Estimation of the joint distribution of random effects for a discretely observed diffusion with random effects

Maud Delattre, Valentine Genon-Catalot, Catherine Larédo

Estimation of the joint distribution of random effects for a discretely observed diffusion with random effects.

Maud Delattre[1], Valentine Genon-Catalot[2] and Catherine Larédo[3*]

[1] AgroParisTech, France

[2] UMR CNRS 8145, Laboratoire MAP5, Université Paris Descartes, Sorbonne Paris Cité, France

[3] Laboratoire MaIAGE, I.N.R.A. and LPMA, Université Denis Diderot, UMR CNRS 7599.

January 25, 2017

Abstract

Mixed effects models are popular tools for analyzing longitudinal data from several individuals simultaneously. Individuals are described by N independent stochastic processes $(X_i(t), t \in [0,T])$, $i = 1, \ldots, N$, defined by a stochastic differential equation with random effects. We assume that the drift term depends linearly on a random vector Φ_i and the diffusion coefficient depends on another linear random effect Ψ_i. For the random effects, we consider a joint parametric distribution leading to explicit approximate likelihood functions for discrete observations of the processes X_i on a fixed time interval. The asymptotic behaviour of the associated estimators is studied when both the number of individuals and the number of observations per individual tend to infinity. The estimation methods are investigated on simulated and real neuronal data.

Key Words: asymptotic properties, discrete observations, estimating equations, parametric inference, random effects models, stochastic differential equations.

1 Introduction

Mixed effects models are popular tools for analyzing longitudinal data from several individuals simultaneously. In these models, the individuals are described by the same structural model and the inclusion of some random effects enables the description of the inter-individual variability which is inherent to the data (see Pinheiro and Bates (2000) for instance). The qualification "mixed-effects" comes from the fact that these models often include two kinds of parameters: the fixed effects that are common to all individuals and the random effects that vary from one individual to another according to a given probability distribution. Such models are especially suited in the biomedical field where typical experiments often consist of repeated measurements on several patients or experimental animals. 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 as well as in SDEMEs is the estimation

*Corresponding author

Email addresses: maud.delattre@agroparistech.fr (Maud Delattre), valentine.genon-catalot@parisdescartes.fr (Valentine Genon-Catalot), catherine.laredo@jouy.inra.fr (Catherine Larédo)

of the parameters of the distribution of the random effects. It is often difficult in practice due to the untractable likelihood function. In the SDEMEs framework, many methods have now been proposed. In Ditlevsen and De Gaetano (2005), the special case of a mixed-effects Brownian motion with drift is treated. For general models, methods are based on various numerical approximations of the likelihood (Picchini et al. (2010), Picchini and Ditlevsen (2011), Delattre and Lavielle (2013)). For what concerns the exact likelihood approach or explicit estimating equations, the main contributions to our knowledge are from Delattre et al. (2013), Delattre et al. (2015) and Delattre et al. (2016). In this paper, we extend the framework of these three papers, considering discrete observations adding simultaneous random effects in the drift and diffusion coefficients by means of a joint parametric distribution. The problem of estimating simultaneously the distributions of random effects in the drift and in the diffusion coeffcient has never been studied theoretically to our knowledge.
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):

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

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. The processes $(X_i(t))$ are supposed to be discretely observed on the fixed-length time interval $[0, T]$ at n times $t_j = jT/n$. With this parametric distribution for the random effects Ψ_i, an explicit approximate likelihood is built and the associated estimators are studied assuming that both the number N of individuals and the number n of observations per individual tend to infinity. In Delattre et al. (2016), the previous papers are extended and improved by considering for model (1) the cases $\Phi_i \sim \mathcal{N}_d(\boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Omega})$, $\Psi_i = \gamma^{-1/2}$ non random but unknown or $\Phi_i = \varphi$ non random and unknown and $\Psi_i = \Gamma_i^{-1/2}$ with Γ_i following a Gamma distribution. The joint estimation of respectively $(\gamma, \boldsymbol{\mu}, \boldsymbol{\Omega})$ and (λ, a, φ) is studied.
Our aim here is to study the case where both Φ_i and Ψ_i are random and have a joint parametric distribution. We assume that each process $(X_i(t))$ is discretely observed on a fixed-length time interval $[0, T]$ at n times $t_j = jT/n$. We consider a joint parametric distribution for the random effects (Φ_i, Ψ_i) and 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. Joining the two cases of Delattre

et al. (2016) in a single one, we assume that (Φ_i, Ψ_i) has the following distribution:

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

Thus, Φ_i and Ψ_i are dependent and the unknown parameter is $\vartheta = (\lambda, a, \boldsymbol{\mu}, \boldsymbol{\Omega})$. Let us stress that, 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. We obtain consistent and $\sqrt{N}$-asymptotically Gaussian estimators for all parameters under the condition $N/n \to 0$. We compare the results with estimation for direct observation of the random effects; we also compare the results with the analogous ones based on continuous observation of the diffusions sample path.
The structure of the paper is the following. Some working assumptions and two approximations of the model likelihood are introduced in Section 2. Two estimation methods are derived from these approximate likelihoods in Section 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. Section 5 concerns the implementation on a real neuronal dataset. Some concluding remarks are given in Section 6. Theoretical proofs are gathered in the Appendix.

2 Approximate Likelihood.

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 $(\Phi_i, \Psi_i), i = 1, \ldots, N$ are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and we set $(\mathcal{F}_t = \sigma(\Phi_i, \Psi_i, W_i(s), s \leq t, i = 1, \ldots, N), t \geq 0)$. The set of assumptions is the same as in Delattre et al. (2016). 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 $\sigma(.)$ is lower bounded : $\exists\, \sigma_0 > 0, \forall x \in \mathbb{R}, \sigma(x) \geq \sigma_0$.

(H2) There exists a constant K such that, $\forall x \in \mathbb{R}$, $\|b(x)\| + |\sigma(x)| \leq K$.
($\|.\|$ denotes the Euclidian norm of $\mathbb{R}^d$.)

Assumption (H1) ensures existence and uniqueness of solutions of (1) (see *e.g.* Delattre et al. (2013)). Under (H1), for $i = 1, \ldots, N$, for all deterministic $(\varphi, \psi) \in \mathbb{R}^d \times (0, +\infty)$, the stochastic differential equation

$$d\, X_i^{\varphi,\psi}(t) = \varphi' b(X_i^{\varphi,\psi}(t))dt + \psi\sigma(X_i^{\varphi,\psi}(t))\, dW_i(t), \quad X_i^{\varphi,\psi}(0) = x \tag{3}$$

admits a unique strong solution process $(X_i^{\varphi,\psi}(t), t \geq 0)$ 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 $(\Phi_i, \Psi_i, X_i(t), t \geq 0)$ is strong Markov and the conditional distribution of $(X_i(t))$ given $\Phi_i = \varphi, \Psi_i = \psi$ is identical to the distribution of (3). The processes $(\Phi_i, \Psi_i, (X_i(t), t \geq 0)), i = 1, \ldots, N$ are *i.i.d.*. Assumption (H2) simplifies proofs.
The processes $(X_i(t))$ are discretely observed with sampling interval Δ_n on a fixed time interval $[0, T]$

and we set:

$$\Delta_n = \Delta = \frac{T}{n},\ X_{i,n} = X_i = (X_i(t_{j,n}),\ \ t_{j,n} = t_j = jT/n,\ \ j = 1,\ldots,n). \tag{4}$$

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.
The 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 integrate 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 (3). Setting $\psi = \gamma^{-1/2}$, it is given by:

$$L_n(X_{i,n},\gamma,\varphi) = L_n(X_i,\gamma,\varphi) = \gamma^{n/2}\exp\left[-\frac{\gamma}{2}(S_{i,n} + \varphi' V_{i,n}\varphi - 2\varphi' U_{i,n})\right], \tag{5}$$

where

$$S_{i,n} = S_i = \frac{1}{\Delta}\sum_{j=1}^n \frac{(X_i(t_j) - X_i(t_{j-1}))^2}{\sigma^2(X_i(t_{j-1}))}, \tag{6}$$

$$V_{i,n} = V_i = \left(\sum_{j=1}^n \Delta \frac{b_k(X_i(t_{j-1}))b_\ell(X_i(t_{j-1}))}{\sigma^2(X_i(t_{j-1}))}\right)_{1\le k,\ell\le d}, \tag{7}$$

$$U_{i,n} = U_i = \left(\sum_{j=1}^n \frac{b_k(X_i(t_{j-1}))(X_i(t_j) - X_i(t_{j-1}))}{\sigma^2(X_i(t_{j-1}))}\right)_{1\le k\le d}. \tag{8}$$

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

$$\mathcal{L}_{N,n}(\vartheta) = \prod_{i=1}^N \mathcal{L}_n(X_{i,n},\vartheta), \qquad \mathcal{L}_n(X_{i,n},\vartheta) = \int L_n(X_{i,n},\gamma,\varphi)\nu_\vartheta(d\gamma, d\varphi).$$

In general, this does not lead to a closed-form formula. The joint distribution (2) proposed here allows to obtain an explicit formula for the above integral. Some additional assumptions are required.

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

$$V_i(T) = \left(\int_0^T \frac{b_k(X_i(s))b_\ell(X_i(s))}{\sigma^2(X_i(s))}ds\right)_{1\le k,\ell\le d}. \tag{9}$$

(H4) The parameter set satisfies, for constants $m, c_0, c_1, \ell_0, \ell_1, \alpha_0, \alpha_1$,
$\|\boldsymbol{\mu}\| \le m,\ \ \lambda_{max}(\boldsymbol{\Omega}) \le c_1,\ 0 < \ell_0 \le \lambda \le \ell_1,\ 0 < \alpha_0 \le a \le \alpha_1,$
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 (7) 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 $\mathbf{\Omega}$ may be non invertible which allows including non random effects in the drift term.

2.1 First approximation of the likelihood

Let us now compute the approximate likelihood $\mathcal{L}_n(X_{i,n}, \vartheta)$ of $X_{i,n}$ integrating $L_n(X_{i,n}, \gamma, \varphi)$ with respect to the distribution (2). For this, we first integrate $L_n(X_i, \gamma, \varphi)$ with respect to the Gaussian distribution $\mathcal{N}_d(\boldsymbol{\mu}, \gamma^{-1}\mathbf{\Omega})$. 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 a subset $E_{i,n}(\vartheta)$ precised now. Let I_d denote the identity matrix of $\mathbb{R}^d$ and set, for $i = 1, \ldots, N$,

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

which is well defined and invertible under (H3). Define

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

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

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

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

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

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 $T_{i,n}(\boldsymbol{\mu}, \mathbf{\Omega})/n$ 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),

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

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,*

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

Consequently, 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.

