2\psi'(a) & \lambda\psi'(a) \\ \lambda\psi'(a) & a \end{pmatrix}. \tag{50}$$

If a large, $a/(a\psi'(a) - 1) = O(a^2)$.

However, natural parameters for Gamma distributions are $m = a/\lambda$, $t = \psi(a) - \log\lambda$ with unbiased estimators $\hat{m} = N^{-1}\sum_{i=1}^N \Gamma_i$, $\hat{t} = N^{-1}\sum_{i=1}^N \log\Gamma_i$ such that the vector $\sqrt{N}(\hat{m} - m, \hat{t} - t)$ is asymptotically Gaussian with limiting covariance matrix

$$\begin{pmatrix} \frac{a}{\lambda^2} & \frac{1}{\lambda} \\ \frac{1}{\lambda} & \psi'(a) \end{pmatrix}. \tag{51}$$

The asymptotic variance of $\hat{t}$ is $\psi'(a) = O(a^{-1})$ and both parameters (m, t) are well estimated.

References

Berglund M, Sunnaker M, Adiels M, Jirstrand M, Wennberg B (2001) Investigations of a compartmental model for leucine kinetics using non-linear mixed effects models with ordinary and stochastic differential equations. Math Med Biol. https://doi.org/10.1093/imammb/dqr021

Delattre M, Dion C (2017) MsdeParEst: Parametric estimation in mixed-effects stochastic differential equations. https://CRAN.R-project.org/package=MsdeParEst. Accessed 16 Sept 2017

Delattre M, Genon-Catalot V, Larédo C (2017) Parametric inference for discrete observations of diffusion with mixed effects in the drift or in the diffusion coefficient. Stoch Process Appl. https://doi.org/10.1016/j.spa.2017.08.016

Delattre M, Genon-Catalot V, Samson A (2013) Maximum likelihood estimation for stochastic differential equations with random effects. Scand J Stat 40:322–343

Delattre M, Genon-Catalot V, Samson A (2015) Estimation of population parameters in stochastic differential equations with random effects in the diffusion coefficient. ESAIM Probab Stat 19:671–688

Delattre M, Lavielle M (2013) Coupling the SAEM algorithm and the extended Kalman filter for maximum likelihood estimation in mixed-effects diffusion models. Stat Interface 6(4):519–532

Dion C (2016) Nonparametric estimation in a mixed-effect Ornstein–Uhlenbeck model. Metrika 79:919–951

Ditlevsen S, De Gaetano A (2005) Mixed effects in stochastic differential equation models. REVSTAT Stat J 3:137–153

Donnet S, Samson A (2008) Parametric inference for mixed models defined by stochastic differential equations. ESAIM P&S 12:196–218

Forman J, Picchini U (2016) Stochastic differential equation mixed effects models for tumor growth and response to treatment. Preprint arXiv:1607.02633

Grosse Ruse M, Samson A, Ditlevsen S (2017) Multivariate inhomogeneous diffusion models with covariates and mixed effects. Preprint arXiv:1701.08284

Jacod J, Protter P (2012) Discretization of processes, vol 67. Stochastic modelling and applied probability. Springer, Berlin

Kessler M, Lindner A, Sørensen ME (2012) Statistical methods for stochastic differential equations, vol 124. Monographs on statistics and applied probability. Chapman & Hall, London

Leander J, Almquist J, Ahlstrom C, Gabrielsson J, Jirstrand M (2015) Mixed effects modeling using stochastic differential equations: Illustrated by pharmacokinetic data of nicotinic acid in obese zucker rats. AAPS J. https://doi.org/10.1208/s12248-015-9718-8

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. J Stat Plan Inference 137:1787–1804

Nie L, Yang M (2005) Strong consistency of the MLE in nonlinear mixed-effects models with large cluster size. Sankhya Indian J Stat 67:736–763

Picchini U, De Gaetano A, Ditlevsen S (2010) Stochastic differential mixed-effects models. Scand J Stat 37:67–90

Picchini U, Ditlevsen S (2011) Practicle estimation of high dimensional stochastic differential mixed-effects models. Comput Stat Data Anal 55:1426–1444

Whitaker GA, Golightly A, Boys RJ, Chris S (2017) Bayesian inference for diffusion driven mixed-effects models. Bayesian Anal 12:435–463