

# Nonparametric Bayesian volatility estimation for gamma-driven stochastic differential equations

DENIS BELOMESTNY<sup>1,2,a</sup>, SHOTA GUGUSHVILI<sup>3,b</sup>, MORITZ SCHAUER<sup>4,c</sup> and PETER SPREIJ<sup>5,6,d</sup>

<sup>1</sup>Faculty of Mathematics, Duisburg-Essen University, Thea-Leymann-Str. 9, D-45127 Essen, Germany

<sup>2</sup>Faculty of Computer Sciences, HSE University. <sup>a</sup>[denis.belomestny@uni-due.de](mailto:denis.belomestny@uni-due.de)

<sup>3</sup>Wageningen University & Research, Postbus 16, 6700 AA Wageningen, The Netherlands.

<sup>b</sup>[shota@yesdatasolutions.com](mailto:shota@yesdatasolutions.com)

<sup>4</sup>Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, SE-412 96 Göteborg, Sweden. <sup>c</sup>[smoritz@chalmers.se](mailto:smoritz@chalmers.se)

<sup>5</sup>Korteweg-de Vries Institute for Mathematics, Universiteit van Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands

<sup>6</sup>Institute for Mathematics, Astrophysics and Particle Physics, Radboud University, Nijmegen, Netherlands.

<sup>d</sup>[spreij@uva.nl](mailto:spreij@uva.nl)

We study a nonparametric Bayesian approach to estimation of the volatility function of a stochastic differential equation driven by a gamma process. The volatility function is modelled a priori as piecewise constant, and we specify a gamma prior on its values. This leads to a straightforward procedure for posterior inference via an MCMC procedure. We give theoretical performance guarantees (minimax optimal contraction rates for the posterior) for the Bayesian estimate in terms of the regularity of the unknown volatility function. We illustrate the method on synthetic and real data examples.

**Keywords:** Gamma process; nonparametric Bayesian estimation; stochastic differential equation

## 1. Introduction

### 1.1. Problem formulation

The goal of the present paper is Bayesian nonparametric estimation of the positive local scale function  $\sigma$  that appears in the Lévy-driven stochastic differential equation

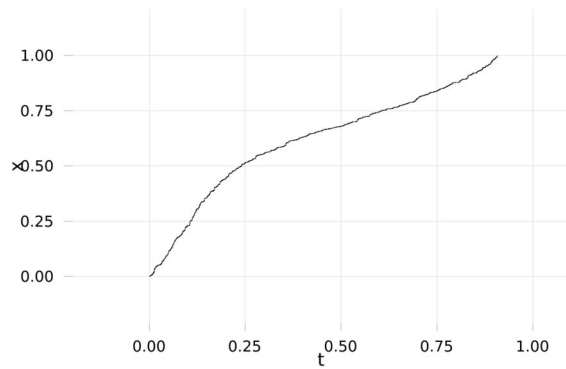
$$dX_t = \sigma(X_{t-}) dL_t, X_0 = 0, \tag{1}$$

from observations of the solution  $X$ . Here  $L$  is a gamma process with  $L_0 = 0$  and therefore  $L$  is a subordinator, i.e. a stochastic process with monotonous sample paths. Furthermore,  $L$  has a Lévy measure  $\nu$  admitting the Lévy density

$$\nu(x) = \alpha x^{-1} \exp(-\beta x), x > 0, \tag{2}$$

where  $\alpha$  and  $\beta$  are two positive constants. The process  $L$  has independent increments, and  $L_t - L_s$  has a  $\text{Gamma}(\alpha(t-s), \beta)$  distribution for  $t > s$  with shape parameter  $\alpha$  and scale parameter  $1/\beta$ , as defined in Section 1.6.

Under the assumption that the function  $\sigma$  (in view of financial applications we refer to it as volatility function) is measurable and satisfies a linear growth condition, it has been shown in Belomestny et al. (2021) that Equation (1) admits a weak solution that is unique in law. Under the stronger condition



**Figure 1.** Observation  $X^n$ ,  $n = 200$ , generated from (1) with the Euler scheme with volatility  $\sigma_0$  given in (3). Parameters of the driving gamma process are  $\alpha = 1$ ,  $\beta = 1$ .

that  $\sigma$  is Lipschitz continuous, Equation (1) even has a unique strong solution, see (Protter, 2004, Theorem V.6). Note that  $X$  is a Markov process.