2.2 Second approximation of the likelihood

Formulae (13)-(14) suggest another approximation of the likelihood which is simpler and related to the likelihood functions obtained in Delattre et al. (2013) and Delattre et al. (2015). We give this approxmation without being rigorous. The formula is justified *a posteriori*. We can write:

$$\mathcal{L}_n(X_{i,n}, \vartheta) = \mathcal{L}_n^{(1)}(X_{i,n}, \vartheta) \times \mathcal{L}_n^{(2)}(X_{i,n}, \vartheta) \tag{17}$$

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

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:

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

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:

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

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 unknown parameters, we can forget it and set:

$$\boldsymbol{V}_n(X_{i,n}, \vartheta) = \boldsymbol{V}_n^{(1)}(X_{i,n}, \lambda, a) + \boldsymbol{V}_n^{(2)}(X_{i,n}, \boldsymbol{\mu}, \boldsymbol{\Omega}) \tag{18}$$

with

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

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.

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. A natural question arises which concerns the comparison with the estimation when direct observation of an *i.i.d.* sample $(\Phi_i,\Gamma_i), i=1,\dots,N$ is available (details are given in Section 8.3).

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

$$Z_{i,n}(a,\lambda,\boldsymbol{\mu},\boldsymbol{\Omega}) = Z_i = \frac{\lambda+(S_{i,n}/2)+(T_{i,n}(\boldsymbol{\mu},\boldsymbol{\Omega})/2)}{a+n/2}. \tag{19}$$

By (H4), using notations (16), we have $\lambda+S_{i,n}/2+T_{i,n}(\boldsymbol{\mu},\boldsymbol{\Omega})/2 \geq (S_{i,n}-M_{i,n})/2$. 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

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

Define the gradient vector

$$\mathcal{G}_{N,n}(\vartheta) = \nabla_\vartheta \mathbf{U}_{N,n}(\vartheta) \tag{20}$$

We consider the estimators defined by the estimating equation:

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

To investigate their asymptotic behaviour, we need to prove that $Z_{i,n}$ defined in (19) 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 $1_{F_{i,n}}Z_{i,n}^{-1}-\Gamma_i$ intoducing $n/S_{i,n}^{(1)}$.
From now on, 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 of $d>1$ does not present any additional difficulty. Let $\vartheta=(\lambda,a,\mu,\omega^2)$.

Lemma 2. *For all $p\geq 1$, we have*

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

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

Using (7),(8), we set

$$A_{i,n}=\frac{U_{i,n}-\mu V_{i,n}}{1+\omega^2V_{i,n}}\ \text{ and }\ B_{i,n}=\frac{V_{i,n}}{1+\omega^2V_{i,n}}. \tag{23}$$

Let $\psi(u)=\frac{\Gamma'(u)}{\Gamma(u)}$ denote the di-gamma function and F^c denote the complementary set of F. We obtain (see (19)):

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

Using (9),(15), we set

$$A_i(T;\mu,\omega^2)=\frac{U_i(T)-\mu V_i(T)}{1+\omega^2V_i(T)}\ \text{ and }\ B_i(T;\omega^2)=\frac{V_i(T)}{1+\omega^2V_i(T)} \tag{24}$$

and

$$J(\vartheta)=\begin{pmatrix}\mathbb{E}_\vartheta\Gamma_1B_1(T;\omega^2) & \mathbb{E}_\vartheta\Gamma_1A_1(T;\mu,\omega^2)B_1(T;\omega^2)\\ \mathbb{E}_\vartheta\Gamma_1A_1(T;\mu,\omega^2)B_1(T;\omega^2) & \mathbb{E}_\vartheta\left(\Gamma_1A_1^2(T;\mu,\omega^2)B_1(T;\omega^2)-\frac{1}{2}B_1^2(T;\omega^2)\right)\end{pmatrix}. \tag{25}$$

Denote

$$I_0(\lambda,a)=\begin{pmatrix}\frac{a}{\lambda^2} & -\frac{1}{\lambda}\\ -\frac{1}{\lambda} & \psi'(a)\end{pmatrix} \tag{26}$$

the Fisher information matrix associated with the direct observation $(\Gamma_1,\ldots,\Gamma_N)$ (see Section 8.3). 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 $N/n \to 0$, $N^{-1/2}\left(\frac{\partial \boldsymbol{U}_{N,n}}{\partial \vartheta_i}(\vartheta), i = 1, \ldots, 4\right)'$ converges in distribution under $\mathbb{P}_\vartheta$ to $\mathcal{N}_4(0, \mathcal{J}(\vartheta))$ where*

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

and the matrix $J(\vartheta)$ defined in (25) is the covariance matrix of the vector

$$\begin{pmatrix} \Gamma_1 A_1(T; \mu, \omega^2) \\ \frac{1}{2}(\Gamma_1 A_1^2(T, \mu, \omega^2) - B_1(T; \omega^2)) \end{pmatrix}. \tag{27}$$

- *Define the opposite of the Hessian of $\boldsymbol{U}_{N,n}(\vartheta)$ by $\mathcal{J}_{N,n}(\vartheta) = (-\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} \boldsymbol{U}_{N,n}(\vartheta))_{i,j=1,\ldots,4}$. Then, if $a > 6$, as n and $N = N(n)$ tend to infinity, $N^{-1}\mathcal{J}_{N,n}(\vartheta)$ converges in $\mathbb{P}_\vartheta$-probability to $\mathcal{J}(\vartheta)$.*

For the proof, we use that $A_{i,n}$ (resp. $B_{i,n}$) converges to $A_i(T; \mu, \omega^2)$ (resp. $B_i(T; \omega^2)$) 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 Section 8.3). 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 $\mathcal{N}_4(0, \mathcal{J}^{-1}(\vartheta_0))$ 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.
Let us stress that 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 the parameters coming from the random effects in the diffusion coefficient. For the other parameters (μ, ω^2), which come from the random effects in the drift, the constraint is $N/n \to 0$ and there is a loss of information (see (67)) w.r.t. the direct observation of $((\Phi_i, \Gamma_i), i = 1, \ldots, N)$. For instance, if ω^2 is known, one can see that

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

On the other hand, if for all i, $\Gamma_i = \gamma$ is deterministic , there is no loss of information, for the estimation of (μ, ω^2), w.r.t. 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 (18). We set

$$\xi_{i,n} = \frac{\lambda + (S_{i,n}/2)}{a + n/2}, \tag{28}$$

so that

$$\mathbf{V}_n^{(1)}(X_{i,n},\lambda,a) = a\log\lambda - \log\Gamma(a) + \log\Gamma(a+(n/2)) - (a+(n/2))\log(a+(n/2)) - (a+(n/2))\log\xi_{i,n}.$$

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

$$\mathbf{W}_n(X_{i,n},\boldsymbol{\mu},\boldsymbol{\Omega}) = -\frac{n}{2\,S_{i,n}}1_{S_{i,n}\geq k\sqrt{n}}(\boldsymbol{\mu} - V_{i,n}^{-1}U_{i,n})'R_{i,n}^{-1}(\boldsymbol{\mu} - V_{i,n}^{-1}U_{i,n}) - \frac{1}{2}\log\left(det(I_d + V_{i,n}\boldsymbol{\Omega})\right), \quad (29)$$

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

$$\mathcal{H}_{N,n}(\vartheta) = (\nabla_{\lambda,a}\mathbf{V}_{N,n}^{(1)}(\lambda,a), \nabla_{\boldsymbol{\mu},\boldsymbol{\Omega}}\mathbf{W}_{N,n}(\boldsymbol{\mu},\boldsymbol{\Omega})). \quad (30)$$

We consider the estimators defined by the estimating equation:

$$\mathcal{H}_{N,n}(\vartheta^*_{N,n}) = 0. \quad (31)$$

The estimators $(\lambda^*_{N,n}, a^*_{N,n})$ were studied in Delattre et al. (2015) in the case $b(.) = 0$. We see below that the non null drift term does not modify their asymptotic behavior.
If we replace $U_{i,n}, V_{i,n}$ in (29) by $U_i(T), V_i(T)$ (see (15) and (9)) and $\frac{n}{S_{i,n}}1_{S_{i,n}\geq k\sqrt{n}}$ by a constant value γ, the formula for $\mathbf{W}_{N,n}$ becomes the exact log-likelihood studied in Delattre et al. (2013) for γ known. Thus, this second approximation of the loglikelihood clarifies (14).
We have the following result.

Proposition 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 $\vartheta^*_{N,n}$ to (31) exists which is consistent and such that $\sqrt{N}(\vartheta^*_{N,n} - \vartheta_0)$ converges in distribution to $\mathcal{N}_4(0, \mathcal{J}^{-1}(\vartheta_0))$ under $\mathbb{P}_{\vartheta_0}$ for all ϑ_0 (see (25), (26) 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.

4 Simulation study

4.1 Models

Several models are simulated. We investigate both the case of singular (models 2-4) and non singular (models 1-3) conditional covariance matrix for the random effects Φ_i in the drift. Only Model (1) satisfies assumptions (H1)-(H2). However, the other models are classically used and the estimation results show that our methods perform well even when (H1)-(H2) are not satified.

Model 1. Mixed effects Brownian motion:

$$dX_i(t) = \Phi_i dt + \Gamma_i^{-1/2}dW_i(t), X_i(0) = 0\ ,\ \Gamma_i \underset{i.i.d}{\sim} G(a,\lambda)\ ,\ \Phi_i|\Gamma_i = \gamma \sim \mathcal{N}(\mu, \gamma^{-1}\omega^2) \quad (32)$$

Model 2. Mixed Ornstein-Uhlenbeck process :

$$dX_i(t) = (\rho - \Phi_i X_i(t))dt + \Gamma_i^{-1/2} dW_i(t), X_i(0) = 0\ ,\ \Gamma_i \underset{i.i.d}{\sim} G(a, \lambda)\ ,\ \Phi_i|\Gamma_i = \gamma \sim \mathcal{N}(\mu, \gamma^{-1}\omega^2) \quad (33)$$

Model 3. Mixed Ornstein-Uhlenbeck process (continued):

$$\begin{aligned} dX_i(t) &= (\Phi_{i,1} X_i(t) + \Phi_{i,2})dt + \Gamma_i^{-1/2} dW_i(t), X_i(0) = 0\ ,\ \Gamma_i \underset{i.i.d}{\sim} G(a, \lambda)\ ,\ \Phi_{i,1}|\Gamma_i = \gamma \sim \mathcal{N}(\mu_1, \gamma^{-1}\omega_1^2) \\ &\Phi_{i,2}|\Gamma_i = \gamma \sim \mathcal{N}(\mu_2, \gamma^{-1}\omega_2^2) \end{aligned} \quad (34)$$

with $\Phi_{i,1}, \Phi_{i,2}$ conditionnally independent.

