ESTIMATION OF POPULATION PARAMETERS IN STOCHASTIC DIFFERENTIAL EQUATIONS WITH RANDOM EFFECTS IN THE DIFFUSION COEFFICIENT

Maud Delattre[1], Valentine Genon-Catalot[2] and Adeline Samson[3]

Abstract. We consider N independent stochastic processes $(X_i(t), t \in [0, T])$, $i = 1, \ldots, N$, defined by a stochastic differential equation with diffusion coefficients depending linearly on a random variable ϕ_i. The distribution of the random effect ϕ_i depends on unknown population parameters which are to be estimated from a discrete observation of the processes (X_i). The likelihood generally does not have any closed form expression. Two estimation methods are proposed: one based on the Euler approximation of the likelihood and another based on estimations of the random effects. When the distribution of the random effects is Gamma, the asymptotic properties of the estimators are derived when both N and the number of observations per component X_i tend to infinity. The estimators are computed on simulated data for several models and show good performances.

Mathematics Subject Classification. 62F10, 62F12.

Received March 2, 2014. Revised December 6, 2014.

1. Introduction

Stochastic differential equations (SDEs) with random effects have been the subject of several recent contributions, with various applications such as pharmacokinetic/pharmacodynamic, neuronal modeling [2,5,10]. Several estimation methods have been proposed to provide estimators in these complex models. For general mixed effects models (not only SDE), the main reference for a theoretical study of the exact maximum likelihood is Nie [9]. But the likelihood has no explicit expression except in some special cases and assumptions are then very difficult to verify. In Delattre *et al.* [3], the case of a linear random effect in the drift together with a specific distribution for the random effects is investigated. In this case, the exact maximum likelihood estimator is explicit and studied. In the general case, Picchini *et al.* [10]; Picchini and Ditlevsen [11] propose approximations of the likelihood based on Hermite expansion and Gaussian quadrature. All these references work with random effects in the drift, and not in the diffusion coefficient, except Delattre and Lavielle [2] who incorporate measurement error and propose an approximation of the likelihood with the extended Kalman filter.

[1] AgroParisTech, UMR518, 75005 Paris, France. maud.delattre@agroparistech.fr

[2] UMR CNRS 8145, Laboratoire MAP5, Université Paris Descartes, Sorbonne Paris Cité, France. valentine.genon-catalot@parisdescartes.fr

[3] Laboratoire Jean Kuntzmann, UMR CNRS 5224, Université Grenoble-Alpes, France. adeline.leclercq-samson@imag.fr

Here, we focus on discretely observed SDEs with a random effect in the diffusion coefficient. The distribution of the random effect depends on unknown parameters to be estimated. Applications of these models have been considered in population pharmacokinetics or pharmacodynamic where the diffusion coefficient represents the intra-individual variability, which may vary from one individual to another, thus being modeled as a random effect [2]. We assume that there is a linear random effect in the diffusion coefficient. For simplicity, we assume that the drift is zero in Sections 2 and 3, and give the extension to the case of non null drift in Section 4.

Statistical inference for discretely observed SDEs with no random effects has been widely studied (see [8], and references therein). In Genon-Catalot and Jacod [6] the estimation of unknown fixed parameters in the diffusion coefficient is studied with discrete observations of a single trajectory when the sampling interval tends to zero. The likelihood of these observations is not explicit, therefore estimating equations are built based on the Euler approximation of the SDE with drift set to zero. One of the strategies described below follows the same idea, but here the parameters are random. This complicates the definition and the theoretical study of the estimator.

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 SDEs:

$$\mathrm{d}\, X_i(t) = \phi_i\, \sigma(X_i(t))\, \mathrm{d}W_i(t), \quad X_i(0) = x_i^0,\, i = 1, \ldots, N, \tag{1.1}$$

where $(W_i)_{1 \leq i \leq N}$ are N independent Wiener processes, $(\phi_i)_{1 \leq i \leq N}$ are N *i.i.d.* random variables taking values in $(0, +\infty)$, $(\phi_i)_{1 \leq i \leq N}$ and $(W_i)_{1 \leq i \leq N}$ are independent. The function $\sigma(x)$ is known and real-valued. Each process $(X_i(t))$ represents an individual, the variable ϕ_i represents the random effect of individual i. The variables $(\phi_i)_{1 \leq i \leq N}$ have a common distribution $g(\varphi, \theta)\mathrm{d}\nu(\varphi)$ on $(0, +\infty)$ where ν is a dominating measure and θ is a vector of unknown parameters called population parameters, belonging to a set $\Theta \subset \mathbb{R}^p$.

Our aim is to estimate θ from discrete observations $\{X_i(t_j), j = 1, \ldots, n, i = 1, \ldots, N\}$ where $t_j = jT/n$ and T is fixed. In the case of a linear random effect in the diffusion coefficient (1.1), choosing an inverse Gamma distribution leads to explicit estimators. Therefore, we consider the specific case

$$\phi_i = \frac{1}{\Gamma_i^{1/2}} \quad \text{with} \quad \Gamma_i \sim G(a, \lambda), \quad a > 0, \lambda > 0, \quad \theta = (a, \lambda). \tag{1.2}$$

We study the exact maximum likelihood estimator in the case $\sigma(.) \equiv 1$. When $\sigma(.) \not\equiv 1$, we build estimating equations based on the Euler's approximation of the fixed effect diffusion model. The difficulty of these estimating equations is that the Euler's approximation has to be integrated out with respect to the distribution of the random effects. Moreover, we build another type of estimating equations, corresponding to the ideal likelihood of directly observed random effects where estimators of the random effects are plugged in. This second approach has the advantage to be easily generalized to any distribution for the random effects.

The paper is organized as follows. Section 2 introduces some assumptions and gives the exact likelihood and its approximation obtained by Euler's scheme. Our asymptotic framework is when the number N of subjects tends to infinity. In Section 3, we study the asymptotic properties of the estimators. When $\sigma(.) \equiv 1$, the exact maximum likelihood estimator of θ is asymptotically Gaussian with rate $\sqrt{N}$ both for fixed number of measurements per subject n and for n tending to infinity. When $\sigma(.) \not\equiv 1$, we must assume that n depends on N and satisfies the constraint $N/n \to 0$ for the first method, $\sqrt{N}/n \to 0$ for the second. Our estimators are asymptotically Gaussian with rate $\sqrt{N}$. Section 4 deals with a non nul drift. Simulations illustrate the behavior of the estimators and results are presented in Section 5. Section 6 concludes the paper with some extensions. Proofs are gathered in Appendix.

2. Exact and approximate likelihoods

Consider N real valued stochastic processes $(X_i(t), t \geq 0)$, $i = 1, \ldots, N$, with dynamics ruled by (1.1). The processes $(W_i)_{1 \leq i \leq N}$ and the r.v.'s $(\phi_i)_{1 \leq i \leq N}$ are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Consider the filtration $(\mathcal{F}_t, t \geq 0)$ defined by $\mathcal{F}_t = \sigma(\phi_i, W_i(s), s \leq t, i = 1, \ldots, N)$. As $\mathcal{F}_t = \sigma(W_i(s), s \leq t) \vee \mathcal{F}_t^i$,

with $\mathcal{F}_t^i = \sigma(\phi_i, \phi_j, W_j(s), s \leq t, j \neq i)$ independent of W_i, each process W_i is a $(\mathcal{F}_t, t \geq 0)$-Brownian motion. Moreover, the random variables ϕ_i are $\mathcal{F}_0$-measurable. In what follows, we assume that

(H1) The function σ belongs to $C^2(\mathbb{R})$ and for all $x \in \mathbb{R}$, $0 < \sigma_0^2 \leq \sigma^2(x) \leq \sigma_1^2$, $|\sigma'(x)| + |\sigma''(x)| \leq K$.

Under (H1), the process $(X_i(t))$ is well-defined and $(\phi_i, X_i(t))$ is strong Markov adapted to the filtration $(\mathcal{F}_t, t \geq 0)$. The N processes $(\phi_i, X_i(.))_{1 \leq i \leq N}$ are independent. For all φ, and all $x_i^0 \in \mathbb{R}$, the fixed effect SDE

$$d\,X_i^{\varphi, x_i^0}(t) = \varphi\,\sigma(X_i^{\varphi, x_i^0}(t))\,dW_i(t), \quad X_i^{\varphi, x_i^0}(0) = x_i^0 \tag{2.1}$$

admits a unique strong solution process $(X_i^{\varphi, x_i^0}(t), t \geq 0)$ adapted to the filtration $(\mathcal{F}_t, t \geq 0)$. From the Markov property of $(\phi_i, X_i(t))$, we deduce that the conditional distribution of X_i given $\phi_i = \varphi$ is identical to the distribution of X_i^{φ, x_i^0} (for more details, see [3]).

For $i = 1, \ldots, N$, the process $(X_i(t), t \in [0, T])$ is discretely observed at times $t_j = jT/n, j = 0, \ldots, n$ and we set for fixed T:

$$\Delta = \frac{T}{n}, \qquad X_i = (X_i(t_j), j = 1, \ldots, n), \quad i = 1, \ldots, N. \tag{2.2}$$

We start with the exact likelihood of (2.2). The distribution of the observations $(X_i)_{1 \leq i \leq N}$ on $\prod_{i=1}^N \mathbb{R}^n$ has the form $P_\theta = \otimes_{i=1}^N P_\theta^i$ where P_θ^i is the distribution of X_i on $\mathbb{R}^n$. If $Q_{\varphi, x_i^0}^i$ denotes the distribution of $X_i^\varphi = (X_i^{\varphi, x_i^0}(t_j), j = 1, \ldots, n)$ and $p_t(x, y, \varphi)$ the transition density of (2.1), then $Q_{\varphi, x_i^0}^i$ admits the density $\prod_{j=1}^n p_\Delta(x_{i,j-1}, x_{i,j}, \varphi)$ w.r.t. the Lebesgue's measure of $\mathbb{R}^n$ (with $x_{i,0} = x_i^0$). Therefore, the density of P_θ^i w.r.t. the Lebesgue's measure of $\mathbb{R}^n$ is given by:

$$\lambda_i(\theta, \mathbf{x_i}) = \int_0^{+\infty} g(\varphi, \theta) \prod_{j=1}^n p_\Delta(x_{i,j-1}, x_{i,j}, \varphi)\,d\nu(\varphi),$$

where $\mathbf{x_i} = (x_{i,j}, j = 1, \ldots, n)$. The exact likelihood is $\Lambda_N(\theta) = \prod_{i=1}^N \lambda_i(\theta, \mathbf{x_i})$. Here, we are faced with two problems. First, the transition densities of (2.1) are generally not explicit. Second, and this is specific to random effect models (SDE or more generally non Gaussian models), even if these transition densities were explicit, it is generally not possible to get a closed-form expression for the marginal density of X_i, which corresponds to the integral $\lambda_i(\theta, \mathbf{x_i})$. Therefore the exact likelihood is not explicit and difficult to study theoretically and numerically.