**Example 1.** To get an impression of how observations of  $X$  look like, we take  $X$  to solve the Lévy SDE (1) with the volatility function

$$\sigma_0(x) = \frac{1}{500} \left( \frac{3}{2} + \sin(2\pi x) \right). \quad (3)$$

For the driving gamma process  $L$  we take parameters  $\alpha = 1$ ,  $\beta = 1$ . We used the Euler scheme to generate a single trajectory to serve as observation. We simulated a path traversing the unit interval on a time grid with time step-size  $\delta t = 0.0001$  starting in  $X_0 = 0 = b_0$ . In Section 2.2 we will introduce bins with boundaries  $b_k$  and their hitting times  $\tau_k$ . Approximate hitting times were obtained setting  $\tau'_k = \inf\{i\delta t : X_{i\delta t} \geq b_k\}$ . Figure 1 shows the sample path until the hitting time  $\tau'_K$  of  $b_K = 1$ .

## 1.2. Motivation

Gamma processes, that form a special class of Lévy processes (see, e.g., Kyprianou (2014)), are a fundamental modelling tool in several fields, e.g. reliability (see van Noortwijk (2009)) and risk theory (see Dufresne, Gerber and Shiu (1991)). Since the driving gamma process  $L$  in (1) has non-decreasing sample paths and the volatility function  $\sigma$  is non-negative, also the process  $X$  has non-decreasing sample paths. Such processes find applications across various fields. A reliability model as in (1) has been thoroughly investigated from a probabilistic point of view in Wenocur (1989b), and constitutes a far-reaching generalisation of a basic gamma model. Furthermore, non-decreasing processes are ideally suited to model revenues from an innovation: in Chance, Hillebrand and Hilliard (2008), the authors study the question of pricing options on movie box office revenues that are modelled through a gamma-like stochastic process. Another potential application is in modelling the evolution of forest fire sizes over time, as in Reed and McKelvey (2002).

Any practical application of the model (1) would require knowledge of the volatility function  $\sigma$ , that has to be inferred from observations on the process  $X$ . In this paper, we will approach estimation of  $\sigma$  nonparametrically. The latter comes in handy when no apparent functional form for the volatility is available, which usually is the case in practice. Such an approach reduces the risk of model misspecification and allows a honest representation of uncertainties in inferential conclusions (see Silverman

(1986) and Müller and Mitra (2013)). Recent years have witnessed a tremendous growth of interest and rapid advances in nonparametric Bayesian methods (see, e.g., two monographs Ghosal and van der Vaart (2017) and Müller et al. (2015)). Coherence, elegance and automatic uncertainty quantification are some of the widely acknowledged attractive features of a Bayesian approach to statistics. Hence our decision to follow a Bayesian method in this paper. However, we also note that with a careful choice of a prior a nonparametric Bayesian method enjoys very favourable frequentist properties; see, e.g., Ghosal and van der Vaart (2017).

On the theoretical side, we are ultimately interested in asymptotic properties of our Bayesian procedure for estimating the volatility function  $\sigma$ . To that end we need a sufficiently rich set of observations and this can be accomplished by the scaled observation process in (4) below, instead of  $X$  satisfying (1). So we consider the process  $X^n$  given as the solution to

$$dX_t^n = \frac{1}{n} \sigma(X_{t-}^n) dL_t, X_0^n = 0. \tag{4}$$

The scaling factor  $\frac{1}{n}$  causes for large values of  $n$  a ‘slow growth’ of the process  $X^n$  and ‘long times’ to reach certain levels. We will thus assume that  $X^n$  is observed on a long time interval  $[0, T^n]$ , where  $T^n \rightarrow \infty$ . Later we will specify  $T^n$  and we will see that  $T^n$  grows roughly proportionally with  $n$ . Asymptotic results will be derived for  $n \rightarrow \infty$ . The setup above for having a rich set of observations allows for a different, but equivalent description of the model. Let  $Y_t^n = X_{nt}^n$ . Then  $Y^n$  satisfies the SDE