Model 4. Mixed C.I.R. process :

$$dX_i(t) = (\rho - \Phi_i X_i(t))dt + \Gamma_i^{-1/2} \sqrt{|X_i(t)|} dW_i(t), X_i(0) = x > 0\ . \quad (35)$$

Some justification is required for this process. Let $X_i^{\varphi,\gamma}$ denote the process corresponding to deterministic values of $\Phi_i = \varphi, \Gamma_i = \gamma$:

$$dX_i^{\varphi,\gamma}(t) = (\rho - \varphi X_i^{\varphi,\gamma}(t))dt + \gamma^{-1/2} \sqrt{|X_i^{\varphi,\gamma}(t)|} dW_i(t), X_i(0) = x > 0. \quad (36)$$

Statistical inference for the parameters of this process has been extensively studied (see *e.g.* Overbeck (1998), Ben Alaya and Kebaier (2012) and Ben Alaya and Kebaier (2013)). For $\rho > 0$ and $\varphi \in \mathbb{R}$, the stochastic differential equation (36) admits a unique strong solution such that $X_i^{\varphi,\gamma}(t) \geq 0$ for all $t \geq 0$. Let $\tau_0 = \inf\{t \geq 0, X_i^{\varphi,\gamma}(t) = 0\}$ denote the first hitting time of the boundary 0. If $\rho \geq 1/(2\gamma)$, $\tau_0 = +\infty$ *a.s.* and the process is strictly positive. Otherwise, the process can hit the value 0 and in this case, the boundary 0 is instantaneously reflecting. In case $0 < \rho < 1/(2\gamma)$ and $\varphi \geq 0$, $\tau_0 < +\infty$ *a.s.*. In case $0 < \rho < 1/(2\gamma)$ and $\varphi < 0$, $\mathbb{P}(\tau_0 < +\infty) \in (0, 1)$. To consider mixed effects in C.I.R. processes within our framework (see (2)), we assume that ρ is deterministic and positive and that (Φ_i, Γ_i) are such that $\Gamma_i \underset{i.i.d}{\sim} G(a, \lambda)$, $\Phi_i|\Gamma_i = \gamma \sim \mathcal{N}(\mu, \gamma^{-1}\omega^2)$. It is an example of singular $\mathbf{\Omega}$. The process (35) is thus well defined and nonnegative. As $\mathbb{P}(\rho < 1/(2\Gamma_i)) = \mathbb{P}(\Gamma_i > 1/(2\rho)) < 1$, the sample path $X_i(t)$ will hit 0 in finite time with positive probability with an instantaneous reflection.

4.2 Simulation design and results

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 path are simulated with a very small discretization step-size $\delta = 0.001$. Exact simulation is performed for Models 1, 2 and 3 whereas the approximate discretization scheme described in Alfonsi (2005) is used for Model 4 and yields positive values. The time interval between consecutive observations is taken equal to $\Delta = 0.01$ or 0.05 with a resulting time interval $[0, n\Delta]$. The model parameters are then estimated by Method 1 (section 3.1) and Method 2 (section 3.2) from each simulated dataset. The empirical mean and standard deviation of the estimates are computed from the 100 datasets (Tables 3 to 6) and are compared with those of the estimates based on a direct observation of the random effects (Tables 1 and 2). As the parameters of Gamma distributions are uneasy to estimate in practice, we use an other parametrization in terms of $m = a/\lambda$ and $t = \psi(a) - \log(\lambda)$, and we present the estimates of m and t rather than the estimates of a and λ.

Both estimation procedures require some truncations. Let us make some remarks about the truncation used in the first estimation method. Although it performs well in theory, the truncation defined by means of the sets $F_{i,n}$ is not adequate for a practical use. We have indeed observed that, even for very small values of α, a large number of simulated trajectories do not belong to the sets $F_{i,n}$ (see (16)), leading to poor estimation performances. For implementing the first method, we have thus replaced $1_{F_{i,n}}$ by $1_{(Z_{i,n} \geq c/\sqrt{n})}$ in the definition of $\mathbf{U}_n(X_{i,n}, \vartheta)$ and then numerically optimized the corresponding new contrast. Here, we choose $c = 0.1$. The interest of introducing the sets $F_{i,n}$ was that they do not depend on the parameters and this simplifies differentiation. However, the main theoretical proof only requires the condition $Z_{i,n} \geq c/\sqrt{n})$ (see Lemma 2). The present simulation study shows that the method performs well in practice. Concerning the second estimation method, we use $k = 0.1$ (see (29)).

We observe from Tables 3 to 6 that the two estimation methods have similar performances overall. Both estimate the parameters with very little bias whatever the model and the values of n and N. The standard deviations of the estimates are very close for the two methods as well, except in Model 4 where the estimates are more variable with method 1 than with method 2. We guess that the numerical optimisation of the first contrast is more difficult than the one of the second contrast which decouples the estimation of the parameters of the conditional Gaussian distribution from the estimation of the parameters of the Gamma distribution. We can remark that the estimates for m and t obtained from the observations of the SDEs have similar standard deviations as those obtained from a direct observation of the random effects, whereas they are higher for parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ (except for Model 1 where they are equal). This is expected from the theory (see remark following Theorem 2). The bias and the standard deviations of the estimates generally decrease as N increases. When N is fixed, increasing the number n of observations per trajectory decreases bias but does not have any impact on the estimates standard deviations. This is in accordance with the theory since the asymptotic distribution of the estimates is obtained when $N/n \to 0$ or $N/n^2 \to 0$ according to the parameters, and the rate of convergence is $\sqrt{N}$. Although $n = 200$ does not fulfill $N/n \to 0$ in our simulation design, the results in this case are very satisfactory which encourages the use of these two estimation methods not only for high but also moderate values of n. Let us stress that assumptions (H1)-(H2) are not satified by Models 2, 3 and 4. This does not deteriorate the performances of the estimation methods. Note also that although the matrix $\boldsymbol{\Omega}$ is singular in Models 2 and 4, the model parameters are well estimated.

Direct observations of N i.i.d. (Φ_i, Γ_i) s.t. $\Gamma_i \sim G(a,\lambda)$, $\Phi_i \vert \Gamma_i = \gamma \sim \mathcal{N}(\mu, \gamma^{-1}\omega^2)$							
		$N = 50$			$N = 100$		
Parameter	True value	Mean	s.d.	Theor. s.d.	Mean	s.d.	Theor. s.d.
μ	-0.50	-0.51	0.06	0.05	-0.50	0.04	0.04
ω^2	0.50	0.50	0.09	0.10	0.49	0.08	0.07
m	4.00	3.99	0.23	0.20	3.99	0.14	0.14
t	1.32	1.32	0.06	0.05	1.32	0.04	0.04
μ	-0.50	-0.50	0.02	0.02	-0.50	0.02	0.02
ω^2	0.10	0.10	0.02	0.02	0.10	0.01	0.01
m	4.00	4.00	0.21	0.20	4.02	0.14	0.14
t	1.32	1.32	0.05	0.05	1.33	0.03	0.04
μ	-1.00	-1.00	0.02	0.02	-1.00	0.02	0.02
ω^2	0.10	0.10	0.02	0.02	0.10	0.02	0.01
m	4.00	3.98	0.19	0.20	4.01	0.14	0.14
t	1.32	1.32	0.05	0.05	1.33	0.04	0.04

Table 1: Empirical mean and standard deviation from 100 datasets; theoretical s.d. of estimators for different values of N, μ and ω^2, $a = 8$, $\lambda = 2$, $m = a/\lambda$, $t = \psi(a) - \log \lambda$.

Direct observations of N i.i.d. (Φ_i, Γ_i) s.t. $\Gamma_i \sim G(a,\lambda)$, $\Phi_i \vert \Gamma_i = \gamma \sim \mathcal{N}(\mu, \gamma^{-1}\Omega)$ with $\mu = (\mu_1, \mu_2)'$ and $\Omega = \mathrm{diag}(\omega_1^2, \omega_2^2)$							
		$N = 50$			$N = 100$		
Parameter	True value	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

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

Model 1: $dX_i(t) = \Phi_i dt + \Gamma_i^{-1/2} dW_i(t), X_i(0) = 0$ $\Gamma_i \underset{i.i.d}{\sim} G(a,\lambda)$, $\Phi_i \vert \Gamma_i = \gamma \sim \mathcal{N}(\mu, \gamma^{-1}\omega^2)$					
Parameter	True value	$N = 50$		$N = 100$	
		$\Delta = 0.025$ ($n = 200$)	$\Delta = 0.005$ ($n = 1000$)	$\Delta = 0.025$ ($n = 200$)	$\Delta = 0.005$ ($n = 1000$)
Method 1					
μ	-0.50	-0.51 (0.06)	-0.51 (0.06)	-0.50 (0.05)	-0.50 (0.05)
ω^2	0.50	0.49 (0.12)	0.50 (0.13)	0.48 (0.10)	0.48 (0.10)
m	4.00	4.00 (0.24)	3.99 (0.23)	4.02 (0.15)	4.00 (0.15)
t	1.32	1.28 (0.06)	1.31 (0.06)	1.28 (0.04)	1.31 (0.04)
Method 2					
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 3: Empirical mean and (empirical s.d.) of estimators for different values of N and n, truncations $c = k = 0.1$, $m = a/\lambda$, $t = \psi(a) - \log \lambda$.

Model 2: $dX_i(t) = (\rho - \Phi_i X_i(t))dt + \Gamma_i^{-1/2} dW_i(t), X_i(0) = 0$ $\Gamma_i \underset{i.i.d}{\sim} G(a,\lambda)$, $\Phi_i \vert \Gamma_i = \gamma \sim \mathcal{N}(\mu, \gamma^{-1}\omega^2)$					
Parameter	True value	$N = 50$		$N = 100$	
		$\Delta = 0.025$ ($n = 200$)	$\Delta = 0.005$ ($n = 1000$)	$\Delta = 0.025$ ($n = 200$)	$\Delta = 0.005$ ($n = 1000$)
Method 1					
ρ	1.00	1.00 (0.08)	1.00 (0.08)	0.99 (0.05)	1.00 (0.05)
μ	0.50	0.50 (0.06)	0.50 (0.06)	0.49 (0.04)	0.50 (0.05)
ω^2	0.10	0.10 (0.05)	0.09 (0.04)	0.10 (0.04)	0.10 (0.04)
m	4.00	4.10 (0.22)	4.01 (0.21)	4.12 (0.16)	4.03 (0.16)
t	1.32	1.30 (0.05)	1.31 (0.05)	1.31 (0.04)	1.32 (0.04)
Method 2					
ρ	1.00	0.99 (0.07)	1.00 (0.08)	0.98 (0.05)	0.99 (0.05)
μ	0.50	0.49 (0.06)	0.50 (0.06)	0.48 (0.04)	0.49 (0.04)
ω^2	0.10	0.08 (0.04)	0.09 (0.04)	0.09 (0.03)	0.09 (0.04)
m	4.00	3.95 (0.21)	3.99 (0.21)	3.97 (0.14)	4.01 (0.14)
t	1.32	1.26 (0.05)	1.30 (0.05)	1.27 (0.04)	1.31 (0.03)

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

Model 3: $dX_i(t) = (\Phi_{i,1}X_i(t) + \Phi_{i,2})dt + \Gamma_i^{-1/2}dW_i(t), X_i(0) = 0$ $\Gamma_i \underset{i.i.d}{\sim} G(a,\lambda)$, $\Phi_{i,1}\|\Gamma_i = \gamma \sim \mathcal{N}(\mu_1, \gamma^{-1}\omega_1^2)$, $\Phi_{i,2}\|\Gamma_i = \gamma \sim \mathcal{N}(\mu_2, \gamma^{-1}\omega_2^2)$					
Parameter	True value	$N = 50$		$N = 100$	
		$\Delta = 0.025$ ($n = 200$)	$\Delta = 0.005$ ($n = 1000$)	$\Delta = 0.025$ ($n = 200$)	$\Delta = 0.005$ ($n = 1000$)
Method 1					
μ_1	-0.50	-0.50 (0.10)	-0.50 (0.08)	-0.51 (0.08)	-0.51 (0.05)
μ_2	1.00	1.01 (0.14)	1.01 (0.13)	1.01 (0.10)	1.00 (0.07)
ω_1^2	0.10	0.13 (0.14)	0.14 (0.22)	0.16 (0.20)	0.12 (0.07)
ω_2^2	0.50	0.61 (0.35)	0.59 (0.35)	0.63 (0.45)	0.55 (0.20)
m	4.00	4.09 (0.31)	4.02 (0.22)	4.05 (0.20)	4.01 (0.15)
t	1.32	1.31 (0.08)	1.33 (0.06)	1.30 (0.05)	1.32 (0.04)
Method 2					
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)

Table 5: Empirical mean and (empirical s.d.) of estimators for different values of N and n, truncations $c = k = 0.1$, $m = a/\lambda$, $t = \psi(a) - \log\lambda$.