Instead of using the exact transition densities of (2.1), it is now standard to use the approximation given by the transition densities of the corresponding Euler's scheme, *i.e.* the one-step discretisation of (2.1) (see *e.g.* [4,6,8]), although for small n, it may give a bias. Therefore, we introduce

$$\tilde{L}_i(X_i, \varphi) = \frac{1}{\varphi^n \prod_{j=1}^n \sigma(X_i(t_{j-1}))} \exp\left(-\frac{S_i}{2\varphi^2}\right) \propto \frac{1}{\varphi^n} \exp\left(-\frac{S_i}{2\varphi^2}\right),$$

with, for $i = 1, \ldots, N$,

$$S_i = \frac{1}{\Delta} \sum_{j=1}^n \frac{(X_i(t_j) - X_i(t_{j-1}))^2}{\sigma^2(X_i(t_{j-1}))}. \tag{2.3}$$

Note that when $\sigma(.) \equiv 1$, $\tilde{L}_i(X_i, \varphi)$ is the exact density of $(X_i^{\varphi, x_i^0}(t_j), j = 1, \ldots, n)$ (see (2.1)). To estimate θ, instead of the exact likelihood, we introduce the approximate likelihood, corresponding to the Euler's scheme integrated with respect to the random effects distribution:

$$\tilde{\Lambda}_N(\theta) = \prod_{i=1}^N \int_0^{+\infty} \varphi^{-n} \exp\left(-\frac{S_i}{2\varphi^2}\right) g(\varphi, \theta)\,d\nu(\varphi). \tag{2.4}$$

When $\sigma(.) \equiv 1$, the problem reduces to $S_i/n = \phi_i^2 \frac{1}{n} \sum_{j=1}^n \frac{(W_i(t_j) - W_i(t_{j-1}))^2}{\Delta}$ and $\frac{1}{n\Delta} \sum_{j=1}^n (W_i(t_j) - W_i(t_{j-1}))^2$ has a distribution $\chi^2(n)/n$, and is independent of ϕ_i. Thus the model can be seen as an observation of ϕ_i with a multiplicative error. In both cases of $\sigma(.) \equiv 1$ and $\sigma(.) \not\equiv 1$, a theoretical study of the estimators based on $\tilde{\Lambda}_N(\theta)$ could be possible using the approach developed by [9] but his assumptions are generally difficult to verify. Below, as in Delattre *et al.* [3], we rather introduce a specific distribution for the random effects allowing to obtain an explicit formula for (2.4). In Section 3, we are able to directly study the corresponding estimators.

Remark 2.1. Except in the case $\sigma(.) \equiv 1$ where (2.4) is the exact likelihood, our approach based on an approximate likelihood imposes a double asymptotic framework where both N and n tend to infinity. As $n \to \infty$, note that the statistic S_i based on the ith trajectory provides an estimator of the random effect ϕ_i^2. Indeed, let $M_i(t) = \int_0^t \sigma(X_i(s)) \mathrm{d}W_i(s)$, and

$$R_i = \sum_{j=1}^n (M_i(t_j) - M_i(t_{j-1}))^2 / \sigma^2(X_i(t_{j-1})). \tag{2.5}$$

By standard properties of quadratic variations, $R_i/T \to 1$ in probability as $n \to \infty$. Thus, using $X_i(t_j) = X_i(t_{j-1}) + \int_{t_{j-1}}^{t_j} \phi_i \sigma(X_i(s)) \mathrm{d}W_i(s)$, $S_i/n = \phi_i^2 R_i/T$ tends to ϕ_i^2.

3. A specific distribution for the random effect

For a general distribution $g(\varphi, \theta) \mathrm{d}\nu(\varphi)$ of the random effect ϕ_i, the integral in (2.4) has no explicit expression. However, for the conjugate distribution, namely the inverse Gamma (1.2), an explicit expression is obtained. The unknown parameter is then $\theta = (a, \lambda) \in \Theta = \mathbb{R}^+ \times \mathbb{R}^+$. The true value is denoted by θ_0.

Let us start with the ideal case of directly observed random effects ϕ_i (or Γ_i). Then, the exact log-likelihood of $(\Gamma_1, \ldots, \Gamma_N)$ is given by:

$$\ell_N(\theta) = Na \log \lambda - N \log \Gamma(a) + (a-1) \sum_{i=1}^N \log \Gamma_i - \lambda \sum_{i=1}^N \Gamma_i \tag{3.1}$$

with associated score function $\mathcal{S}_N(\theta) = \left(\frac{\partial}{\partial \lambda} \ell_N(\theta) \ \ \frac{\partial}{\partial a} \ell_N(\theta) \right)'$ where

$$\frac{\partial}{\partial \lambda} \ell_N(\theta) = \sum_{i=1}^N \left(\frac{a}{\lambda} - \Gamma_i \right), \frac{\partial}{\partial a} \ell_N(\theta) = \sum_{i=1}^N (-\psi(a) + \log \lambda + \log \Gamma_i),$$

where $\psi(z) = \frac{\Gamma'(z)}{\Gamma(z)}$ is the di-gamma function. By standard properties of Gamma distributions, we have, under the true value θ_0, $(1/\sqrt{N}) \mathcal{S}_N(\theta_0) \to_{\mathcal{D}} \mathcal{N}_2(0, \mathcal{I}(\theta_0))$, where $\mathcal{I}(\theta)$ is

$$\mathcal{I}(\theta) = \begin{pmatrix} \frac{a}{\lambda^2} & -\frac{1}{\lambda} \\ -\frac{1}{\lambda} & \psi'(a) \end{pmatrix}. \tag{3.2}$$

Note that using properties of the di-gamma function (see Sect. 6), $\mathcal{I}(\theta)$ is invertible for all $\theta \in (0, +\infty)^2$. The maximum likelihood estimator based on the observation of $\Gamma_1, \ldots, \Gamma_N$, denoted $\theta_N = \theta_N(\Gamma_1, \ldots, \Gamma_N)$ is consistent and satisfies $\sqrt{N}(\theta_N - \theta_0) \to_{\mathcal{D}} \mathcal{N}_2(0, \mathcal{I}^{-1}(\theta_0))$ as N tends to infinity.

But the Γ_i's are not observed. Two different strategies are studied. Following Remark 2.1, a natural idea consists in plugging in $\ell_N(\theta)$ the estimator n/S_i of Γ_i. This reveals to be more complex than expected (Sect. 3.2) and we will need to truncate the estimator n/S_i. The other strategy (Sect. 3.1) is based on (2.4). We provide asymptotic results when n is fixed and $N \to \infty$ in the case $\sigma(.) \equiv 1$, and when both $n, N \to \infty$ for a general $\sigma(.)$.

3.1. Estimation based on the Euler approximation of the likelihood

Let $\tilde{\mathcal{L}}_N(\theta) = \log \tilde{\Lambda}_N(\theta)$ be the log Euler contrast (see (2.3) and (2.4)).

Proposition 3.1. *Under* (H1) *and* (1.2), *we have:*

$$\tilde{\mathcal{L}}_N(\theta) = \sum_{i=1}^{N} \log \left(\frac{\Gamma(a+n/2)}{\Gamma(a)} \right) + aN \log \lambda - \sum_{i=1}^{N} (a+n/2) \log \left(\lambda + \frac{1}{2} S_i \right). \tag{3.3}$$

The associated gradient vector (pseudo-score function)

$$\tilde{G}_N(\theta) = \left(\frac{\partial}{\partial \lambda} \tilde{\mathcal{L}}_N(\theta) \ \ \frac{\partial}{\partial a} \tilde{\mathcal{L}}_N(\theta) \right)' \tag{3.4}$$

is given by $\frac{\partial}{\partial \lambda} \tilde{\mathcal{L}}_N(\theta) = \sum_{i=1}^{N} \left(\frac{a}{\lambda} - \frac{a+n/2}{\lambda+S_i/2} \right)$ and $\frac{\partial}{\partial a} \tilde{\mathcal{L}}_N(\theta) = \sum_{i=1}^{N} (\psi(a+n/2) - \psi(a)) + \sum_{i=1}^{N} \log \left(\frac{\lambda}{\lambda+S_i/2} \right)$.
For the Hessian matrix (pseudo Fisher information matrix)

$$\widetilde{\mathcal{I}}_N(\theta) = - \left(\begin{array}{cc} \frac{\partial^2}{\partial \lambda^2} \tilde{\mathcal{L}}_N(\theta) & \frac{\partial^2}{\partial \lambda \partial a} \tilde{\mathcal{L}}_N(\theta) \\ \frac{\partial^2}{\partial \lambda \partial a} \tilde{\mathcal{L}}_N(\theta) & \frac{\partial^2}{\partial a^2} \tilde{\mathcal{L}}_N(\theta) \end{array} \right), \tag{3.5}$$

we get $\frac{\partial^2}{\partial \lambda^2} \tilde{\mathcal{L}}_N(\theta) = -\sum_{i=1}^{N} \left(\frac{a}{\lambda^2} - \frac{a+n/2}{(\lambda+S_i/2)^2} \right)$, $\frac{\partial^2}{\partial \lambda \partial a} \tilde{\mathcal{L}}_N(\theta) = \sum_{i=1}^{N} \left(\frac{1}{\lambda} - \frac{1}{\lambda+S_i/2} \right)$ and
$\frac{\partial^2}{\partial a^2} \tilde{\mathcal{L}}_N(\theta) = -N (\psi'(a) - \psi'(a+n/2))$. We study the estimators defined by the estimating equation:

$$\tilde{G}_N(\tilde{\theta}_N) = 0. \tag{3.6}$$

We consider two asymptotics: n fixed (Sect. 3.1.1) and $n \to \infty$ (Sect. 3.1.2).

3.1.1. Fixed number of observations per subject

We assume that the number n of observations per subject is fixed and that the number of subjects N tends to infinity. The only model that can be studied with this asymptotic is the special case $\sigma(.) \equiv 1$. We denote by the upper index 1 all the quantities associated to this model: $dX_i^1(t) = \phi_i dW_i(t)$, and the statistic is

$$S_i^1 = \sum_{j=1}^{n} (X_i^1(t_j) - X_i^1(t_{j-1}))^2 / \Delta. \tag{3.7}$$

The distribution of S_i^1 can be explicitly computed.

Proposition 3.2. *Under* P_θ, *the random variables* $\beta_i^1(\lambda) = \frac{\lambda}{\lambda+S_i^1/2}, i = 1, \ldots, N$, *are independent and* $\beta_i^1(\lambda)$ *has distribution beta of the first kind on* $(0,1)$ *with parameters* $(a, n/2)$. *The random variables* $S_i^1/(2\lambda)$ *are independent with distribution on* $(0, +\infty)$ *beta of the second kind with parameters* $(n/2, a)$.

Then, $\tilde{\mathcal{L}}_N(\theta) = \tilde{\mathcal{L}}_N^1(\theta)$ where S_i is replaced by S_i^1 in the exact log-likelihood. Define the associated exact maximum likelihood estimator as any solution of:

$$\hat{\theta}_N^1 = \text{Argsup}_\theta \ \tilde{\mathcal{L}}_N^1(\theta). \tag{3.8}$$

Proposition 3.3. *Assume that* n *is fixed. Then, the maximum likelihood estimator* $\hat{\theta}_N^1$ (3.8) *is consistent. Let*

$$\mathcal{I}_n(\theta) = \left(\begin{array}{cc} \frac{a(n/2)}{\lambda^2(a+1+n/2)} & -\frac{n/2}{\lambda(a+n/2)} \\ -\frac{n/2}{\lambda(a+n/2)} & \psi'(a) - \psi'(a+n/2) \end{array} \right). \tag{3.9}$$

Then, the matrix $\mathcal{I}_n(\theta_0)$ *is invertible and under* P_{θ_0} $\sqrt{N}(\hat{\theta}_N^1 - \theta_0) \to_{\mathcal{D}} \mathcal{N}_2(0, \mathcal{I}_n^{-1}(\theta_0))$.