$$Y_t^n = \int_0^t \sigma(Y_{s-}^n) dL_s^n, \tag{5}$$

where  $L_t^n = \frac{L_{nt}}{n}$ . Note that  $L^n$  is again a gamma process with a Lévy density

$$v_n(x) = \frac{n\alpha}{x} \exp(-n\beta x),$$

and (for  $t > s \geq 0$ )  $L_t^n - L_s^n$  has a  $\text{Gamma}(n\alpha(t - s), n\beta)$  distribution. Here we see a smoother behaviour of  $L^n$  and hence of  $Y^n$  for growing  $n$ . In fact, it can be shown that  $L^n$  weakly converges to the function  $L^\infty$  given by  $L_t^\infty = \frac{\alpha}{\beta}t$ , and as a consequence  $Y^n$  should then converge to a deterministic limit as well. This behaviour of the process  $Y^n$  bears some similarity to the properties of a diffusion process  $Y^\varepsilon$  with small diffusion coefficient ( $W$  is a Wiener process),

$$Y_t^\varepsilon = \int_0^t a(Y_s^\varepsilon) ds + \varepsilon W_t,$$

which also has a deterministic limit as  $\varepsilon \rightarrow 0$ . The similarity becomes more pronounced, if one writes the semimartingale decomposition of  $Y^n$ ,

$$Y_t^n = \frac{\alpha}{\beta} \int_0^t \sigma(Y_{s-}^n) ds + M_t^n,$$

where, under appropriate conditions, the (local) martingale  $M^n$  vanishes for  $n \rightarrow \infty$ . This conceptual similarity with drift estimation in a well-known diffusion model and the statistical problem there, estimation of the function  $a$ , should be understood as a mathematical motivation for our study. Although it is possible to present all results that follow in terms of properties of the process  $Y^n$ , we opted to give them for  $X^n$ .

### 1.3. Literature overview

Statistical inference for Lévy driven SDEs is an active research with many contributions. Let us mention recent works by Gushchin, Pavlyukevich and Ritsch (2019), Jasra, Kamatani and Masuda (2019), Jasra et al. (2011), Uehara (2019b) and Eguchi and Uehara (2020). Such models are popular in finance and econometrics, see e.g. Todorov (2011). In this paper we study the problem of Bayesian inference for the volatility coefficient of a Lévy-driven SDE.

Nonparametric Bayesian literature on inference in the model (1) is non-existent, but somewhat related problems have been considered in several papers. Thus, Belomestny et al. (2019) study a nonparametric Bayesian approach to estimation of the Lévy measure for Lévy processes with monotonous sample paths (subordinators), while Gugushvili, van der Meulen and Spreij (2015, 2018) and Gugushvili, Mariucci and van der Meulen (2020) investigate the same problem for compound Poisson processes. There exists also a limited amount of work on nonparametric Bayesian volatility estimation in diffusion models, e.g. Batz, Ruttor and Opper (2018) and Nickl and Söhl (2017), but a paper that is most related to the present one is Gugushvili et al. (2019). Finally, Koskela, Spanò and Jenkins (2019) is a theoretical contribution, where frequentist consistency of a Bayesian approach to inference in jump-diffusion models is established.

### 1.4. Our contribution

Our work is the first contribution to nonparametric Bayesian volatility estimation for Lévy-driven SDEs. The method we propose is easy to understand, and leads to good practical results in synthetic and real data examples. We expect it to open up new research directions in inference for stochastic differential equations driven by jump processes, both from the practical and theoretical point of views. Our approach is based on a *piecewise constant approximation* with a proper prior on the corresponding parameters. We show the contraction of the posterior for Hölder continuous volatility coefficients and propose a MCMC procedure for sampling.