Model 4: $dX_i(t) = (\rho X_i(t) - \Phi_i)dt + \Gamma_i^{-1/2}\sqrt{\|X_i(t)\|}dW_i(t), X_i(0) = 5$ $\Gamma_i \underset{i.i.d}{\sim} G(a,\lambda)$, $\Phi_i\|\Gamma_i = \gamma \sim \mathcal{N}(\mu, \gamma^{-1}\omega^2)$					
Parameter	True value	$N = 50$		$N = 100$	
		$\Delta = 0.025$ ($n = 200$)	$\Delta = 0.005$ ($n = 1000$)	$\Delta = 0.025$ ($n = 200$)	$\Delta = 0.005$ ($n = 1000$)
Method 1					
ρ	4.00	3.97 (0.68)	4.04 (0.69)	3.85 (0.58)	3.88 (0.61)
μ	1.00	0.99 (0.16)	1.01 (0.17)	0.96 (0.14)	0.97 (0.15)
ω^2	0.10	0.10 (0.06)	0.10 (0.06)	0.11 (0.06)	0.10 (0.05)
m	4.00	4.09 (0.25)	3.98 (0.21)	4.14 (0.23)	4.02 (0.15)
t	1.32	1.30 (0.06)	1.31 (0.05)	1.32 (0.05)	1.32 (0.04)
Method 2					
ρ	4.00	3.95 (0.49)	4.02 (0.48)	3.82 (0.33)	3.90 (0.34)
μ	1.00	0.99 (0.12)	1.00 (0.12)	0.95 (0.08)	0.97 (0.09)
ω^2	0.10	0.09 (0.04)	0.10 (0.04)	0.09 (0.03)	0.09 (0.03)
m	4.00	4.03 (0.20)	3.99 (0.19)	4.08 (0.15)	4.02 (0.15)
t	1.32	1.28 (0.05)	1.31 (0.05)	1.30 (0.04)	1.32 (0.04)

Table 6: Empirical mean and (empirical s.d.) of estimators for different values of N and n, truncations $c = k = 0.1$, $m = a/\lambda$, $t = \psi(a) - \log\lambda$.

5 Implementation on real data

We now analyze some real neuronal data that are available in the R-package `mixedsde`. These data were first studied by Picchini et al. (2008, 2010). A complete description can be found in Dion (2016). The dataset is composed of $N = 240$ membrane potential trajectories. The observations are expressed in Volts. Each trajectory is observed at $n = 2000$ equidistant time points on a common time interval $[0, T]$ with $T = 0.3$s. Hence, the sampling time interval is $\Delta = 0.00015$s. Models of the form:

$$dX_i(t) = (\Phi_{i,1} - \Phi_{i,2}X_i(t))dt + \Psi_i\sigma(X_i(t))dW_i(t) \quad , \quad X_i(0) = x_i, \tag{37}$$

have been proposed for describing this kind of data. This formalism includes the Ornstein-Uhlenbeck model (OU): $\sigma(x) = 1$, and the Cox-Ingersoll-Ross model (CIR): $\sigma(x) = \sqrt{|x|}$. Note that fixed effect Ornstein-Uhlenbeck diffusions are often used for neuronal data. The fixed effect CIR model was investigated by Höpfner (2007).
The parameters in (37) have a biological meaning: $\Phi_{i,1}$ represents the local average input received by the neuron after the i^{th} spike, and $\Phi_{i,2}$ is the time constant of the neuron. The parameters $\Phi_{i,1}$, $\Phi_{i,2}$ and Ψ_i can either be fixed or random but to our knowledge, only models with fixed effects in the diffusion coefficient were used for describing the present dataset (see Picchini et al. (2008, 2010); Dion (2016); Dion et al. (2016)). Here, we try to answer two questions:

1) Which of the OU or the CIR is the most appropriate SDE for describing the data?

2) Which parameter should be fixed or random?

For that purpose, we estimate the OU in different configurations:

a. $\Phi_{i,1}$ random s.t. $\Phi_{i,1}|\Gamma_i \sim \mathcal{N}(\mu_1, \omega_1^2/\Gamma_i)$, $\Phi_{i,2} = \mu_2$ fixed, $\Psi_i = 1/\sqrt{\Gamma_i}$ random, $\Gamma_i \sim G(a, \lambda)$, ,

b. $\Phi_{i,1} = \mu_1$ fixed, $\Phi_{i,2}$ random s.t. $\Phi_{i,2}|\Gamma_i \sim \mathcal{N}(\mu_2, \omega_2^2/\Gamma_i)$, $\Psi_i = 1/\sqrt{\Gamma_i}$ random, $\Gamma_i \sim G(a, \lambda)$,

c. $\Phi_{i,1}$ and $\Phi_{i,2}$ random s.t. $\Phi_{i,1}|\Gamma_i \sim \mathcal{N}(\mu_1, \omega_1^2/\Gamma_i)$ and $\Phi_{i,2}|\Gamma_i \sim \mathcal{N}(\mu_2, \omega_2^2/\Gamma_i)$, $\Phi_{i,1}, \Phi_{i,2}$ independent, $\Psi_i = 1/\sqrt{\Gamma_i}$ random, $\Gamma_i \sim G(a, \lambda)$.

and the CIR in configuration b. (see remarks about the CIR in the simulation section).
To allow the estimation of both CIR and OU, the two trajectories taking negative values are omitted, leading to $N = 238$. The observations take very small values (9×10^{-3} Volts on average). This can cause numerical difficulties such as lack of precision of the calculations. To avoid this, we change volts in millivolts and seconds in milliseconds ($T = 300$ ms). Parameter estimation can be performed by means of the methods described above. We favor the second method which is more stable numerically. We then identify the most appropriate model by comparing the BIC values. BIC is defined as $-2\log\widehat{\mathcal{L}} + d\log(N)$ where $\widehat{\mathcal{L}}$ denotes the likelihood of the observations evaluated at the parameter estimate and d is the number of model parameters. Here, $\mathcal{L}$ is not explicit, thus approximated by the approximate likelihood given by equations (13) and (14). The different BIC values and the parameter estimates are reported in Table 5. Note that $E(1/\Gamma_i) = \lambda/(a-1)$ is a more relevant value in this model as it represents the expectation of the square of the random effect Ψ_i. From the BIC criterion, the Cox-Ingersoll-Ross model better suits to the data than Ornstein-Uhlenbeck models. To our knowledge, there does not exist any

Model	μ_1	ω_1^2	μ_2	ω_2^2	a	λ	$\lambda/(a-1)$	BIC
OU.a	0.380	-	0.039	2.89×10^{-4}	16.203	2.932	0.193	8.80×10^8
OU.b	0.379	0.015	0.037	-	16.203	2.932	0.193	8.80×10^8
OU.c	0.377	0.017	0.038	3.37×10^{-5}	16.203	2.932	0.193	8.80×10^8
CIR.b	0.482	-	0.049	0.003	12.836	0.349	0.029	1.05×10^8

Table 7: Neuronal data: BIC values of the different candidate models.

analysis of these data by means of Cox-Ingersoll-Ross models. The parameter estimations obtained with the Ornstein-Uhlenbeck models are nevertheless consistent with those available in the literature (Picchini et al. (2008, 2010); Dion (2016); Dion et al. (2016)) while taking into account the change of scale.

6 Concluding remarks

In this paper, we study one-dimensional stochastic differential equations including linear random effects Φ_i in the drift and one multiplicative random effect $\Gamma_i^{-1/2}$ in the diffusion coefficient. Such framework covers many examples of the literature in which it is often relevant to consider random effects in the drift and in the diffusion coefficient. We assume that Γ_i follows a Gamma distribution, whereas conditional on Γ_i, Φ_i is multivariate Gaussian. In this specific case, we propose two methods for the estimation of the model parameters from discrete observations of N trajectories on a fixed length time interval. The asymptotic properties of the parameter estimates are studied in both cases and lead to asymptotically equivalent estimators. A simulation study shows that the two methodologies perform well in practice. Both methods are implemented on a real dataset.

The proofs are derived by assuming the same initial condition for the N trajectories, but the two estimation methods can also be used with different initial conditions, which occurs in the application on the neuronal data presented in the last section.

The present study opens up many perspectives. First, a direct extension of this work would be to consider multidimensional diffusion with mixed effects:

$$d\,X_i(t) = B(X_i(t))\Phi_i dt + \Psi_i \Sigma(X_i(t))\,dW_i(t), \quad X_i(0) = x,\ i = 1, \ldots, N, \tag{38}$$

where $X_i(t) \in \mathbb{R}^k$, for $B(x)$ is a (k,d)- matrix, Φ_i is a d-dimensional random effect, $\Sigma(x)$ is a (k,k)-matrix and W_i is a $\mathbb{R}^k$-Brownian motion. Under the assumption that $C(x) = \Sigma(x)\Sigma(x)'$ is invertible, the likelihood of the Euler scheme is easily computed and can be integrated with respect to the distribution (2) of (Φ_i, Ψ_i) to obtain the extension of the formula (14).

Second, the proposed estimation methods require a large number of observations per trajectory. Building specific estimation methods for fixed n would be an interesting working perspective from a practical point of view.

Finally, as illustrated in the application, real data analysis raises the difficult question of model selection and especially the choice of considering fixed or random effects in the drift and in the diffusion coefficient. We have suggested the use of the Bayesian Information Criterion (BIC) but as pointed out in Delattre et al. (2014) and Delattre and Poursat (2016), the use of BIC may be inadequate in mixed models.

Building statistical testing methods to decide which parameter is a challenging perspective.

References

Alfonsi, A. (2005). On the discretization schemes for the CIR (and Bessel squared) processes. *Monte Carlo methods and Applications*, 11:355–384.