Remark that $\mathcal{I}_n(\theta) = \mathcal{I}(\theta) + O(\frac{1}{n})$. The proof of Proposition 3.3 is standard and omitted. It is simply the asymptotic study of the maximum likelihood estimator based on the *i.i.d.* sample $(S_i^1, i = 1, \ldots, N)$ whose distribution is specified by Proposition 3.2.

3.1.2. Number of observations per subject goes to infinity

Now, we assume that both n and N tend to infinity for some fixed T. The strategy consists in studying the case $\sigma(.) \equiv 1$ where computations can all be done explicitly and then studying the difference between the general case and the case $\sigma(.) \equiv 1$. Some preliminary results on the moments of R_i/T are needed (Prop. 3.4). For these results specifically, we do not assume that the ϕ_i^2's have inverse Gamma distribution. As already said before, $S_i/n = \phi_i^2 R_i/T$ tends to ϕ_i^2 as n tends to infinity (see (2.5) for the definition of R_i). Let us define the equivalent of R_i for the model $\sigma(.) \equiv 1$: $R_i^1 = \sum_{j=1}^n (W_i(t_j) - W_i(t_{j-1}))^2$.

We denote by $\mathbb{P}_\theta = \otimes_{i \geq 1} \mathbb{P}_\theta^i$ the distribution of the sequence of processes $(\phi_i, (X_i(t), t \in [0, T])), i \geq 1$ on $(0, +\infty) \times \prod_{i \geq 1} C([0, T])$, by $\mathbb{E}_\theta$ the corresponding expectation. Note that P_θ is the marginal distribution of $(X_i, i = 1, \ldots, N)$ under $\mathbb{P}_\theta$.

Both R_i/T and R_i^1/T tend to 1 in probability as $n \to \infty$. Furthermore:

Proposition 3.4. *Under* (H1), *for all θ, we have* $\mathbb{E}_\theta(\frac{R_i^1}{T} - 1|\phi_i) = 0$, $|\mathbb{E}_\theta(\frac{R_i}{T} - 1|\phi_i)| \leq C\frac{T}{n}\phi_i^2$, *and* $|\mathbb{E}_\theta(\frac{R_i}{T} - \frac{R_i^1}{T}|\phi_i)| \leq C\frac{T}{n}\phi_i^2$, *and for all $p \geq 1$,* $\mathbb{E}_\theta((\frac{R_i}{T} - 1)^{2p}|\phi_i) \leq C(\frac{T^p}{n^p} + \frac{T^{2p}}{n^{2p}}(\phi_i^{2p} + \phi_i^{4p}))$ *and* $\mathbb{E}_\theta((\frac{R_i}{T} - \frac{R_i^1}{T})^{2p}|\phi_i) \leq C\frac{T^{2p}}{n^{2p}}(\phi_i^{2p} + \phi_i^{4p})$.

We now study the score function (3.4) and the Fisher information matrix (3.5).

Proposition 3.5. *Recall $\tilde{G}_N(\theta_0)$ defined by* (3.4) *and $\mathcal{I}(\theta_0)$ given in* (3.2).
For $\sigma(.) \equiv 1$, as N, n tend to infinity, under P_{θ_0}, $\tilde{G}_N(\theta_0)/\sqrt{N}$ converges in distribution to $\mathcal{N}_2(0, \mathcal{I}(\theta_0))$.
In the general case, if $\mathbb{E}_{\theta_0}\phi_i^8 < +\infty$, i.e. if $a_0 > 4$ and N, n tend to infinity in such a way that N/n tends to 0, the same result holds.

The convergence of the Fisher information matrix is as follows:

Proposition 3.6. *In the case $\sigma(.) \equiv 1$ and the general case, the pseudo Fisher information matrix given in* (3.5), *$\widetilde{\mathcal{I}}_N(\theta_0)/N$, converges in probability to $\mathcal{I}(\theta_0)$ (see (3.2)) as N, n tend to infinity, under P_{θ_0}.*

Now we study the estimator $\tilde{\theta}_N$ defined by (3.6).

Proposition 3.7. *Assume that $n, N \to +\infty$ in such a way that N/n tends to 0. Then, an estimator $\tilde{\theta}_N$ which solves* (3.6) *exists with probability tending to one as N tends to infinity under P_{θ_0} and is weakly consistent. The matrix $\mathcal{I}(\theta_0)$ is invertible and under P_{θ_0}, $\sqrt{N}(\tilde{\theta}_N - \theta_0) \to_\mathcal{D} \mathcal{N}_2(0, \mathcal{I}^{-1}(\theta_0))$.*
Moreover, the estimator $\tilde{\theta}_N$ is asymptotically equivalent to the MLE $\theta_N = \theta_N(\Gamma_1, \ldots, \Gamma_N)$ based on the direct observation of $(\Gamma_1, \ldots, \Gamma_N)$.

If we discretely observe N trajectories of a fixed effect SDE, *i.e.* $dX_i(t) = \varphi\sigma(X_i(t))dW_i(t)$, the rate of convergence for φ is $\sqrt{Nn}$ with the same constraint $N/n = o(1)$. This is because all the N discrete trajectories are used to estimate a unique parameter φ. In the case of a random effect SDE, the rate of convergence of θ is $\sqrt{N}$, which is the same as the one obtained when the random effects Γ_i are directly observed.

3.2. Approach based on estimators of the random effects

In this section, we exploit directly the fact that the random effect $\phi_i^2 = \Gamma_i^{-1}$ can be estimated using the vector $X_i = (X_i(t_j), j \leq n)$ by S_i/n. The idea is simply to replace the random variables Γ_i by their estimator n/S_i in the likelihood (3.1) of $(\Gamma_1, \ldots, \Gamma_N)$. But this works only when $\sigma(.) \equiv 1$. More precisely, in this case, let us set:

$$U_N(\theta) = Na\log\lambda - N\log\Gamma(a) + (a-1)\sum_{i=1}^N \log(n/S_i^1) - \lambda\sum_{i=1}^N (n/S_i^1).$$

and consider the estimators θ_N^* given by

$$\nabla U_N(\theta_N^*) = 0. \tag{3.10}$$

Otherwise, we must define estimators of $\log \Gamma_i$ and Γ_i which have appropriate moment properties. These properties are obtained by associating a truncation to the estimators $\log(n/S_i)$ and n/S_i. Let us define:

$$\widetilde{\log}\Gamma_i = \log(n/S_i)\mathbb{1}_{(S_i/n \geq k/\sqrt{n})}, \qquad \widetilde{\Gamma}_i = (n/S_i)\mathbb{1}_{(S_i/n \geq k/\sqrt{n})}.$$

where k is a constant. Note that, by the above definitions, $\widetilde{\log}\Gamma_i$ and $\widetilde{\Gamma}_i$ are set to 0 outside the set $(S_i/n \geq k/\sqrt{n})$ where S_i/n is not bounded from below. The following holds.

Lemma 3.8. *Assume that $\mathbb{E}_{\theta_0}\phi_i^8 < +\infty$, i.e. $a_0 > 4$ and that $n > 4$. Then,*

$$\mathbb{E}_{\theta_0}\left(\frac{n}{S_i}\mathbb{1}_{(S_i/n \geq k/\sqrt{n})} - \frac{n}{S_i^1}\right)^2 + \mathbb{E}_{\theta_0}\left(\log\frac{S_i}{n}\mathbb{1}_{(S_i/n \geq k/\sqrt{n})} - \log\frac{S_i^1}{n}\right)^2 \leq \frac{C}{n^2}.$$

Then we consider the function

$$V_N(\theta) = Na\log\lambda - N\log\Gamma(a) + (a-1)\sum_{i=1}^N \widetilde{\log}\Gamma_i - \lambda\sum_{i=1}^N \widetilde{\Gamma}_i, \tag{3.11}$$

and the associated estimator θ_N^{**} defined by the estimating equation:

$$\nabla V_N(\theta_N^{**}) = 0. \tag{3.12}$$

Proposition 3.9. *Assume that $\sigma(.) \equiv 1$. If N, n tend to infinity in such a way that $\sqrt{N}/n$ tends to 0, then an estimator θ_N^* which solves (3.10) exists with probability tending to 1 under $\mathbb{P}_{\theta_0}$ and is weakly consistent. Moreover, $\sqrt{N}(\theta_N^* - \theta_0)$ converges in distribution to $\mathcal{N}_2(0, \mathcal{I}^{-1}(\theta_0))$ and θ_N^* is asymptotically equivalent to the exact MLE θ_N associated to $(\Gamma_1, \ldots, \Gamma_N)$, i.e. $\sqrt{N}(\theta_N^* - \theta_N) = o_{P_{\theta_0}}(1)$.*

When $\sigma(.)$ is not equal to 1, the same result holds for θ_N^{**} under the condition $\mathbb{E}_{\theta_0}\phi_i^8 < +\infty$, i.e. $a_0 > 4$.

Note that in this approach, even when $\sigma(.) \equiv 1$, the constraint $\sqrt{N}/n \to 0$ is required.

4. EXTENSION TO NONNUL DRIFT

The estimation results obtained above can be extended to more general SDEs with a drift term. Indeed, assume that the observed processes are given by:

$$\mathrm{d}\,X_i(t) = b(X_i(t))\mathrm{d}t + \phi_i\sigma(X_i(t))\,\mathrm{d}W_i(t), \quad X_i(0) = x_i^0,\ i = 1,\ldots,N,$$

where the drift function $b : \mathbb{R} \to \mathbb{R}$ may be known or unknown. We define, for $i = 1, \ldots, N$, S_i with the same formula (2.3)

$$S_i = \frac{1}{\Delta}\sum_{j=1}^n \frac{(X_i(t_j) - X_i(t_{j-1}))^2}{\sigma^2(X_i(t_{j-1}))}.$$

We consider the same estimation procedures using S_i. If b is bounded, it is easy to see that Proposition 3.4 holds.

The extension of our estimation results to the case of b or σ unbounded is not straightforward because ϕ_i is not bounded. Nevertheless, the simulation results for Examples 2−4 below show that good results are obtained even for unbounded functions b, σ.

TABLE 1. Example 1: $\sigma(.) \equiv 1$. Empirical mean and standard deviation (in brackets) of $\tilde{\theta}$ (method 1) and θ^* (method 2) computed from 100 datasets.