**Example 2.** We continue Example 1 and graphically illustrate our inferential results. In agreement with the notation pertaining to the asymptotic regime detailed in (4) we took  $X^n$  as the solution to the Lévy SDE (1) with the volatility function

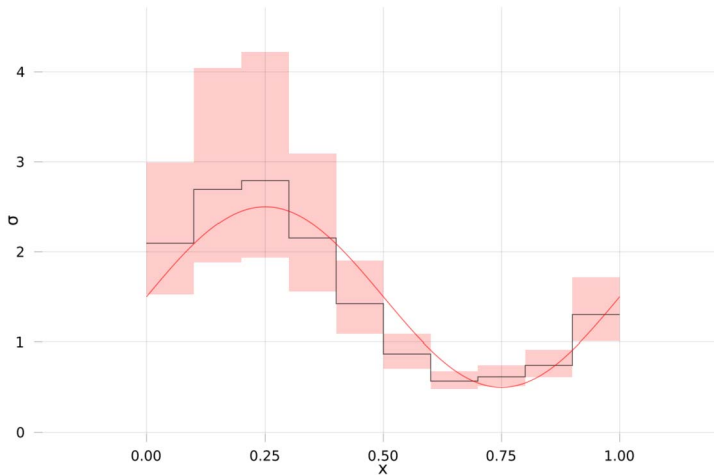
$$\sigma_0(x) = \frac{3}{2} + \sin(2\pi x) \quad (6)$$

and scaled it with  $1/n$ ,  $n = 500$ . For details of our Bayesian procedure we refer to Section 2.2, in the present example this comes down to the following.

We have partitioned the unit interval into 10 bins setting  $b_k = k\delta x$ ,  $\delta x = 0.1$ ,  $k = 1, \dots, 10$ , and used the piecewise constant prior of the form  $\sigma(x) = \sum_k \xi_k \mathbf{1}_{B_k}(x)$ , where  $\xi_1, \dots, \xi_{10}$  are i.i.d. random variables with an inverse gamma distribution. The posterior is given in closed form by (10). Figure 2 contrasts the corresponding marginal 90%-posterior credible bands for  $\sigma$  with the true volatility function  $\sigma_0$  from (6).

### 1.5. Structure of the paper

The paper is organised as follows: in Section 2 we describe in detail our Bayesian method for volatility estimation. In Section 3 we present asymptotic properties of the posterior distribution when the



**Figure 2.** Red: true volatility  $\sigma_0$  as a function of  $x$ . Shaded: marginal 90 %-posterior credible band for the piecewise constant posterior  $\sigma$ . Black: marginal posterior median.

volatility function  $\sigma$  is piecewise constant, whereas in Section 4 we present such properties when the volatility function is Hölder continuous. The contraction rate we obtain in the latter case is shown to be minimax optimal. Real data examples are considered in Section 5. In Section 6 we consider some extensions and variations of our approach, in particular paying attention to the setting of discrete time observations, and propose a Metropolis-Hastings approach for simulating gamma process bridges.

### 1.6. Notation

We denote the gamma distribution with shape parameter  $a > 0$  and rate parameter  $b > 0$  (hence scale  $1/b$ ) by  $\text{Gamma}(a, b)$ . Recall that its density is given by

$$x \mapsto \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, \quad x > 0,$$

where  $\Gamma$  is the gamma function. The inverse gamma distribution with shape parameter  $a > 0$  and scale parameter  $b > 0$  will be denoted by  $\text{IG}(a, b)$ . The corresponding density is

$$x \mapsto \frac{b^a}{\Gamma(a)} x^{-a-1} e^{-b/x}, \quad x > 0,$$

and its expectation and variance are  $\frac{b}{a-1}$  and  $\frac{b^2}{(a-1)^2(a-2)}$ , respectively. Following a standard Bayesian convention, we will often use lowercase letters to write random variables. Conditioning of  $x$  on  $y$  will be denoted by  $x | y$ .

## 2. Bayesian approach

To compute a posterior distribution, a likelihood ratio is needed. In this section we study likelihood ratios and existence of a weak solution to Equation (1).