Ben Alaya, M. and Kebaier, M. (2012). Parameter estimation for the square-root diffusions: Ergodic and nonergodic cases. *Stochastic Models*, 28:609–634.

Ben Alaya, M. and Kebaier, M. (2013). Asymptotic behavior of the maximum likelihood estimator for ergodic and nonergodic square-root diffusions. *Stochastic Analysis and Applications*, 31:552–573.

Delattre, M., Genon-Catalot, V., and Larédo, C. (2016). Parametric inference for discrete observations of diffusion with mixed effects in the drift or in the diffusion coefficient. *preprint MAP5 2016-15*.

Delattre, M., Genon-Catalot, V., and Samson, A. (2013). Maximum likelihood estimation for stochastic differential equations with random effects. *Scandinavian Journal of Statistics*, 40:322–343.

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

Delattre, M. and Lavielle, M. (2013). Coupling the saem algorithm and the extended kalman filter for maximum likelihood estimation in mixed-effects diffusion models. *Statistics and its interface*, 6(4):519–532.

Delattre, M. and Poursat, M. (2016). Bic strategies for model choice in a population approach. *preprint*.

Delattre, M., Poursat, M., and Lavielle, M. (2014). A note on BIC in mixed-effects models. *Electronic Journal of Statistics*, 8:456–475.

Dion, C. (2016). Nonparametric estimation in a mixed-effect ornstein-uhlenbeck model. *Metrika*.

Dion, C., Hermann, S., and Samson, A. (2016). Mixedsde : a r package to fit mixed stochastic differential equations. *Preprint hal-01305574 https://hal.archives-ouvertes.fr/hal-01305574*.

Ditlevsen, S. and De Gaetano, A. (2005). Mixed effects in stochastic differential equation models. *REVSTAT Statistical Journal*, 3:137–153.

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

Forman, J. and Picchini, U. (2016). Stochastic differential equation mixed effects models for tumor growth and response to treatment. *preprint*.

Höpfner, R. (2007). On a set of data for the membrane potential in a neuron. *Math. Biosci.*, 207(2):275–301.

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

Overbeck, L. (1998). Estimation for continuous branching processes. *Scandinavian Journal of Statistics*, 25:111–126.

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.

Picchini, U., Ditlevsen, S., and De Gaetano, A. (2008). Parameters of the diffusion leaky integrate-and-fire neuronal model for slowly fluctuating signal. *Neural Computation*, 20:2696–2714.

Pinheiro, J. and Bates, D. (2000). *Mixed-effect models in S and Splus.* Springer-Verlag.

The corresponding author is:
Catherine Larédo (catherine.laredo@jouy.inra.fr)
Laboratoire MaIAGE, I.N.R.A., Jouy-en-Josas, France.

7 Appendix

7.1 Proof of Proposition 1

Assume first that $\mathbf{\Omega}$ is invertible. Integrating (5) with respect to the distribution $\mathcal{N}(\boldsymbol{\mu}, \gamma^{-1}\mathbf{\Omega})$ yields the expression:

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

where

$$T_{i,n}(\boldsymbol{\mu}, \mathbf{\Omega}) = \boldsymbol{\mu}'\mathbf{\Omega}^{-1}\boldsymbol{\mu} - \mathbf{m}'_{i,n}\mathbf{\Sigma}^{-1}_{i,n}\mathbf{m}_{i,n}, \tag{39}$$

and

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

Computations using matrices equalities and (7),(8), (10) yield that $T_{i,n}(\boldsymbol{\mu}, \mathbf{\Omega})$ is equal to the expression given in (11), *i.e.*:

$$T_{i,n}(\boldsymbol{\mu}, \mathbf{\Omega}) = (\boldsymbol{\mu} - V^{-1}_{i,n}U_{i,n})'R^{-1}_{i,n}(\boldsymbol{\mu} - V^{-1}_{i,n}U_{i,n}) - U'_{i,n}V^{-1}_{i,n}U_{i,n}.$$

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

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

We multiply $\Lambda_n(X_{i,n}, \gamma, \boldsymbol{\mu}, \mathbf{\Omega})$ 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 $\mathbf{\Omega}$. Consequently, we can consider $\mathcal{L}_n(X_{i,n}, \vartheta)$ as an approximate likelihood for non invertible $\mathbf{\Omega}$. $\square$

We now consider the set $\mathbf{F}_{N,n}$ introduced in (16) which is defined under (H4). Let us stress that this set does not depend on unknown parameters.

7.2 Proof of Lemma 1

By (10)-(11), we have, using $\boldsymbol{\mu} - V_{i,n}^{-1}U_{i,n} = V_{i,n}^{-1}(V_{i,n}\boldsymbol{\mu} - U_{i,n})$,

$$T_{i,n}(\mu, \mathbf{\Omega}) = \boldsymbol{\mu}'V_{i,n}(I_d + \mathbf{\Omega}V_{i,n})^{-1}\boldsymbol{\mu} + U_{i,n}'\left[(I_d + \mathbf{\Omega}V_{i,n})^{-1} - I_d\right]V_{i,n}^{-1}U_{i,n} - 2\boldsymbol{\mu}'V_{i,n}(I_d + \mathbf{\Omega}V_{i,n})^{-1}V_{i,n}^{-1}U_{i,n}.$$

For two nonnegative symmetric matrices A, B, we have $A \leq B$ if and only if for all vector x, $x'Ax \leq x'Bx$ and in this case, for all vectors x, y, $2x'Ay \leq (x+y)'B(x+y)$. We write:

$$(I_d + \mathbf{\Omega}V_{i,n})^{-1} - I_d = (I_d + \mathbf{\Omega}V_{i,n})^{-1}\left[I_d - (I_d + \mathbf{\Omega}V_{i,n})\right] = -(I_d + \mathbf{\Omega}V_{i,n})^{-1}\mathbf{\Omega}V_{i,n} \leq 0.$$

Therefore, using (H4) and the previous inequality with $A = V_{i,n}(I_d + \mathbf{\Omega}V_{i,n})^{-1}V_{i,n}^{-1}, B = I_d$ yields

$$U_{i,n}'\left[(I_d + \mathbf{\Omega}V_{i,n})^{-1} - I_d\right]V_{i,n}^{-1}U_{i,n} = -U_{i,n}'(I_d + \mathbf{\Omega}V_{i,n})^{-1}\mathbf{\Omega}U_{i,n} \geq -U_{i,n}'\mathbf{\Omega}U_{i,n} \geq -c_1\|U_{i,n}\|^2,$$

$$-2\boldsymbol{\mu}'V_{i,n}(I_d + \mathbf{\Omega}V_{i,n})^{-1}V_{i,n}^{-1}U_{i,n} \geq -(\boldsymbol{\mu} + U_{i,n})'(\boldsymbol{\mu} + U_{i,n}) \geq -2(\|U_{i,n}\|^2 + m^2).$$

This yields

$$T_{i,n}(\mu, \mathbf{\Omega}) \geq -\left[(c_1 + 2)\|U_{i,n}\|^2 + 2m^2\right].$$

We consider (see (12)-(16)): $M_{i,n} = \max\{c_1 + 2, 2m^2\}(\|U_{i,n}\|^2 + 1)$, and for $\alpha > 0$, the set $F_{i,n} = \{\frac{1}{n}(S_{i,n} - M_{i,n}) \geq \frac{\alpha}{\sqrt{n}}\}$. Then, for all ϑ, we have $F_{i,n} \subset E_{i,n}(\vartheta)$.
For simplicity of notations, we continue the proof for $d = 1$.
Fix a value ϑ_0. We study the probability of $F_{i,n}$ and $\mathbf{F}_{N,n}$ under $\mathbb{P}_{\vartheta_0}$. For all fixed $i = 1, \ldots, N$, as n tends to infinity, $U_{i,n}$ tends to $U_i(T)$ and $V_{i,n}$ tends to $V_i(T)$. Hence, $M_{i,n}/n$ tends to 0 in probability. Moreover, $\frac{1}{n}S_{i,n}$ tends to Γ_i^{-1}. Thus, $\mathbb{P}_{\vartheta_0}(F_{i,n})$ tends to 1 as n tends to infinity. This is not enough for our purpose because we have to deal with the N trajectories simultaneously. Thus, we look for a condition on N, n ensuring that $\mathbb{P}_{\vartheta_0}(\mathbf{F}_{N,n}) \to_{N,n\to+\infty} 1$. As $\mathbb{P}_{\vartheta_0}(\cup_{i=1}^N F_{i,n}^c) \leq N\mathbb{P}_{\theta_0}(F_{1,n}^c)$, we will find a condition ensuring that $N\mathbb{P}_{\vartheta_0}(F_{1,n}^c)$ tends to 0. More precisely, we prove that, if $a_0 > 4$,

$$\mathbb{P}_{\vartheta_0}(F_{1,n}^c) \lesssim \frac{1}{n^2} \tag{40}$$

which explains the constraint $N/n^2 \to 0$.
Let us study the set $F^c_{1,n} = \{\frac{1}{n}(S_{1,n} - M_{1,n}) \leq \frac{\alpha}{\sqrt{n}}\}$. For the rest of the proof, as we consider only the process X_1 and let n tend to infinity, we set $X_1 = X$, $\Psi_1 = \Psi, \Gamma_1 = \Gamma = \Psi^{-2}, \Phi_1 = \Phi$, $S_{1,n} = S_n(X)$, $M_{1,n} = M_n(X)$, $U_{1,n} = U_n(X), V_{1,n} = V_n(X)$. We split successively by $\{M_n(X)/n > \alpha/\sqrt{n}\}$ and by $\{\Psi^2 < 4\alpha/\sqrt{n}\}$ and obtain:

$$\begin{aligned}\mathbb{P}_{\vartheta_0}(\frac{1}{n}(S_n(X) - M_n(X)) &\leq \frac{\alpha}{\sqrt{n}}) \leq \mathbb{P}_{\vartheta_0}(\frac{M_n(X)}{n} > \frac{\alpha}{\sqrt{n}}) + \mathbb{P}_{\vartheta_0}(\frac{S_n(X)}{n} \leq \frac{2\,\alpha}{\sqrt{n}}) \\ &\leq \mathbb{P}_{\vartheta_0}(\frac{M_n(X)}{n} > \frac{\alpha}{\sqrt{n}}) + \mathbb{P}_{\vartheta_0}(\Psi^2 < \frac{4\,\alpha}{\sqrt{n}}) + \mathbb{P}_{\vartheta_0}(\Psi^2 - \frac{S_n(X)}{n} \geq \frac{1}{2}\Psi^2).\end{aligned}$$

The last part is obtained as follows. Note that $(\Psi^2 \geq \frac{4\,\alpha}{\sqrt{n}}) = (\Psi^2 - \frac{2\,\alpha}{\sqrt{n}} \geq \frac{\Psi^2}{2})$. Therefore,

$$(\frac{S_n(X)}{n} \leq \frac{2\,\alpha}{\sqrt{n}}, \Psi^2 \geq \frac{4\,\alpha}{\sqrt{n}})) = (\Psi^2 - \frac{S_n(X)}{n} \geq \Psi^2 - \frac{2\,\alpha}{\sqrt{n}}, \Psi^2 \geq \frac{4\,\alpha}{\sqrt{n}})) \subset (\Psi^2 - \frac{S_n(X)}{n} \geq \frac{1}{2}\Psi^2).$$