	$N = 50$		$N = 100$	
	$n = 100$	$n = 500$	$n = 100$	$n = 500$
	$(a_0 = 5, \lambda_0 = 3)$			
$\tilde{a}$	5.16 (1.33)	5.21 (1.17)	5.05 (0.75)	5.04 (0.64)
$\tilde{\lambda}$	3.12 (0.87)	3.14 (0.76)	3.06 (0.47)	3.05 (0.41)
a^*	4.70 (0.99)	5.10 (1.12)	4.60 (0.62)	4.94 (0.62)
λ^*	2.79 (0.64)	3.07 (0.73)	2.73 (0.39)	2.98 (0.39)
	$(a_0 = 6, \lambda_0 = 1)$			
$\tilde{a}$	6.11 (1.61)	6.09 (1.31)	5.94 (0.86)	5.93 (0.94)
$\tilde{\lambda}$	1.02 (0.29)	1.01 (0.23)	0.99 (0.15)	0.99 (0.14)
a^*	5.43 (1.31)	5.95 (1.25)	5.32 (0.68)	5.80 (0.79)
λ^*	0.89 (0.23)	0.98 (0.22)	0.87 (0.12)	0.96 (0.14)

TABLE 2. Example 2: $\sigma^2(x) = 1 + x^2$. Empirical mean and standard deviation (in brackets) of $\tilde{\theta}$ (method 1) and θ^{**} (method 2) computed from 100 datasets.

	$N = 50$		$N = 100$	
	$n = 100$	$n = 500$	$n = 100$	$n = 500$
	$(a_0 = 5, \lambda_0 = 3)$			
$\tilde{a}$	4.80 (1.16)	4.95 (0.99)	4.89 (0.81)	5.08 (0.74)
$\tilde{\lambda}$	2.94 (0.75)	2.99 (0.64)	2.96 (0.49)	3.04 (0.43)
a^{**}	4.35 (1.08)	4.85 (0.95)	4.47 (0.68)	4.99 (0.71)
λ^{**}	2.61 (0.68)	2.93 (0.61)	2.65 (0.40)	2.97 (0.42)
	$(a_0 = 6, \lambda_0 = 1)$			
$\tilde{a}$	6.11 (1.51)	6.20 (1.23)	5.80 (0.93)	5.83 (0.77)
$\tilde{\lambda}$	1.01 (0.26)	1.02 (0.22)	0.98 (0.17)	0.98 (0.14)
a^{**}	5.43 (1.18)	6.05 (1.17)	5.22 (0.75)	5.71 (0.96)
λ^{**}	0.88 (0.21)	1.00 (0.20)	0.86 (0.13)	0.96 (0.13)

5. NUMERICAL SIMULATION RESULTS

We compare the performances of both estimation methods on simulated data for several models. Two sets of population parameters θ_0 are used: $(a_0 = 6, \lambda_0 = 1)$ and $(a_0 = 5, \lambda_0 = 3)$. In each case, 100 datasets are generated with an Euler's scheme with sampling interval $\delta = 10^{-4}T$ on time interval $[0, T]$, with $T = 5$, and $N = 50, 100$ subjects, $n = 100, 500$. The parameter θ_0 is estimated *via* $\tilde{\theta}_N$ (method 1) and *via* either θ_N^* or θ_N^{**} (method 2). The empirical mean and standard deviation are computed from the 100 datasets. We consider:

Example 1. $dX_i(t) = \phi_i dW_i(t)$, $X_i(0) = 0$.

Example 2. $dX_i(t) = \phi_i \sqrt{1 + X_i^2(t)} dW_i(t)$, $X_i(0) = 0$.

Example 3. $dX_i(t) = -\rho X_i(t) dt + \phi_i dW_i(t)$, $X_i(0) = 0$.

Example 4. $dX_i(t) = -\rho X_i(t) dt + \phi_i \sqrt{1 + X_i^2(t)} dW_i(t)$, $X_i(0) = 0$.

For Example 1, estimation method 1 leads to the exact MLE of θ_0. Examples 2−4 rely on Section 4.

Both methods require to optimize a criterion. The two criteria use the function gamma, which takes large values. Thus, the optimization is implemented vanishing the gradients of the two criteria, which are more stable.

The results for Examples 1−4 are displayed in Tables 1−4 respectively. The results are satisfactory overall and similar for the 4 models, even when the model includes a drift. Method 1 estimators are biased for $n = 100$.

TABLE 3. Example 3 ($\rho = 1$): Empirical mean and standard deviation (in brackets) of $\tilde{\theta}$ (method 1) and θ^{**} (method 2) computed from 100 datasets.

	$N = 50$		$N = 100$	
	$n = 100$	$n = 500$	$n = 100$	$n = 500$
	$(a_0 = 5, \lambda_0 = 3)$			
$\tilde{a}$	5.16 (1.26)	5.12 (1.14)	5.01 (0.71)	4.98 (0.68)
$\tilde{\lambda}$	3.03 (0.75)	3.06 (0.68)	2.93 (0.43)	2.97 (0.41)
a^{**}	4.68 (1.03)	5.02 (1.09)	4.57 (0.60)	4.89 (0.65)
λ^{**}	2.70 (0.61)	2.99 (0.65)	2.62 (0.36)	2.90 (0.39)
	$(a_0 = 6, \lambda_0 = 1)$			
$\tilde{a}$	6.44 (1.35)	6.35 (1.23)	6.10 (0.99)	6.04 (0.88)
$\tilde{\lambda}$	1.07 (0.23)	1.07 (0.21)	0.98 (0.19)	1.00 (0.17)
a^{**}	5.72 (1.07)	6.20 (1.17)	5.44 (0.81)	5.90 (0.84)
λ^{**}	0.93 (0.18)	1.04 (0.19)	0.86 (0.15)	0.97 (0.15)

TABLE 4. Example 4. ($\rho = 1$) Empirical mean and standard deviation (in brackets) of $\tilde{\theta}$ (method 1) and θ^{**} (method 2) computed from 100 datasets.

	$N = 50$		$N = 100$	
	$n = 100$	$n = 500$	$n = 100$	$n = 500$
	$(a_0 = 5, \lambda_0 = 3)$			
$\tilde{a}$	4.97 (0.90)	5.08 (0.83)	4.86 (0.74)	4.97 (0.74)
$\tilde{\lambda}$	2.90 (0.58)	3.04 (0.55)	2.83 (0.47)	2.96 (0.46)
a^{**}	4.54 (0.75)	4.98 (0.81)	4.45 (0.61)	4.88 (0.71)
λ^{**}	2.60 (0.48)	2.98 (0.53)	2.54 (0.39)	2.90 (0.44)
	$(a_0 = 6, \lambda_0 = 1)$			
$\tilde{a}$	6.21 (1.28)	6.17 (1.15)	6.01 (0.81)	6.01 (0.76)
$\tilde{\lambda}$	1.01 (0.22)	1.02 (0.20)	0.98 (0.15)	1.00 (0.14)
a^{**}	5.48 (1.17)	6.02 (1.10)	5.38 (0.66)	5.88 (0.73)
λ^{**}	0.88 (0.20)	1.00 (0.19)	0.86 (0.12)	0.98 (0.13)

When $\sigma(.) \not\equiv 1$, this is expected due to the Euler approximation of the likelihood. Nevertheless, for fixed N, we observe the convergence of the estimators to the true value when n increases. We also observe the convergence of the estimators to the true values when both N and n increase. This clearly illustrates consistency of the estimators when both n and N tend to infinity. Finally, the implementation of the method 2 in Examples $2-4$ requires to choose a value for the threshold k. The results are displayed for $k = 0.5$. Simulations with various values of k have not shown any significant impact of k on the estimators performances.

6. EXTENSIONS AND CONCLUDING REMARKS

In this paper, we study the estimation of population parameters in a SDE with a linear random effect in the diffusion coefficient from discrete observations of N i.i.d. trajectories on a fixed length time interval. We especially study the case of a null drift and of $\phi_i = 1/\Gamma_i^{1/2}$ with $\Gamma_i \sim G(a, \lambda)$. This leads to estimators using two different approaches. The first method is based on an approximation of the exact likelihood relying on the Euler's scheme of the SDE. The second method uses a plug-in of estimators of the random effects in the likelihood of $(\phi_1, \ldots, \phi_N)$.

Several extensions are possible. The theoretical results extend easily to N individual processes observed on $[0, T_i]$, $i = 1, \ldots, N$ at n discrete time points $t_{i,j} = jT_i/n$ with all T_i's fixed and $T_i \leq T$ for some fixed T. Another direction for extensions is to look at other distributions for the random effects. In particular, the plug-in

method applies for any distribution provided that we introduce appropriate truncations as is done here. The Euler approximation could also be applied with other distributions but numerical integration would then be necessary. For a more general model for the diffusion coefficient including a non linear random effect, the two approaches studied here could be extended, in particular the second method.

APPENDIX A. PROOFS

Proof of Proposition 3.1. Using the fact that ϕ_i^{-2} has Gamma distribution $G(a, \lambda)$, we get the result as:

$$\tilde{\lambda}_i(\theta, X_i) = \int_{(0,+\infty)} \frac{\lambda^a \gamma^{a-1+n/2}}{\Gamma(a)} \exp\left[-\gamma\left(\lambda + \frac{1}{2}S_i\right)\right] d\gamma = \frac{\lambda^a \Gamma(a+n/2)}{\Gamma(a)\left(\lambda + \frac{1}{2}S_i\right)^{a+n/2}}. \qquad \square$$

Proof of Proposition 3.2. Let $\chi_i = \sum_{j=1}^n (W_i(t_j) - W_i(t_{j-1}))^2/\Delta = R_i^1/\Delta$. As $S_i^1 = \Gamma_i^{-1}\chi_i$, (see 3.7), $\beta_i^1(\lambda) = \frac{\lambda \Gamma_i}{\lambda \Gamma_i + \frac{1}{2}\chi_i}$, $\frac{S_i^1}{2\lambda} = \frac{\chi_i/2}{\lambda \Gamma_i}$ Under P_θ, Γ_i and χ_i are independent, Γ_i is $G(a, \lambda)$ and χ_i is $\chi^2(n) = G(n/2, 1/2)$. Hence the results using Proposition B.1. $\qquad \square$

Proof of Proposition 3.4. We need the following Lemma and Proposition:

Lemma A.1. *For all θ, $\mathbb{E}_\theta((X_i(t) - X_i(s))^{2p}|\phi_i) \leq C(2p)\sigma_1^{2p}\phi_i^{2p}|t-s|^p$ where $C(2p)$ is a numerical constant.*

Proposition A.2. $(R_i^1/T) - 1 = T^{-1} \int_0^T H_{i,1}^n(s) dW_i(s)$ and

$$(R_i/T) - 1 = T^{-1}\left(\int_0^T H_i^n(s) dW_i(s) + \int_0^T K_i^n(s) dW_i(s) + \int_0^T L_i^n(s) ds\right),$$

where, for $j = 1, \ldots, n$ and $s \in]t_{j-1}, t_j]$, $H_{i,1}^n(s) = 2(W_i(s) - W_i(t_{j-1}))$,

$$H_i^n(s) = 2\frac{(M_i(s) - M_i(t_{j-1}))\sigma(X_i(s))}{\sigma^2(X_i(t_{j-1}))}, \quad K_i^n(s) = 2\phi_i(t_j - s)\frac{\sigma^2(X_i(s))\sigma'(X_i(s))}{\sigma^2(X_i(t_{j-1}))},$$

$$L_i^n(s) = \phi_i^2 \kappa(X_i(s))(t_j - s)\frac{\sigma^2(X_i(s))}{\sigma^2(X_i(t_{j-1}))}, \quad with \quad \kappa = \sigma\sigma'' + (\sigma')^2.$$

Lemma A.1 and Proposition A.2 yield $\mathbb{E}_\theta(\frac{R_i^1}{T} - 1|\mathcal{F}_0) = 0$ and

$$\mathbb{E}_\theta\left(\frac{R_i}{T} - 1|\mathcal{F}_0\right) = \mathbb{E}_\theta\left(\frac{R_i}{T} - \frac{R_i^1}{T}|\mathcal{F}_0\right) = \frac{1}{T}\int_0^T \mathbb{E}_\theta(L_i^n(s)|\mathcal{F}_0) ds.$$

Using (H1), we get $|L_i^n(s)| \leq C\phi_i^2 \sum_{j=1}^n \mathbb{1}_{]t_{j-1}, t_j]}(s)(t_j - s) \leq C\phi_i^2 \Delta \mathbb{1}_{]0,T]}(s)$, for C depending on σ_0, σ_1, K. Thus, the first inequality of Proposition 3.4.

As $(\frac{R_i}{T} - 1)^{2p} = (A_1 + A_2 + A_3)^{2p} \leq 3^{2p-1}\sum_{i=1}^3 A_i^{2p}$, we study separately the three terms A_i^{2p}. We have $A_3^{2p} = \left(\frac{1}{T}\int_0^T L_i^n(s) ds\right)^{2p} \leq (C\phi_i^2 \Delta)^{2p}$. Next, we use the Burkholder–Davies–Gundy (BDG), the Hölder's inequalities and (H1):

$$T^{2p}\mathbb{E}_\theta(A_2^{2p}|\mathcal{F}_0) \leq C(2p)\mathbb{E}_\theta\left(\left(\int_0^T (K_i^n(s))^2 ds\right)^p|\mathcal{F}_0\right) \leq C(2p)T^{p-1}\mathbb{E}_\theta\left(\int_0^T \mathbb{E}_\theta((K_i^n(s))^{2p}|\mathcal{F}_0) ds\right)$$

where $(K_i^n(s))^{2p} \leq C\phi_i^{2p}\Delta^{2p}\mathbb{1}_{]0,T]}(s)$. Finally, for $T^{2p}\mathbb{E}_\theta(A_1^{2p}|\mathcal{F}_0)$ we study $\left(\int_0^T (H_i^n(s))^2 ds\right)^p$. By the Holder's inequality, we have,

$$\left(\int_0^T (H_i^n(s))^2 ds\right)^p \leq C^p n^{p-1}\sum_{j=1}^n \Delta^{p-1}\int_{t_{j-1}}^{t_j} (M_i(s) - M_i(t_{j-1}))^{2p} ds.$$

Consequently, for constants C depending on σ_0, σ_1, K,

$$\mathbb{E}_\theta\left(\left(\int_0^T (H_i^n(s))^2 \mathrm{d}s\right)^p |\mathcal{F}_0\right) \leq C \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \mathbb{E}_\theta(\left(\int_{t_{j-1}}^s \sigma^2(X_i(u))\mathrm{d}u\right)^p |\mathcal{F}_0)\mathrm{d}s \leq C\Delta^p.$$

Finally, to study the difference $R_i - R_i^1$, we only need to study the term: $\int_0^T (H_i^n(s) - H_{i,1}^n(s))\mathrm{d}W_i(s) = \int_0^T 2\sum_{k=1}^3 \sum_{j=1}^n \mathbb{1}_{]t_{j-1},t_j]}(s)Z_{i,j}^k(s)\mathrm{d}W_i(s)$ with

$$Z_{i,j}^1(s) = \frac{\sigma(X_i(s)) - \sigma(X_i(t_{j-1}))}{\sigma^2(X_i(t_{j-1}))} \int_{t_{j-1}}^s (\sigma(X_i(u)) - \sigma(X_i(t_{j-1})))\mathrm{d}W_i(u),$$

$$Z_{i,j}^2(s) = \frac{\sigma(X_i(s)) - \sigma(X_i(t_{j-1}))}{\sigma(X_i(t_{j-1}))} \int_{t_{j-1}}^s \mathrm{d}W_i(u),$$

$$Z_{i,j}^3(s) = \frac{1}{\sigma(X_i(t_{j-1}))} \int_{t_{j-1}}^s (\sigma(X_i(u)) - \sigma(X_i(t_{j-1})))\mathrm{d}W_i(u),$$

These terms are studied analogously using the BDG and Cauchy−Schwarz's inequalities. □

Proof of Lemma A.1. Recall that ϕ_i is $\mathcal{F}_0$-measurable and when dealing with the process $(X_i(t))$, conditioning on ϕ_i is equal to conditioning on $\mathcal{F}_0$. We have $\mathbb{E}_\theta((X_i(t) - X_i(s))^{2p}|\mathcal{F}_0) = \phi_i^{2p}\mathbb{E}_\theta((M_i(t) - M_i(s))^{2p}|\mathcal{F}_0)$. By the BDG inequality and (H1), for $s \leq t$,

$$\mathbb{E}_\theta((M_i(t) - M_i(s))^{2p}|\mathcal{F}_0) \leq C(2p)\mathbb{E}\left(\left(\int_s^t \sigma^2(X_i(u))\mathrm{d}u\right)^p |\mathcal{F}_0\right) \leq C(2p)\sigma_1^{2p}(t - s)^{2p}. \quad \square$$

Proof of Proposition A.2. By the Ito's formula, we have:

$$(M_i(t_j) - M_i(t_{j-1}))^2 = 2\int_{t_{j-1}}^{t_j} (M_i(s) - M_i(t_{j-1}))\sigma(X_i(s))\mathrm{d}W_i(s) + \int_{t_{j-1}}^{t_j} \sigma^2(X_i(s))\mathrm{d}s.$$

We split: $\sigma^2(X_i(s)) = \sigma^2(X_i(t_{j-1})) + \sigma^2(X_i(s)) - \sigma^2(X_i(t_{j-1}))$ and use the Ito formula: $\sigma^2(X_i(s)) - \sigma^2(X_i(t_{j-1})) = \phi_i \int_{t_{j-1}}^s (\sigma^2)'(X_i(u))\sigma(X_i(u))\mathrm{d}W_i(u) + \frac{1}{2}\phi_i^2 \int_{t_{j-1}}^s (\sigma^2)''(X_i(u))\sigma^2(X_i(u))\mathrm{d}u$. Integrating, Fubini formula yields:

$$\int_{t_{j-1}}^{t_j} (\sigma^2(X_i(s)) - \sigma^2(X_i(t_{j-1})))\mathrm{d}s = \phi_i \int_{t_{j-1}}^{t_j} (t_j - u)(\sigma^2)'(X_i(u))\sigma(X_i(u))\mathrm{d}W_i(u)$$
$$+ \frac{1}{2}\phi_i^2 \int_{t_{j-1}}^{t_j} (t_j - u)(\sigma^2)''(X_i(u))\sigma^2(X_i(u))\mathrm{d}u.$$

The result for R_i^1 corresponds to $\sigma(.) \equiv 1$. Hence the results. □

Proof of Proposition 3.5. Recall that $S_i = n\Gamma_i^{-1}R_i/T$. We have by (3.4) $\tilde{G}_N(\theta_0) = \mathcal{S}_N(\theta_0) + \left(\sum_{i=1}^N Y_i(\theta_0) \; \sum_{i=1}^N Z_i(\theta_0)\right)'$ where $Y_i(\theta_0) = \Gamma_i - \frac{(a_0 + \frac{n}{2})\Gamma_i}{\lambda_0\Gamma_i + C_i}$ and $Z_i(\theta_0) = \psi(a_0 + \frac{n}{2}) - \log(\lambda_0\Gamma_i + C_i)$.

Therefore, we have to prove that $\frac{1}{\sqrt{N}}\left(\sum_{i=1}^N Y_i(\theta_0) \; \sum_{i=1}^N Z_i(\theta_0)\right)'$ tends to 0 in $\mathbb{P}_{\theta_0}$-probability as n, N tend to infinity. To distinguish the two cases $\sigma(.) \equiv 1$ and $\sigma(.) \not\equiv 1$, we introduce the random variables $Y_i^1(\theta_0), Z_i^1(\theta_0)$

where we replaced S_i by S_i^1. We proceed on two steps:

(1) $\frac{1}{\sqrt{N}} \left(\sum_{i=1}^{N} Y_i^1(\theta_0) \ \ \sum_{i=1}^{N} Z_i^1(\theta_0) \right)' = o_{\mathbb{P}_{\theta_0}}(1)$ as $N, n \to \infty$.

(2) $\frac{1}{\sqrt{N}} \left(\sum_{i=1}^{N} (Y_i(\theta_0) - Y_i^1(\theta_0)) \ \ \sum_{i=1}^{N} (Z_i(\theta_0) - Z_i^1(\theta_0)) \right)' = o_{\mathbb{P}_{\theta_0}}(1)$ as $N, n \to \infty$ under the constraints $N/n \to 0$ and $\mathbb{E}_{\theta_0} \phi_i^8 < +\infty$.

Proof of (1). Let $C_i^1 = nR_i^1/(2T) = \Gamma_i S_i^1/2$ and $G_i^0 = \lambda_0 \Gamma_i$. We have $Y_i^1(\theta_0) = \Gamma_i - \frac{a_0 + n/2}{\lambda_0} \frac{G_i^0}{G_i^0 + C_i^1}$. In what follows, we use repeatedly the fact that G_i^0 and C_i^1 are independent, that $G_i^0 \sim G(a_0, 1)$ and $C_i^1 \sim G(n/2, 1)$. Hence, $G_i^0 + C_i^1$ and $G_i^0/(G_i^0 + C_i^1)$ are independent, the latter with distribution $\beta^{(1)}(a_0, (n/2))$, the former with distribution $G(a_0 + (n/2), 1)$.

The r.v. $Y_i^1(\theta_0), i = 1, \ldots, N$ are *i.i.d.* with $\mathbb{E}_{\theta_0}(Y_i^1(\theta_0)) = 0$, $\mathbb{E}_{\theta_0}(Y_i^1(\theta_0)^2) = \frac{a_0(a_0+1)}{\lambda_0^2(a_0+1+n/2)}$. Therefore, $N^{-1/2} \sum_{i=1}^{N} Y_i^1(\theta_0) = o_{\mathbb{P}_{\theta_0}}(1)$.

Analogously, $Z_i^1(\theta_0) = \psi(a_0 + n/2) - \log(G_i^0 + C_i^1)$ satisfies $\mathbb{E}_{\theta_0}(Z_i^1(\theta_0)) = 0, \mathbb{E}_{\theta_0}(Z_i^1(\theta_0))^2 = 1/(a_0 + n/2) + o(1/n)$. Thus, $N^{-1/2} \sum_{i=1}^{N} Z_i^1(\theta_0) = o_{\mathbb{P}_{\theta_0}}(1)$.