We study successively the rates of convergence to 0 of the three terms of the bound of $\mathbb{P}_{\vartheta_0}(\frac{1}{n}(S_n(X) - M_n(X) \leq \frac{\alpha}{\sqrt{n}})$. Using the Markov inequality for the middle term,

$$\mathbb{P}_{\vartheta_0}(\Psi^2 < \frac{4\,\alpha}{\sqrt{n}}) = \mathbb{P}_{\vartheta_0}(\Gamma > \frac{\sqrt{n}}{4\,\alpha}) \leq (\frac{4\alpha}{\sqrt{n}})^4 \mathbb{E}_{\vartheta_0}\Gamma^4 \lesssim n^{-2}.$$

For the first term, we now check that, if $\mathbb{E}_{\vartheta_0}\Gamma^{-4} < +\infty$, *i.e.* $a_0 > 4$,

$$\mathbb{P}_{\vartheta_0}(M_n(X) \geq \alpha\sqrt{n}) \lesssim n^{-2}. \tag{41}$$

For this, we have to study $\mathbb{E}_{\vartheta_0}(U_n(X)^8|\Phi, \Psi)$ (see (16)). We write $U_n(X) = U_n^{(1)} + U_n^{(2)}$, where

$$U_n^{(1)} = \Phi \sum_{j=1}^{n} \frac{b(X(t_{j-1}))}{\sigma^2(X(t_{j-1}))} \int_{t_{j-1}}^{t_j} b(X(s))ds, \quad |U_n^{(1)}| \leq |\Phi| \frac{K^2}{\sigma_0^2} T.$$

Therefore, $\mathbb{E}_{\vartheta_0}(U_n^{(1)})^8|\Phi, \Psi) \lesssim \Phi^8$. Next,

$$U_n^{(2)} = \int_0^T J_s^n dW_s, \quad J_s^n = \Psi \sum_{j=1}^{n} 1_{(t_{j-1},t_j]}(s) \frac{b(X(t_{j-1}))}{\sigma^2(X(t_{j-1}))} \sigma(X(s)).$$

For $s \in [0,T]$, $|J_s^n| \leq \Psi K^2/\sigma_0^2$. This yields, by the B-D-G inequality, $\mathbb{E}_{\vartheta_0}(((U_n^{(2)})^8|\Phi, \Psi) \leq CK^{16}\sigma_0^{-16}\Psi^8 T^4$. Now, $\mathbb{E}_{\vartheta_0}\Phi^8 \lesssim \mu^8 + \omega^8 \mathbb{E}_{\vartheta_0}\Gamma^{-4}$ and $\mathbb{E}_{\vartheta_0}\Psi^8 = \mathbb{E}_{\vartheta_0}\Gamma^{-4}$. Hence (41).
It remains to study $\mathbb{P}_{\vartheta_0}(\Psi^2 - \frac{S_n(X)}{n} \geq \frac{1}{2}\Psi^2)$ which is the most difficult term.

Lemma 3. *Under (H1)-(H2), if* $a_0 > 2$,

$$\mathbb{P}_{\vartheta_0}(\Psi^2 - \frac{S_n(X)}{n} \geq \frac{1}{2}\Psi^2) \lesssim n^{-2}.$$

Proof of Lemma 3. We use the following classical development (see Comte *et al.* 2007, p.522):

$$\begin{aligned}\frac{(X((j+1)\Delta)) - X(j\Delta))^2}{\Delta} &= \Psi^2\sigma^2(X(j\Delta)) + \Psi^2 V^{(1)}_{j\Delta}(X) + 2\Psi^3 V^{(2)}_{j\Delta}(X) \\ &+ 2\Phi\Psi V^{(3)}_{j\Delta}(X) + R_{j\Delta}(\Phi,\Psi,X),\end{aligned}$$

where

$$V^{(1)}_{j\Delta}(X) = \frac{1}{\Delta}\left[\left(\int_{j\Delta}^{(j+1)\Delta}\sigma(X(s))dW(s)\right)^2 - \int_{j\Delta}^{(j+1)\Delta}\sigma^2(X(s))ds\right] \tag{42}$$

$$V^{(2)}_{j\Delta}(X) = \frac{1}{\Delta}\int_{j\Delta}^{(j+1)\Delta}((j+1)\Delta - u)\sigma'(X(u))\sigma^2(X(u)dW(u) \tag{43}$$

$$V^{(3)}_{j\Delta}(X) = b(X(j\Delta))\int_{j\Delta}^{(j+1)\Delta}\sigma(X(s))dW(s) \tag{44}$$

$$\begin{aligned}R_{j\Delta}(\Phi,\Psi,X) &= \frac{\Phi^2}{\Delta}\left(\int_{j\Delta}^{(j+1)\Delta} b(X(s))ds\right)^2 \\ &+\frac{2\Phi\Psi}{\Delta}\int_{j\Delta}^{(j+1)\Delta}(b(X(s) - b(X(j\Delta)))ds\int_{j\Delta}^{(j+1)\Delta}\sigma(X(s))dW(s) \\ &+\frac{1}{\Delta}\int_{j\Delta}^{(j+1)\Delta}((j+1)\Delta - u)K_{\Phi,\Psi}(X(u))du\end{aligned} \tag{45}$$

where $K_{\Phi,\Psi}(.)) = \Psi^2[\Phi 2b\sigma\sigma' + \Psi^2\sigma^2(\sigma^2)'']$. Therefore,

$$\frac{S_n(X)}{n} = \Psi^2 + \nu_n^{(1)} + \nu_n^{(2)} + \nu_n^{(3)} + \nu_n^{(4)}. \tag{46}$$

And

$$\mathbb{P}_{\vartheta_0}(\Psi^2 - \frac{S_n(X)}{n} \geq \frac{1}{2}\Psi^2|\Phi,\Psi) \leq \sum_{i=1}^{4}\mathbb{P}_{\vartheta_0}(-\nu_n^{(i)} \geq \frac{\Psi^2}{8}|\Phi,\Psi)$$

For the term containing $\nu_n^{(1)}$, we apply Lemma 3 of Comte *et al.* (2007, p.536) which states that, for a continous process X adapted to $(\mathcal{F}_t)$, when the function $\sigma(x)$ is upper bounded by K and for $t(x)$ a continous function:

$$\mathbb{P}(\sum_{j=1}^{n} t(X(j\Delta))V^{(1)}_{j\Delta}(X) \geq n\epsilon, \|t\|_n^2 \leq v^2) \leq \exp\left[-Cn\frac{\epsilon^2/2}{2K^4v^2 + \epsilon\|t\|_\infty K^2 v}\right]$$

where C is a universal constant and $\|t\|_n^2 = n^{-1}\sum_{j=1}^n t^2(X(j\Delta))$.
We apply this result conditoning on $\Phi = \varphi$, $\Psi = \psi$ with $t(x) = -1/\sigma^2(x)$, $v^2 = \sigma_0^{-4}$, $\|t\|_\infty = \sigma_0^{-2}$. Noting that

$$-\nu_n^{(1)} = \Psi^2\frac{1}{n}\sum_{j=0}^{n-1}\frac{-1}{\sigma^2(X(j\Delta))}V^{(1)}_{j\Delta}(X), \tag{47}$$

we choose $\epsilon = 1/8$ and simplifying by Ψ^2, this yields

$$\mathbb{P}_{\vartheta_0}(-\nu_n^{(1)} \geq \Psi^2/8 | \Phi = \varphi, \Psi = \psi) \leq \exp\left(-Cn\frac{1/(2\times 8^2)}{2K^4\sigma_0^{-4} + (1/8)\sigma_0^{-4}K^2}\right).$$

The upper bound is deterministic so $\mathbb{P}_{\vartheta_0}(-\nu_n^{(1)} \geq \Psi^2/8) \leq \exp(-cn)$, where c depends on K and σ_0.
Next, we study $\nu_n^{(2)}$:

$$-\nu_n^{(2)} = \frac{\Psi^3}{n}\int_0^{n\Delta} H_n(s)dW(s), \quad \text{with} \tag{48}$$

$$H_n(s) = -\sum_{j=0}^{n-1}\frac{((j+1)\Delta - s)\sigma'(X(s))\sigma^2(X(s))}{\Delta\sigma^2(X(j\Delta))}1_{(j\Delta,(j+1)\Delta]}(s), \quad |H_n(s)| \leq \frac{LK^2}{\sigma_0^2}.$$

Hence,

$$\begin{aligned}\mathbb{P}_{\vartheta_0}(-\nu_n^{(2)} \geq \Psi^2/8 | \Phi = \varphi, \Psi = \psi)) &= \mathbb{P}_{\theta_0}(\int_0^{n\Delta} H_n(s)dW(s) \geq n/(8\Psi) |\ \Phi = \varphi, \Psi = \psi) \\ &\leq \left(\frac{8\Psi}{n}\right)^2 \mathbb{E}_{\theta_0}(\int_0^{n\Delta} H_n^2(s)ds | \Phi = \varphi, \Psi = \psi) \leq \left(\frac{8\Psi}{n}\frac{LK^2}{\sigma_0^2}\right)^2 T.\end{aligned}$$

Therefore, provided that $\mathbb{E}_{\vartheta_0}\Gamma^{-1} < +\infty$, *i.e.* $a_0 > 1$, $\mathbb{P}_{\vartheta_0}(-\nu_n^{(2)} \geq \Psi^2/8) \lesssim n^{-2}$.
Next, we study

$$-\nu_n^{(3)} = \frac{\Psi}{n}\int_0^{n\Delta} K_n(s)dW(s) \quad \text{with} \tag{49}$$

$$K_n(s) = -\sum_{j=0}^{n-1} 1_{(j\Delta,(j+1)\Delta]}(s)\frac{2\Phi b(X(j\Delta))\sigma(X(s))}{\sigma^2(X(j\Delta))}, \quad |K_n(s)| \leq 2|\Phi|\frac{K^2}{\sigma_0^2}.$$

Thus,

$$\begin{aligned}\mathbb{P}_{\vartheta_0}(-\nu_n^{(3)} \geq \Psi^2/8 | \Phi = \varphi, \Psi = \psi)) &= \mathbb{P}_{\vartheta_0}(\int_0^{n\Delta} K_n(s)dW(s) \geq n\Psi/8 | \Phi = \varphi, \Psi = \psi) \\ &\leq \left(\frac{8}{n\psi}\right)^2 \times 4\varphi^2 T(\frac{K^2}{\sigma_0^2})^2\end{aligned}$$