Proof of (2). We introduce $C_i = nR_i/(2T)$. We have $Y_i(\theta_0) - Y_i^1(\theta_0) = \frac{(a_0 + \frac{n}{2}) \Gamma_i}{G_i^0 + C_i^1} - \frac{(a_0 + \frac{n}{2}) \Gamma_i}{G_i^0 + C_i}$, and $Z_i(\theta_0) - Z_i^1(\theta_0) = \log(G_i^0 + C_i^1) - \log(G_i^0 + C_i)$. Thus,

$$Y_i(\theta_0) - Y_i^1(\theta_0) = \frac{a_0 + \frac{n}{2}}{\lambda_0} \frac{G_i^0}{(G_i^0 + C_i)(G_i^0 + C_i^1)} \frac{n}{2} \left(\frac{R_i}{T} - \frac{R_i^1}{T} \right)$$

We introduce the set $\Omega_i = \{|(R_i/T) - 1| \leq 1/2\}$. On Ω_i, we use $G_i^0 + C_i \geq (n/4)$. So

$$\mathbb{E}_{\theta_0} |Y_i(\theta_0) - Y_i^1(\theta_0)| \mathbb{1}_{\Omega_i} \leq 2 \frac{a_0 + \frac{n}{2}}{\lambda_0} \mathbb{E}_{\theta_0} \left(\frac{G_i^0}{G_i^0 + C_i^1} \left| \frac{R_i}{T} - \frac{R_i^1}{T} \right| \right).$$

Then, if $\mathbb{E}_{\theta_0} \phi_i^4 < +\infty$, we have by Proposition 3.4

$$\mathbb{E}_{\theta_0} \left(\frac{G_i^0}{G_i^0 + C_i^1} \right)^2 = O\left(\frac{1}{n^2} \right) \quad \text{and} \quad \mathbb{E}_{\theta_0} \left(\frac{R_i}{T} - \frac{R_i^1}{T} \right)^2 \leq (T/n)^2 \mathbb{E}_{\theta_0} \phi_i^4.$$

Therefore, using the Cauchy–Schwarz's inequality, $\mathbb{E}_{\theta_0} |Y_i(\theta_0) - Y_i^1(\theta_0)| \mathbb{1}_{\Omega_i} \leq Cn \left(\frac{C}{n^4} \right)^{1/2} \leq \frac{C}{n}$.

On Ω_i^c, we use $G_i^0/(G_i^0 + C_i) \leq 1$. Therefore,

$$|Y_i(\theta_0) - Y_i^1(\theta_0)| \mathbb{1}_{\Omega_i^c} \leq \frac{\frac{n}{2}(a_0 + \frac{n}{2})}{\lambda_0} \frac{1}{G_i^0 + C_i^1} \left| \frac{R_i}{T} - \frac{R_i^1}{T} \right| \mathbb{1}_{\Omega_i^c}.$$

We have if $\mathbb{E}_{\theta_0} \phi_i^8 < +\infty$, by Proposition 3.4, $\mathbb{E}_{\theta_0} \left(\frac{1}{G_i^0 + C_i^1} \right)^4 = O\left(\frac{1}{n^4} \right)$, $\mathbb{E}_{\theta_0} \left(\frac{R_i}{T} - \frac{R_i^1}{T} \right)^4 \leq (T/n)^4 \mathbb{E}_{\theta_0}(\phi_i^4 + \phi_i^8)$, and $\mathbb{P}_{\theta_0}(\Omega_i^c) \leq 2^{2p} \mathbb{E}_{\theta_0} |\frac{R_i}{T} - 1|^{2p}$. Using the Cauchy–Schwarz's inequality twice, the above inequality with $p = 2$ and Proposition 3.4 with the condition $\mathbb{E}_{\theta_0} \phi_i^8 < +\infty$, we get:

$$\mathbb{E}_{\theta_0} |Y_i(\theta_0) - Y_i^1(\theta_0)| \mathbb{1}_{\Omega_i^c} \leq Cn^2 (C/n^4)^{1/4} (C/n^4)^{1/4} \left(\mathbb{P}_{\theta_0}(\Omega_i^c) \right)^{1/2} \leq C/n.$$

We can conclude that under the condition $\mathbb{E}_{\theta_0} \phi_i^8 < +\infty$,

$$\mathbb{E}_{\theta_0} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (Y_i(\theta_0) - Y_i^1(\theta_0)) \right| \leq C\sqrt{N}/n. \tag{A.1}$$

We now turn to the other difference. We have by the Taylor's formula

$$Z_i(\theta_0) - Z_i^1(\theta_0) = \frac{n}{2} \frac{1}{G_i^0 + n/2} \left(\frac{R_i^1}{T} - \frac{R_i}{T} \right) + \frac{n}{2} \left(\frac{R_i^1}{T} - \frac{R_i}{T} \right) \int_0^1 f_i(s) \mathrm{d}s,$$

where $f_i(s) = -\dfrac{s\frac{n}{2}\left(\frac{R_i^1}{T}-1\right)+(1-s)\frac{n}{2}\left(\frac{R_i}{T}-1\right)}{(G_i^0+n/2)(G_i^0+s\frac{n}{2}\frac{R_i^1}{T}+(1-s)\frac{n}{2}\frac{R_i}{T})}$. We use that:

$$\left| \mathbb{E}_{\theta_0} \frac{n}{2} \frac{1}{G_i^0 + n/2} \left(\frac{R_i^1}{T} - \frac{R_i^1}{T} \right) |\mathcal{F}_0 \right| \leq \frac{n}{2} \frac{C}{n} \frac{1}{G_i^0 + n/2} \phi_i^2 = C \frac{1}{\Gamma_i^2(G_i^0 + n/2)}$$

and if $\mathbb{E}_{\theta_0}\phi_i^4 < +\infty$, $\mathbb{E}_{\theta_0} \frac{1}{\Gamma_i^2(G_i^0+n/2)} \leq C \left(\frac{1}{(a_0+n/2-1)(a_0+n/2-2)} \right)^{1/2}$, to obtain that,

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbb{E}_{\theta_0} \left| \mathbb{E}_{\theta_0} \frac{n}{2} \frac{1}{G_i^0 + n/2} \left(\frac{R_i^1}{T} - \frac{R_i}{T} \right) |\mathcal{F}_0 \right| \leq \frac{C\sqrt{N}}{n}.$$

On the other hand, noticing that for $s \in [0,1]$: $|f_i(s)| \leq \frac{n/2}{(G_i^0+n/2)G_i^0} \left(|\frac{R_i^1}{T} - 1| + |\frac{R_i}{T} - 1| \right)$, if $\mathbb{E}_{\theta_0}\phi_i^8 < +\infty$,

$$\mathbb{E}_{\theta_0} \left(\int_0^1 f_i(s) \mathrm{d}s \right)^2 \leq \left(\mathbb{E}_{\theta_0} \frac{1}{(G_i^0)^4} \mathbb{E}_{\theta_0} \left(\left| \frac{R_i^1}{T} - 1 \right|^4 + \left| \frac{R_i}{T} - 1 \right|^4 \right) \right)^{1/2} \leq C \left(\frac{T}{n} \right).$$

Finally, we have $\mathbb{E}_{\theta_0} \frac{n}{2} | \left(\frac{R_i^1}{T} - \frac{R_i}{T} \right) \int_0^1 f_i(s) \mathrm{d}s | \leq \left(\frac{C}{n} \right)^{1/2}$. Therefore, $\mathbb{E}_{\theta_0} |\frac{1}{\sqrt{N}} \sum_{i=1}^N (Z_i(\theta_0) - Z_i^1(\theta_0))| \leq C\sqrt{N}(1/n + 1/n^{1/2})$. $\square$

Proof of Proposition 3.6. To obtain $\widetilde{\mathcal{I}}_N(\theta_0)/N = \mathcal{I}(\theta_0) + o_{\mathbb{P}_{\theta_0}}(1)$, we have to prove that $\frac{1}{N} \sum_{i=1}^N A_i(\theta_0) \to 0$, and $\frac{1}{N} \sum_{i=1}^N B_i(\theta_0) \to 0$, in $\mathbb{P}_{\theta_0}$ − probability where

$$A_i(\theta_0) = \frac{a_0 + n/2}{(\lambda_0 + S_i/2)^2} = \frac{a_0 + n/2}{\lambda_0^2} \frac{(G_i^0)^2}{(G_i^0 + C_i)^2}, \quad B_i(\theta_0) = \frac{1}{\lambda_0 + S_i/2} = \frac{\Gamma_i}{G_i^0 + C_i}. \tag{A.2}$$

As in the previous proposition, we separate the cases $\sigma(.) \equiv 1$ and $\sigma(.) \not\equiv 1$ and define the random variables $A_i^1(\theta_0), B_i^1(\theta_0)$ where S_i is replaced by S_i^1:

$$A_i^1(\theta_0) = \frac{a_0 + n/2}{\lambda_0^2} \frac{(G_i^0)^2}{(G_i^0 + C_i^1)^2}, \quad B_i^1(\theta_0) = \frac{\Gamma_i}{G_i^0 + C_i^1}.$$

Recall that $C_i^1 \sim G(n/2, 1)$ and is independent of Γ_i. Thus,

$$\mathbb{E}_{\theta_0} A_i^1(\theta_0) = \frac{a_0(a_0 + 1)}{\lambda_0^2(a_0 + n/2 + 1)} = O\left(\frac{1}{n} \right), \quad \mathbb{E}_{\theta_0} B_i^1(\theta_0) = \frac{a_0}{\lambda_0(a_0 + n/2)} = O\left(\frac{1}{n} \right).$$

This implies $\frac{1}{N} \sum_{i=1}^N \mathbb{E}_{\theta_0} A_i^1(\theta_0) = O(\frac{1}{n})$, $\frac{1}{N} \sum_{i=1}^N \mathbb{E}_{\theta_0} B_i^1(\theta_0) = O(\frac{1}{n})$. Next, we study the differences $A_i(\theta_0) - A_i^1(\theta_0)$, $B_i(\theta_0) - B_i^1(\theta_0)$.

$$A_i(\theta_0) - A_i^1(\theta_0) = \frac{a_0 + n/2}{\lambda_0^2} (C_i - C_i^1) \left(\frac{(G_i^0)^2}{(G_i^0 + C_i)(G_i^0 + C_i^1)^2} + \frac{(G_i^0)^2}{(G_i^0 + C_i)^2(G_i^0 + C_i^1)} \right).$$

Thus: $|A_i(\theta_0) - A_i^1(\theta_0)| \le \frac{a_0 + n/2}{\lambda_0^2}|C_i^1 - C_i|\frac{2G_i^0}{(G_i^0 + C_i)(G_i^0 + C_i^1)}$. We introduce again the set $\Omega_i = \{|(R_i/T) - 1| \le 1/2\}$. On Ω_i, using $G_i^0 + C_i \ge (n/4)$, we have

$$\mathbb{E}_{\theta_0}|A_i(\theta_0) - A_i^1(\theta_0)|\mathbb{1}_{\Omega_i} \le 2\frac{a_0 + n/2}{\lambda_0^2(n/4)}\left[\mathbb{E}_{\theta_0}\left(\frac{G_i^0}{G_i^0 + C_i^1}\right)^2 \mathbb{E}_{\theta_0}\left(C_i^1 - C_i\right)^2\right]^{1/2} \le C/n.$$