We have $\mathbb{E}_{\vartheta_0}(\Psi^{-2}\Phi^2) = \mathbb{E}_{\vartheta_0}(\mu\sqrt{\Gamma} + \omega\varepsilon)^2 \leq 2(\mu^2\mathbb{E}_{\vartheta_0}\Gamma + \omega^2) < +\infty$. Therefore, $\mathbb{P}_{\vartheta_0}(-\nu_n^{(3)} \geq \Psi^2/8) \lesssim n^{-2}$.
There remains to study:

$$\nu_n^{(4)} = \frac{1}{n}\sum_{j=0}^{n-1}\frac{R_{j\Delta}(\Phi, \Psi, X)}{\sigma^2(X(j\Delta))}. \tag{50}$$

We have:

$$\begin{aligned}\mathbb{P}_{\vartheta_0}(-\nu_n^{(4)} \geq \Psi^2/8 | \Phi = \varphi, \Psi = \psi) &\leq \frac{8^2}{\psi^4}\mathbb{E}_{\vartheta_0}(\nu_n^{(4)})^2 || \Phi = \varphi, \Psi = \psi) \\ &\lesssim \frac{8^2}{\psi^4}\mathbb{E}_{\vartheta_0} R_{j\Delta}^2(\Phi, \Psi, X) | \Phi = \varphi, \Psi = \psi),\end{aligned}$$

where (see (45)) $R^2_{j\Delta}(\Phi,\Psi,X) \lesssim \Delta^2\left(\Phi^4 + (\Phi\Psi^2+\Psi^4)^2\right) + \Delta^{-2}4\Phi^2\Psi^2 I_j^2$ and

$$I_j = \int_{j\Delta}^{(j+1)\Delta} (b(X(s) - b(X(j\Delta)))ds \int_{j\Delta}^{(j+1)\Delta} \sigma(X(s))dW(s) \tag{51}$$

Using Lemma 5 below, we find $\mathbb{E}_{\vartheta_0}(I_j^2|\Phi=\varphi,\Psi=\psi) \lesssim \Delta^4(\varphi^2+\psi^2)$. Finally,

$$\frac{8^2}{\psi^4}\mathbb{E}_{\vartheta_0}R^2_{j\Delta}(\Phi,\Psi,X)|\Phi=\varphi,\Psi=\psi) \lesssim \Delta^2\left(\psi^{-4}\varphi^4+\psi^4+\varphi^2+\psi^{-2}\varphi^4)\right). \tag{52}$$

Therefore, we have $\mathbb{E}_{\vartheta_0}8^2\Psi^{-4}4R^2_{j\Delta}(\Phi,\Psi,X) \lesssim \Delta^2$ if $\mathbb{E}_{\vartheta_0}(\Psi^{-4}\Phi^4+\Psi^4+\Phi^2+\Psi^{-2}\Phi^4) < +\infty$ *i.e.* $\mathbb{E}_{\vartheta_0}(\Gamma^{-2}) < +\infty$ which holds for $a_0 > 2$.
This ends the proof of Lemma 3. □
The proof of Lemma 1 is now complete. □

7.3 Proof of Theorem 1

Proof of Lemma 2
Recall that we consider $d=1$ (univariate effect in the drift). For simplicity of notations, we omit the index n but keep the index i for the i-th sample path. We have (see (19 and (22)):

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

Note that, using (11):

$$T_i(\mu,\omega^2) = \frac{V_i}{(1+\omega^2 V_i)}\left(\mu - \frac{U_i}{V_i}\right)^2 - \frac{U_i^2}{V_i} = \frac{U_i^2}{(1+\omega^2V_i)}(\mu^2-\omega^2) - 2\mu\frac{U_i}{(1+\omega^2V_i)}, \tag{53}$$

By the Hölder inequality, we obtain using (8) and (53):

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

Now, we apply Lemmas 8 and 6 of Section 8.2 and (64) of Section 8.1 and this yields the first inequality. For the second inequality, we write:

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

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

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

Consequently,

$$\left|Z_i^{-1} - \frac{n}{S_i^{(1)}}\right|^p 1_{F_i} \lesssim \left(\frac{n}{C_i^{(1)}}\right)^{2p} \Psi_i^{-4p} \left(n^{p/2}\left(\frac{S_i^{(1)}}{n} - Z_i\right)^{2p} + \left|\frac{S_i^{(1)}}{n} - Z_i\right|^p\right). \tag{54}$$

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 (64) of Section 8.1):

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

And, we apply the first inequality to get the result. □

Now, we start proving Theorem 1. Omitting the index n in $A_{i,n}, B_{i,n}$, we have (see (20), (23), (24))

$$\begin{aligned}
N^{-1/2}\frac{\partial \mathbf{U}_{N,n}}{\partial \lambda}(\vartheta) &= N^{-1/2}\sum_{i=1}^N \left(\frac{a}{\lambda} - \Gamma_i\right) + R_1,\\
N^{-1/2}\frac{\partial \mathbf{U}_{N,n}}{\partial a}(\vartheta) &= N^{-1/2}\sum_{i=1}^N (-\psi(a) + \log\lambda + \log\Gamma_i) + R_2 + R_2',\\
N^{-1/2}\frac{\partial \mathbf{U}_{N,n}}{\partial \mu}\vartheta) &= N^{-1/2}\sum_{i=1}^N \Gamma_i A_i + R_3 = N^{-1/2}\sum_{i=1}^N \Gamma_i A_i(T;\mu,\omega^2) + R_3' + R_3\\
N^{-1/2}\frac{\partial \mathbf{U}_{N,n}}{\partial \omega^2}\vartheta) &= N^{-1/2}\frac{1}{2}\sum_{i=1}^N \left(\Gamma_i A_i^2 - B_i\right) + R_4\\
&= N^{-1/2}\frac{1}{2}\sum_{i=1}^N (\Gamma_i A_i^2(T;\mu,\omega^2) - B_i(T,\omega^2)) + R_4' + R_4.
\end{aligned}$$

The remainder terms are:

$$\begin{aligned}
R_1 &= N^{-1/2}\sum_{i=1}^N (\Gamma_i - Z_i^{-1}1_{F_i}), \quad R_2 = N^{-1/2}\sum_{i=1}^N (\log Z_i^{-1}\, 1_{F_i} - \log\Gamma_i),\\
R_2' &= N^{1/2}(\psi(a+(n/2)) - \log(a+(n/2))) + N^{-1/2}\sum_{i=1}^N 1_{F_i^c},\\
R_3 &= N^{-1/2}\sum_{i=1}^N A_i(Z_i^{-1}1_{F_i} - \Gamma_i), \quad R_4 = N^{-1/2}\frac{1}{2}\sum_{i=1}^N A_i^2(Z_i^{-1}1_{F_i} - \Gamma_i),\\
R_3' &= N^{-1/2}\sum_{i=1}^N \Gamma_i(A_i - A_i(T;\mu,\omega^2)),\\
R_4' &= N^{-1/2}\frac{1}{2}\sum_{i=1}^N (B_i - B_i(T,\omega^2)) - \Gamma_i(A_i^2 - A_i^2(T;\mu,\omega^2))).
\end{aligned}$$

The most difficult remainder terms are R_1, R_2, R_3, R_4. They are treated in Lemma 4 below. For the term R_2', we use that $(\psi(a+(n/2)) - \log(a+(n/2))) = 0(n^{-1})$ (see (63)) and that, for $a > 4$, $\mathbb{P}_\theta(F_i^c) \lesssim n^{-2}$ (see (40)) to get that $R_2' = O_P(\sqrt{N}/n)$.

Using Lemma 6 in Section 8.2, 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:

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

The first two components are exactly the score function corresponding to the exact observation of $(\Gamma_i, i=1,\ldots,N)$ (see section 8.3). 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

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

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 that

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

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

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

Hence, the covariance matrix of the last two components is equal to $J(\vartheta)$ defined in (25).
The proof of the last item relies on the same tools with more cumbersome computations but no additional difficulty. Note that this part only requires that N,n both tend to infinity without further constraint. So the proof is complete. □

Lemma 4. *Recall (19). Then, for $a>4$, (see (16) for the definition of $F_{i,n}$), R_1,R_2 are $O_P(\frac{\sqrt{N}}{n})$, R_3,R_4 are $O_P(\sqrt{\frac{N}{n}})$.*

Proof. The proof goes in several steps. We have introduced

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

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 Section 8)), we obtain:

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

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

Thus,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{n}{S_i^{(1)}}-\Gamma_i\right)=O_P(\frac{\sqrt{N}}{n}),\quad \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\log\frac{n}{S_i^{(1)}}-\log\Gamma_i\right)=O_P(\frac{\sqrt{N}}{n}). \tag{55}$$

Then, we have to study

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(Z_i^{-1}1_{F_i}-\frac{n}{S_i^{(1)}}\right),\quad \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\log{(Z_i^{-1})}1_{F_i}-\log\frac{n}{S_i^{(1)}}\right). \tag{56}$$

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 (40)). This implies, after developping and applying the Cauchy-Schwarz inequality, noting that $\Gamma_i=\Psi_i^{-2}$ has moments of any order,

$$\mathbb{E}_\vartheta\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{n}{S_i^{(1)}}1_{F_i^c}\right)^2=O(\frac{N}{n^2}),\qquad \text{for}\quad a>4. \tag{57}$$

Next, apply Lemma 2,

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

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

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

$$\mathbb{E}_\vartheta\Phi_i^4\lesssim 1+\mathbb{E}_\vartheta\Psi_i^4=1+\mathbb{E}_\vartheta\Gamma_i^{-2}<+\infty.$$

Thus, for $a>4$,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(Z_i^{-1}1_{F_i}-\frac{n}{S_i^{(1)}}\right)=O_P(\frac{\sqrt{N}}{n}). \tag{58}$$

Joining (55)(left)-(56)(left)-(57)-(58), we obtain that, for $a>4$, $R_1=O_P(\frac{\sqrt{N}}{n})$.
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 have

$$\mathbb{E}_\vartheta\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\log\frac{n}{S_i^{(1)}}1_{F_i^c}\right)^2=O(\frac{N}{n^2}),\qquad \text{for}\quad a>4. \tag{59}$$

And:

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

On F_i,

$$\left|\log(Z_i^{-1}) - \log\frac{n}{S_i^{(1)}}\right| \lesssim \frac{1}{\frac{C_i^{(1)}}{n}}\Psi_i^{-2}\left(\left|\frac{S_i^{(1)}}{n} - Z_i\right| + \sqrt{n}(\frac{S_i^{(1)}}{n} - Z_i)^2\right). \tag{60}$$

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:

$$\mathbb{E}_\vartheta\left(\left|\log(Z_i^{-1}) - \log\frac{n}{S_i^{(1)}}\right| |\Phi_i = \varphi, \Psi_i = \psi\right) \lesssim \frac{1}{n}(1 + \psi^{-2}(1+\varphi^2) + \psi^2(1+\varphi^4) + \varphi^2 + \varphi^4 + \psi^6). \tag{61}$$

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:

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

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

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

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

Using Lemma 6 and 7, we obtain:

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

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

Applying now Lemma 2, we obtain:

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

This requires the constraint $\mathbb{E}_\vartheta \Psi_i^3 = \mathbb{E}_\vartheta \Gamma_i^{-3/2} < +\infty$, *i.e.* $a > 3/2$.
We proceed analogously for R_4 and find that $R_4 = O_P(\sqrt{N/n})$ for $a > 2$. $\square$