Next, using $G_i^0/(G_i^0 + C_i) \le 1$ on Ω_i^c, we have $|A_i(\theta_0) - A_i^1(\theta_0)|\mathbb{1}_{\Omega_i^c} \le \frac{a_0 + n/2}{\lambda_0^2}|C_i^1 - C_i|\frac{2}{(G_i^0 + C_i^1)}$. Thus, using the same arguments than in proof of Proposition 3.5

$$\mathbb{E}_{\theta_0}(|A_i(\theta_0) - A_i^1(\theta_0)|\mathbb{1}_{\Omega_i^c}) \le C\left(\mathbb{P}_{\theta_0}(\Omega_i^c)\right)^{1/2} \le \frac{C}{n}.$$

Analogously, $B_i(\theta_0) - B_i^1(\theta_0) = \frac{1}{\lambda_0}(C_i^1 - C_i)\frac{G_i^0}{(G_i^0 + C_i)(G_i^0 + C_i^1)}$. Introducing the set Ω_i again, we obtain $|B_i(\theta_0) - B_i^1(\theta_0)| \le C\left(\frac{G_i^0}{G_i^0 + C_i^1} + \mathbb{1}_{\Omega_i^c}\frac{n}{G_i^0 + C_i^1}\right)|R_i/T - R_i^1/T|$. We conclude $\mathbb{E}_{\theta_0}|B_i(\theta_0) - B_i^1(\theta_0)| \le \frac{C}{n} + (\mathbb{P}_{\theta_0}(\Omega_i^c))^{1/2}$. So the proof is complete. $\qquad\square$

Proof of Proposition 3.7. To obtain the weak consistency of $\tilde{\theta}_N$ and its asymptotic normality, we follow the scheme described in [1] (Thm. 3.4 and Lem. 3.5) and [7] (Thm. 4.1), see also [12]. We must prove that:

(1) Under P_{θ_0}, $\tilde{G}_N(\theta_0)/\sqrt{N} \to_{\mathcal{D}} \mathcal{N}_2(0, \mathcal{I}(\theta_0))$ as $N \to \infty$.
(2) $\widetilde{\mathcal{I}}_N(\theta_0)/N \to \mathcal{I}(\theta_0)$ in P_{θ_0}-probability.
(3) $\sup_{\theta \in M_{c,N}} |\widetilde{\mathcal{I}}_N(\theta)/N - \mathcal{I}(\theta_0)| \to 0$ in P_{θ_0}-probability, where $M_{c,N} = \{\theta \in (0, +\infty)^2, \|\theta - \theta_0\| \le c/\sqrt{N}\}$. (Uniformity condition)

Points (1) and (2) are directly implied by Propositions 3.5 and 3.6. It remains to prove (3). We will prove

(a) $\mathbb{E}_{\theta_0}(\sup_{\theta \in M_{c,N}} |\widetilde{\mathcal{I}}_N^1(\theta)/N - \widetilde{\mathcal{I}}_N^1(\theta_0)/N|) \to 0$.
(b) $\mathbb{E}_{\theta_0}(\sup_{\theta \in M_{c,N}} |\widetilde{\mathcal{I}}_N^1(\theta)/N - \widetilde{\mathcal{I}}_N(\theta)/N|) \to 0$.

Point (a). Let $\varepsilon > 0$ be such that $a_0 - \varepsilon > 0, \lambda_0 - \varepsilon > 0$. Choose N large enough to ensure that $M_{c,N} = \{(a, \lambda) \in (0, +\infty)^2, |a - a_0| \le c/\sqrt{N}, |\lambda - \lambda_0| \le c/\sqrt{N}\} \subset [a_0 - \varepsilon, a_0 + c] \times [\lambda_0 - \varepsilon, \lambda_0 + c]$ and $n > 8$. We have

$$\widetilde{\mathcal{I}}_N^1(\theta)/N - \widetilde{\mathcal{I}}_N^1(\theta_0)/N = \mathcal{I}(\theta) - \mathcal{I}(\theta_0) + \begin{pmatrix} D_N^{11}(\theta, \theta_0) & D_N^{12}(\theta, \theta_0) \\ D_N^{12}(\theta, \theta_0) & D_N^{22}(\theta, \theta_0) \end{pmatrix}$$

where $D_N^{11}(\theta, \theta_0) = \frac{1}{N}\sum_{i=1}^N(A_i^1(\theta) - A_i^1(\theta_0))$, $D_N^{12}(\theta, \theta_0) = \frac{1}{N}\sum_{i=1}^N(B_i^1(\theta) - B_i^1(\theta_0))$, $D_N^{22}(\theta, \theta_0) = -(\psi'(a + n/2) - \psi'(a_0 + n/2))$. We only study D^{11} and D^{12} which are the most difficult. We can write $D_N^{11}(\theta, \theta_0) = c_N + d_N$, with $G_i = \lambda \Gamma_i$ and

$$c_N = (\lambda - \lambda_0)(a_0 + n/2)\frac{1}{N}\sum_{i=1}^N \frac{\Gamma_i(G_i + G_i^0 + 2C_i^1)}{(G_i + C_i^1)^2(G_i^0 + C_i^1)^2}, \quad d_N = (a_0 - a)\frac{1}{N}\sum_{i=1}^N \frac{\Gamma_i}{(G_i + C_i^1)^2}.$$

For $\theta \in M_{c,N}$, we have the bounds $|d_N| \le \frac{c}{(\lambda_0 - \varepsilon)^2\sqrt{N}}$, and

$$|c_N| \le \frac{c(a_0 + n/2)}{\sqrt{N}}\left[(2\lambda_0 + c)\left(\frac{\Gamma_i}{C_i^1}\right)^4 + \left(2\frac{\Gamma_i}{C_i^1}\right)^3\right]$$

We have, for $n > 2k$, $\mathbb{E}_{\theta_0}\left(\frac{\Gamma_i}{2C_i^1}\right)^k = (2\lambda_0)^{-k}\frac{(a_0 + k - 1)(a_0 + k - 2)...a_0}{(n/2 - 1)(n/2 - 2)...(n/2 - k)}$.

Thus, $\mathbb{E}_{\theta_0} \sup_{\theta \in M_{c,N}} |c_N| \leq \frac{C}{\sqrt{N}n^2}$. Thus $\mathbb{E}_{\theta_0} \sup_{\theta \in M_{c,N}} |D_N^{11}(\theta, \theta_0)| = O(1/\sqrt{N})$. For the other term, (a) is proved as we have

$$\sup_{\theta \in M_{c,N}} |D_N^{12}(\theta, \theta_0)| \leq |\lambda - \lambda_0| \frac{1}{\lambda_0(\lambda_0 - \varepsilon)} = O(1/\sqrt{N}).$$

We now prove (b). For this, we prove the convergence to 0 of (see (A.2): $\mathbb{E}_{\theta_0} \sup_{\theta \in M_{c,N}} |A_i(\theta) - A_i^1(\theta)|$ and $\mathbb{E}_{\theta_0} \sup_{\theta \in M_{c,N}} |B_i(\theta) - B_i^1(\theta)|$. We have $|A_i(\theta) - A_i^1(\theta)| \leq \frac{a_0 + n/2 + c}{(\lambda_0 - \varepsilon)^2} |C_i^1 - C_i| \frac{2G_i}{(G_i + C_i)(G_i + C_i^1)}$. On the set $\Omega_i = \{|(R_i/T) - 1| \leq 1/2\}$, $R_i/T \geq 1/2$ and $G_i + C_i \geq n/4$. Thus, as $G_i + C_i^1 > C_i^1$ and G_i and C_i^1 are independent, for $n > 4$,

$$\mathbb{E}_{\theta_0} \sup_{\theta \in M_{c,N}} |A_i(\theta) - A_i^1(\theta)| \mathbb{1}_{\Omega_i}$$

$$\leq 2 \frac{a_0 + n/2 + c}{(\lambda_0 - \varepsilon)^2 (n/4)} (n/2) \left(\mathbb{E}_{\theta_0}(R_i^1/T - R_i/T)^2 \mathbb{E}_{\theta_0}((\lambda_0 + c)\Gamma_i)^2 \mathbb{E}_{\theta_0}(1/C_i^1)^2) \right)^{1/2}$$

$$\leq C \frac{T}{n} \left(\mathbb{E}_{\theta_0}(\phi_i^2 + \phi_i^4) \right)^{1/2} = O\left(\frac{1}{n}\right).$$

Next $(n > 8, \mathbb{E}_{\theta_0} \phi_i^8 < +\infty)$

$$\mathbb{E}_{\theta_0} \sup_{\theta \in M_{c,N}} |A_i(\theta) - A_i^1(\theta)| \mathbb{1}_{\Omega_i^c}$$

$$\leq \frac{2(a_0 + n/2 + c)n/2}{(\lambda_0 - \varepsilon)^2} (\mathbb{P}_{\theta_0}(\Omega_i^c))^{1/2} \left(\mathbb{E}_{\theta_0}(R_i^1/T - R_i/T)^4 \mathbb{E}_{\theta_0} \frac{1}{(nR_i^1/T)^4} \right)^{1/4} = o(1).$$

We have $B_i(\theta) - B_i^1(\theta) = \frac{\Gamma_i}{(\lambda \Gamma_i + \frac{nR_i}{2T})(\lambda \Gamma_i + \frac{nR_i^1}{2T})} (\frac{nR_i^1}{2T} - \frac{nR_i}{2T})$. Using that $\lambda \geq \lambda_0 - \varepsilon$,

$$\mathbb{E}_{\theta_0} \sup_{\theta \in M_{c,N}} |B_i(\theta) - B_i^1(\theta)| \leq C \frac{n}{2(\lambda_0 - \varepsilon)} \frac{T}{n} \left(\frac{1}{(n/2 - 1)(n/2 - 2)} \mathbb{E}_{\theta_0}(\phi_i^2 + \phi_i^4) \right)^{1/2}$$

$$= O\left(\frac{1}{n}\right).$$

Therefore, the proof of the first part of Proposition 3.7 is complete.

The fact that $\sqrt{N}(\tilde{\theta}_n - \theta_N) = o_{\mathbb{P}_{\theta_0}}(1)$ can be deduced from the above proof. $\qquad \square$

Proof of Proposition 3.9. We first consider the case $\sigma(.) \equiv 1$ and the estimating function $\nabla U_N(\theta)$. To get the result, it is enough to prove that:

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{n}{S_i^1} - \Gamma_i \right) = o_{\mathbb{P}_{\theta_0}}(1), \quad \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\log \frac{n}{S_i^1} - \log \Gamma_i \right) = o_{\mathbb{P}_{\theta_0}}(1), \tag{A.3}$$

where we recall that $S_i^1/n = \Gamma_i^{-1} R_i^1/T$ and nR_i^1/T is independent of Γ_i and has distribution $\chi^2(n)$. Using results recalled in Section 6, for $n > 2$, we have $\mathbb{E}_{\theta_0}(\frac{n}{S_i^1} - \Gamma_i) = \mathbb{E}_{\theta_0} \Gamma_i \mathbb{E}_{\theta_0}(\frac{n/2}{C_i^1} - 1) = \frac{a_0}{\lambda_0} O(n^{-1})$. Analogously, for $n > 4$, $\mathbb{E}_{\theta_0}(\frac{n}{S_i^1} - \Gamma_i)^2 = \mathbb{E}_{\theta_0} \Gamma_i^2 \mathbb{E}_{\theta_0}(\frac{n/2}{C_i^1} - 1)^2 = O(n^{-1})$. This implies: $\mathbb{E}_{\theta_0}(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\frac{n}{S_i^1} - \Gamma_i))^2 = O(n^{-1}) + \frac{N-1}{2}(O(n^{-1}))^2$.

Hence, the first part of (A.3) holds provided that $\sqrt{N}/n = o(1)$.

For the second assertion, we compute $\mathbb{E}_{\theta_0}(\log \frac{n}{S_i^1} - \log \Gamma_i) = -\psi(n/2) + \log(n/2) = O(n^{-1})$, and $\mathrm{Var}_{\theta_0}(\log \frac{n}{S_i^1} - \log \Gamma_i) = \mathbb{E}_{\theta_0}(-\log C_i^1 + \psi(n/2))^2 = O(n^{-1}).\psi'(n/2)$ Therefore, the second part of (A.3) also holds for $\sqrt{N}/n = o(1)$.

Next, the result on θ_N^* will follow analogously from the fact that $T_N = (\sqrt{N})^{-1} \sum_{i=1}^{N} (\widetilde{\log} \Gamma_i - \log \Gamma_i)$ and $\tau_N = (\sqrt{N})^{-1} \sum_{i=1}^{N} (\widetilde{\Gamma}_i - \Gamma_i)$ both tend to 0 in probability. The result follows from the first part of the proof and from Lemma 3.8 which we prove now.

Proof of Lemma 3.8. We write: $\widetilde{\Gamma}_i - \frac{n}{S_i^1} = \frac{n}{S_i} \mathbb{1}_{(S_i/n \geq k/\sqrt{n})} - \frac{n}{S_i^1} = \nu_i + \nu_i'$ with

$$\nu_i = \left(\frac{n}{S_i} - \frac{n}{S_i^1} \right) \mathbb{1}_{(S_i/n \geq k/\sqrt{n})}, \quad \nu_i' = -\frac{n}{S_i^1} \mathbb{1}_{(S_i/n < k/\sqrt{n})}. \tag{A.4}$$

And analogously $\log \frac{S_i}{n} \mathbb{1}_{(S_i/n \geq k/\sqrt{n})} - \log \frac{S_i^1}{n} = \tau_i + \tau_i'$ with

$$\tau_i = \left(\log \frac{S_i}{n} - \log \frac{S_i^1}{n} \right) \mathbb{1}_{(S_i/n \geq k/\sqrt{n})}, \quad \tau_i' = -\log \frac{S_i^1}{n} \mathbb{1}_{(S_i/n < k/\sqrt{n})}. \tag{A.5}$$

For $n > 4$ and $\mathbb{E}_{\theta_0} \phi_i^8 < +\infty$, using explicit computations, $\mathbb{E}_{\theta_0} (\frac{n}{S_i^1})^2 = O(1)$, $\mathbb{E}_{\theta_0} \log^2 \frac{S_i^1}{n} = O(1)$. To obtain that $\mathbb{E}_{\theta_0} (\nu_i' + \tau_i')^2 = O(n^{-2})$, we now prove that:

$$\mathbb{P}_{\theta_0}(S_i/n < k/\sqrt{n}) \leq C/n^2. \tag{A.6}$$

Proof of (A.6). We remark:

$$\left(\frac{S_i}{n} < \frac{k}{\sqrt{n}} \right) \subset \left(\left| \phi_i^2 - \frac{S_i}{n} \right| > \phi_i^2 - \frac{k}{\sqrt{n}}, \phi_i^2 \geq 2 \frac{k}{\sqrt{n}} \right) \cup \left(\phi_i^2 < 2 \frac{k}{\sqrt{n}} \right)$$

$$\subset \left(\left| \phi_i^2 - \frac{S_i}{n} \right| > \frac{\phi_i^2}{2} \right) \cup \left(\phi_i^{-2} > \frac{\sqrt{n}}{2k} \right) = \left(\left| 1 - \frac{R_i}{T} \right| > \frac{1}{2} \right) \cup \left(\phi_i^{-2} > \frac{\sqrt{n}}{2k} \right).$$

Consequently, using the Markov's inequality and Proposition 3.4 yields:

$$\mathbb{P}_{\theta_0}(S_i/n < k/\sqrt{n}) \leq C \left(2^4 \left(\frac{T}{n} \right)^2 (1 + \mathbb{E}_{\theta_0}(\phi_i^4 + \phi_i^8) + \frac{(2k)^4}{n^2} \mathbb{E}_{\theta_0} \phi_i^{-4} \right) \leq c'/n^2.$$

So the proof of (A.6) is complete. $\square$

It remains to study the terms ν_i, τ_i. We have on $(S_i/n \geq k/\sqrt{n})$:

$$|\nu_i| = \left| \frac{1}{S_i^1/n} (S_i^1/n - S_i/n) \left(\frac{1}{S_i/n} - \Gamma_i + \Gamma_i \right) \right|$$

$$\leq \frac{1}{R_i^1/T} \left| \left(\frac{R_i^1}{T} - \frac{R_i}{T} \right) \left(1 - \frac{R_i}{T} \right) \right| \Gamma_i^{-1} \frac{\sqrt{n}}{k} + \frac{\Gamma_i}{R_i^1/T} \left| \frac{R_i^1}{T} - \frac{R_i}{T} \right|.$$

We use Proposition 3.4, the Cauchy−Schwarz's inequality and the exact distribution of Γ_i and $\frac{1}{R_i^1/T}$ to obtain, for $n > 4$ and $\mathbb{E}_{\theta_0} \phi_i^8 < +\infty$:

$$\mathbb{E}_{\theta_0} \nu_i^2 \leq C \left[\frac{n}{k^2} \left(\mathbb{E}_{\theta_0} \left(\frac{R_i^1}{T} - \frac{R_i}{T} \right)^4 \mathbb{E}_{\theta_0} (1 - \frac{R_i}{T})^4 \right)^{1/2} + \left(\mathbb{E}_{\theta_0} \left(\frac{R_i^1}{T} - \frac{R_i}{T} \right)^2 \right) \right] \leq \frac{C'}{n^2}.$$

For the term τ_i, we use the Taylor's formula and get:

$$\tau_i = (S_i/n - S_i^1/n) \int_0^1 \frac{\mathrm{d}s}{s(S_i/n) + (1 - s)(S_i^1/n)}.$$

Then, we split the integral:

$$\int_0^1 \frac{\mathrm{d}s}{s(S_i/n) + (1-s)(S_i^1/n)} = \frac{1}{S_i^1/n} + \int_0^1 \frac{s(S_i^1/n - S_i/n)}{(s(S_i/n) + (1-s)(S_i^1/n))S_i^1/n}\mathrm{d}s.$$

Thus, on $(S_i/n \geq k/\sqrt{n})$, we obtain, after simplifications:

$$|\tau_i| \leq \frac{1}{R_i^1/T}|R_i/T - R_i^1/T| + \frac{\sqrt{n}}{k}\frac{1}{R_i^1/T}\Gamma_i^{-1}(R_i/T - R_i^1/T)^2.$$

This yields $\mathbb{E}_{\theta_0}\tau_i^2 \leq C/n^2$. The proof of Lemma 3.8 is now complete. $\square$

Applying Lemma 3.8, we obtain the result of Proposition 3.9.

Appendix B. Auxiliary results

We recall some properties of Gamma and related distributions. The Gamma distribution with parameters (a, λ) $(a > 0, \lambda > 0)$, $G(a, \lambda)$, has density $\gamma_{a,\lambda}(x) = (\lambda^a/\Gamma(a))x^{a-1}\mathrm{e}^{-x}\mathbb{1}_{(0,+\infty)}(x)$, where $\Gamma(a)$ is the Gamma fonction. The digamma function $\psi(a) = \Gamma'(a)/\Gamma(a)$ admits the following integral representation: $\psi(z) = -\gamma + \int_0^1 (1 - t^{z-1})/(1 - t)\mathrm{d}t$. (where $\gamma = \psi(1) = \Gamma'(1)$). For all positive a, we have $\psi'(a) = -\int_0^1 \frac{\log t}{1-t}\, t^{a-1}\mathrm{d}t$. Consequently, using an integration by part, $-a\psi'(a) = -1 - \int_0^1 t^a g(t)\mathrm{d}t$, where $g(t) = (\log t/(1-t))'$. A simple study yields that $t^a g(t)$ is integrable on $(0, 1)$ and positive except at $t = 1$. Thus, $1 - a\psi'(a) \neq 0$. The following asymptotic expansions as a tends to infinity hold:

$$\log \Gamma(a) = \left(a - \frac{1}{2}\right)\log a - a + \frac{1}{2}\log 2\pi + O\left(\frac{1}{a}\right), \tag{B.1}$$

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

The following results are classical.

Proposition B.1. *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, $\mathbb{E}\log(\lambda X) = \psi(a)$, $Var[\log(\lambda X)] = \psi'(a)$.*

If X, Y are independent, X having distribution $G(a, 1)$ and Y having distribution $G(b, 1)$ $(a, b > 0)$, then, $X + Y$ and $X/(X + Y)$ are independent, $X + Y$ has distribution $G(a + b, 1)$, $T = \frac{X}{X+Y}$ has distribution beta of the first kind with parameters (a, b), denoted by $\beta^{(1)}(a, b)$, and density

$$f_T(t) = \frac{1}{B(a, b)}t^{a-1}(1 - t)^{b-1}\mathbb{1}_{(0,1)}(t),$$

with $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ and $Z = X/Y$ has distribution beta of the second kind with parameters (a, b), denoted by $\beta^{(2)}(a, b)$, and density

$$f_Z(z) = \frac{1}{B(a, b)}\frac{z^{a-1}}{(1 + z)^{a+b}}\mathbb{1}_{(0,+\infty)}(z).$$

We have $E(T) = \frac{a}{a+b}$, $E(T^2) = \frac{a(a+1)}{(a+b)(a+b+1)}$, $Var(T) = \frac{ab}{(a+b)^2(a+b+1)}$, $E\log T = \psi(a) - \psi(a+b)$, $Var(\log T) = \psi'(a) - \psi'(a + b)$, $Cov(T, \log T) = \frac{b}{(a+b)^2}$.

References

[1] O. Barndorff-Nielsen and M. Sorensen, Information quantities in non-classical settings. *Comput. Stat. Data Anal.* **12** (1991) 143–158.

[2] M. Delattre and M. Lavielle, Coupling the SAEM algorithm and the extended kalman filter for maximum likelihood estimation in mixed-effects diffusion models. *Stat. Interface* **6** (2013) 519–532.

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

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

[5] S. Donnet and A. Samson, A review on estimation of stochastic differential equations for pharmacokinetic - pharmacodynamic models. *Adv. Drug Delivery Rev.* **65** (2013) 929–939.

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

[7] V. Genon-Catalot, T. Jeantheau and C. Laredo, Parameter estimation for discretely observed stochastic volatility models. *Bernoulli* **5** (1999) 855–872.

[8] M. Kessler, A. Lindner and M. Sorensen, Statistical Methods for Stochastic Differential Equations. Edited by M. Kessler, A. Lindner and M. Sorensen. CRC Press (2012).

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

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

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

[12] T. Sweeting, Uniform asymptotic normality of the maximum likelihood estimator. *Ann. Stat.* **8** (1980) 1179–1403.