The second order derivatives of $-N^{-1}\mathbf{U}_{N,n}(\vartheta)$ are as follows. First, the derivatives corresponding to (λ, a):

$$\begin{aligned}
-\frac{1}{N}\frac{\partial^2}{\partial\lambda^2}\mathbf{U}_{N,n}(\vartheta) &= \frac{a}{\lambda^2} - \frac{1}{N(a+(n/2))}\sum_{i=1}^N Z_i^{-2} 1_{F_i}, \\
-\frac{1}{N}\frac{\partial^2}{\partial\lambda\partial a}\mathbf{U}_{N,n}(\vartheta) &= -\frac{1}{\lambda} + \frac{1}{N(a+(n/2))}\sum_{i=1}^N Z_i^{-1} 1_{F_i}, \\
-\frac{1}{N}\frac{\partial^2}{\partial a^2}\mathbf{U}_{N,n}(\vartheta) &= \psi'(a) - \psi'(a+n/2) + \frac{1}{N(a+n/2))}\sum_{i=1}^N 1_{F_i^c}.
\end{aligned}$$

Then, the cross derivatives w.r.t. (λ, a), (μ, ω^2):

$$\begin{aligned}
-\frac{1}{N}\frac{\partial^2}{\partial\lambda\partial\mu}\mathbf{U}_{N,n}(\vartheta) &= \frac{1}{N(a+(n/2))}\sum_{i=1}^N 1_{F_i} Z_i^{-2} A_i, \\
-\frac{1}{N}\frac{\partial^2}{\partial\lambda\partial\omega^2}\mathbf{U}_{N,n}(\vartheta) &= \frac{1}{2N(a+(n/2))}\sum_{i=1}^N 1_{F_i} Z_i^{-2} A_i^2, \\
-\frac{1}{N}\frac{\partial^2}{\partial a\partial\mu}\mathbf{U}_{N,n}(\vartheta) &= -\frac{1}{N(a+(n/2))}\sum_{i=1}^N 1_{F_i} Z_i^{-1} A_i, \\
-\frac{\partial^2}{\partial a\partial\omega^2}\mathbf{U}_{N,n}(\vartheta) &= -\frac{1}{2N(a+(n/2))}\sum_{i=1}^N 1_{F_i} Z_i^{-1} A_i^2.
\end{aligned}$$

Then, the derivatives corresponding to (μ, ω^2):

$$\begin{aligned}
-\frac{1}{N}\frac{\partial^2}{\partial\mu^2}\mathbf{U}_{N,n}(\vartheta) &= \frac{1}{N}\sum_{i=1}^N 1_{F_i} Z_i^{-1} B_i - \frac{1}{N(a+(n/2))}\sum_{i=1}^N 1_{F_i} Z_i^{-2} A_i^2, \\
-\frac{1}{N}\frac{\partial^2}{\partial\mu\partial\omega^2}\mathbf{U}_{N,n}(\vartheta) &= \frac{1}{N}\sum_{i=1}^N 1_{F_i} Z_i^{-1} A_i B_i - \frac{1}{N(a+(n/2))}\sum_{i=1}^N 1_{F_i} Z_i^{-2} A_i^3, \\
-\frac{1}{N}\sum_{i=1}^N 1\frac{\partial^2}{\partial\omega^2\partial\omega^2}\mathbf{U}_{N,n}(\vartheta) &= \frac{1}{2N}\sum_{i=1}^N \left(1_{F_i} 2 Z_i^{-1} A_i^2 B_i - B_i^2\right) - \frac{1}{4N(a+(n/2))}\sum_{i=1}^N 1_{F_i} Z_i^{-2} A_i^4.
\end{aligned}$$

By applying Lemmas 7 and 2, we can prove that the terms of the form $N^{-1}\sum_{i=1}^N 1_{F_i} Z_i^{-\alpha} A_i^\beta B_i^\delta$ converge, as N, n tend to infinity, to $\mathbb{E}_\vartheta \Gamma_1^\alpha A_1^\beta(T;\mu,\omega^2) B_1^\delta(T;\omega^2)$. This implies that all terms of the form $N^{-1}(a + (n/2))^{-1}\sum_{i=1}^N 1_{F_i} Z_i^{-\alpha} A_i^\beta B_i^\delta$ converge to 0. So, all cross derivatives tend to 0 and the rest of derivatives gives the limit $\mathcal{J}(\vartheta)$. However, this requires to evaluate the moment conditions implied by Lemmas 7 and 2. The worst term is the last one above and yields that $a > 6$. $\square$

7.4 Proof of Proposition 2

We only give a sketch of the proof and assume $d = 1$ for simplicity. We compute $\mathcal{H}_{N,n}(\vartheta)$ (see (30)) and set $G_{i,n} = \{S_{i,n} \geq k\sqrt{n}\}$. We have:

$$
\begin{aligned}
\frac{\partial \mathbf{V}^{(1)}_{N,n}}{\partial \lambda}(\lambda, a) &= \sum_{i=1}^{N} \left(\frac{a}{\lambda} - \xi_{i,n}^{-1} \right), \\
\frac{\partial \mathbf{V}^{(1)}_{N,n}}{\partial a}(\lambda, a) &= N\left(\psi(a+n/2) - \log(a+n/2) - \psi(a) + \log\lambda\right) - \sum_{i=1}^{N} \log \xi_{i,n}, \\
\frac{\partial \mathbf{W}_{N,n}}{\partial \mu}(\mu, \omega^2) &= \sum_{i=1}^{N} 1_{G_{i,n}}\ \xi_{i,n}^{-1} A_{i,n}, \\
\frac{\partial \mathbf{W}_{N,n}}{\partial \omega^2}(\mu, \omega^2) &= \frac{1}{2}\sum_{i=1}^{N} \left(1_{G_{i,n}} \xi_{i,n}^{-1} A_{i,n}^2 - B_{i,n} \right).
\end{aligned}
$$

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:

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

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:

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

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

8 Auxiliary results

8.1 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 part, $-a\psi'(a) = -1 - \int_0^1 t^a g(t)dt$, where $g(t) = (\log t/(1-t))'$. A simple study yields that $t^a g(t)$ integrable on $(0,1)$ and positive except at $t = 1$. Thus, $1 - a\psi'(a) \neq 0$.

The following asymptotic expansions as a tends to infinity hold:

$$\log\Gamma(a) = (a-\frac{1}{2})\log a - a + \frac{1}{2}\log 2\pi + O(\frac{1}{a}), \tag{62}$$

$$\psi(a) = \log a - \frac{1}{2a} + O(\frac{1}{a^2}), \ \ \psi'(a) = \frac{1}{a} + O(\frac{1}{a^2}). \tag{63}$$

The following results are classical.
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, $\mathbb{E}\log(\lambda X) = \psi(a)$, 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 $n \to +\infty$,

$$\mathbb{E}\left(\frac{X}{n}\right)^p = O(1), \qquad \mathbb{E}\left(\frac{n}{X}\right)^p = O(1). \tag{64}$$

Using the Rosenthal inequality, for all $p \geq 2$

$$\mathbb{E}(\frac{X}{n}-1)^p \leq c_p n^{-p}\left(n\mathbb{E}|\varepsilon_i^2 - 1| + (n\mathbb{E}(\varepsilon_i^2-1)^2)^{p/2}\right) \lesssim O(\frac{1}{n^{p/2}}), \tag{65}$$

and for $n > 4p$,

$$\mathbb{E}(\frac{n}{X}-1)^p \leq \left(\mathbb{E}\left(\frac{n}{X}\right)^{2p}\mathbb{E}\left(\frac{X}{n}-1\right)^{2p}\right)^{1/2} \lesssim O(\frac{1}{n^{p/2}}). \tag{66}$$

8.2 Approximation results for discretizations.

The following lemmas are proved in Delattre et al. (2016). In the first two lemmas, we set $X_1(t) = X(t), \Phi_1 = \Phi, \Psi_1 = \Psi$.

Lemma 5. *Under (H1)-(H2), for $s \leq t$ and $t - s \leq 1$, $p \geq 1$,*

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

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 6. *Assume (H1)-(H2) and $p \geq 1$. 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,*

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

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

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

Lemma 7. *Recall notations (23)-(24). 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 8. *Then, for all $p \geq 1$,*

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

8.3 Direct observation of the random effects

Assume that a sample $(\Phi_i, \Gamma_i), i = 1, \ldots, N$ is observed and that $d = 1$ for simplicity. The Gamma distribution with parameters (a, λ) $(a > 0, \lambda > 0)$ $G(a, \lambda)$, has density $\gamma_{a,\lambda}(x) = (\lambda^a/\Gamma(a))x^{a-1}e^{-x}\mathbb{1}_{(0,+\infty)}(x)$, where $\Gamma(a)$ is the Gamma fonction. We set $\psi(a) = \Gamma'(a)/\Gamma(a)$. The log-likelihood $\ell_N(\vartheta)$ of the N-sample $(\Phi_i, \Gamma_i), i = 1, \ldots, N$ 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

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

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

$$\mathcal{J}_0(\vartheta) = \left(\begin{array}{c|c} I_0(\lambda, a) & \mathbf{0} \\ \hline \mathbf{0} & J_0(\lambda, a, \mu, \omega^2) \end{array} \right),$$

$$I_0(\lambda, a) = \begin{pmatrix} \frac{a}{\lambda^2} & -\frac{1}{\lambda} \\ -\frac{1}{\lambda} & \psi'(a) \end{pmatrix}, \quad J_0(\lambda, a, \mu, \omega^2) = \begin{pmatrix} \frac{a}{\lambda\omega^2} & 0 \\ 0 & \frac{1}{2\omega^4} \end{pmatrix}. \tag{67}$$

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

$$I_0^{-1}(\lambda, a) = \frac{1}{a\psi'(a) - 1} \begin{pmatrix} \lambda^2\psi'(a) & \lambda\psi'(a) \\ \lambda\psi'(a) & a \end{pmatrix}. \tag{68}$$

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 $\hat{m} = N^{-1}\sum_{i=1}^N \Gamma_i,\ \hat{t} = N^{-1}\sum_{i=1}^N \log\Gamma_i$ which are asymptotically Gaussian with limiting covariance matrix

$$\frac{1}{N}\begin{pmatrix} \frac{a}{\lambda^2} & \frac{1}{\lambda} \\ \frac{1}{\lambda} & \psi'(a) \end{pmatrix}. \tag{69}$$

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

Let us stress that the marginal distribution of Φ_i is not Gaussian. It is a translated rescaled Student distribution. Indeed we can write $\Phi_i = \mu + \omega\Gamma_i^{-1/2}\varepsilon_i$ with $\varepsilon_i \sim \mathcal{N}(0,1)$ independent of Γ_i. In particular, $\mathbb{E}|\Phi_i|^p < +\infty$ if and only if $a > p/2$.
Note also that Γ_i converges in probability to the constant m if both $a = a_k$ and $\lambda = \lambda_k$ tend to infinity with k at the same rate, *e.g.* $a = mk, \lambda = k$.