### 2.1. Likelihood

Let  $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$  be a filtered probability space and let  $(L_t)_{t \geq 0}$  be a gamma process adapted to  $\mathbb{F}$ , whose Lévy measure admits the density  $\nu$  given by (2). Assume that  $X$  is a (weak) solution to (1), and assume that  $X$  is observed on an interval  $[0, T]$ . We denote by  $\mathbb{P}_T^\sigma$  (a probability measure on  $\mathcal{F}_T^X = \sigma(X_t, 0 \leq t \leq T)$ ) its law. In agreement with this notation we let  $\mathbb{P}_T^1$  be the law of  $X$  when  $\sigma \equiv 1$ , in which case  $X_t = L_t, t \in [0, T]$ . The measure  $\mathbb{P}_T^1$  will serve as a reference measure. The choice  $\sigma = 1$  for obtaining a reference measure is natural, but also arbitrary. Many other choices for the function  $\sigma$  are conceivable, in particular other constant functions. In Belomestny et al. (2021) the following proposition has been proven.

**Proposition 3.** *Let  $\sigma$  be a positive locally bounded measurable function on  $[0, \infty)$  such that (1) admits a weak solution that is unique in law. Assume moreover that  $\sigma$  is lower bounded by a strictly positive constant. Let  $T$  be a finite (stopping) time  $T > 0$ . Then the laws  $\mathbb{P}_T^\sigma$  and  $\mathbb{P}_T^1$  are equivalent on  $\mathcal{F}_T^X$  and the corresponding Radon-Nikodym derivative  $Z_T$  has the explicit representation:*

$$Z_T = \exp \left( \int_0^T \int_0^\infty \log Y(t, x) \mu^X(dx, dt) - \int_0^T \int_0^\infty (Y(t, x) - 1) \nu(x) dx dt \right), \tag{7}$$

where both double integrals are a.s. finite.

**Remark 4.** For statistical inference on  $\sigma$  one needs a realisation of the random quantity  $Z_T$ , induced by an observed realisation of  $X$ . The realisation of  $Z_T$  is then simply obtained by evaluation of the integrals along a path of  $X$ . This causes no difficulties as the integrals in  $Z_T$  are defined pathwise. We will need (in Corollary 5 below) the result for positive piecewise constant functions  $\sigma$ , and we will see there that for a realisation of  $Z_T$  one only needs the observed values of the  $X_{\tau_k^n}^n$ . The stopping times  $\tau_k^n$  will be specified later.

### 2.2. Piecewise constant volatility prior

In our non-parametric Bayesian approach, we will a priori model  $\sigma$  as a piecewise constant function. Namely,

$$\sigma(x) = \sum_{k=1}^K \xi_k \mathbf{1}_{B_k}(x) \tag{8}$$

for bins  $B_1 = [0, b_1], B_k = (b_{k-1}, b_k], k = 2, \dots, K - 1$ , and  $B_K = (b_{K-1}, b_K)$ , with appropriately chosen increasing sequence of bin endpoints  $\{b_k\}$  and the bin number  $K$ . The bins  $B_k$ 's form a partition of the positive halfline  $[0, \infty)$ . The  $\{\xi_k\}$  are positive numbers (later on positive random variables). Although we use (8) for our model, we emphasize that the 'true'  $\sigma$  does not need to be piecewise constant. As a final remark we note that when  $\sigma$  is given by (8), (4) still has a unique solution, obtained as concatenation of stopped gamma processes.

**Corollary 5.** *Suppose that  $X^n$  is given by (4) with  $\sigma$  given by (8). Let  $\tau_k^n = \inf\{t \geq 0 : X_t^n \geq b_k\}, k = 1, \dots, K$ , and write  $T = \tau_K^n$ . Then*

$$\frac{d\mathbb{P}_T^\sigma}{d\mathbb{P}_T^1} = \exp \left\{ \beta \sum_{k=1}^K \left[ 1 - n\xi_k^{-1} \right] (X_{\tau_k^n}^n - X_{\tau_{k-1}^n}^n) - \alpha \sum_{k=1}^K (\tau_k^n - \tau_{k-1}^n) \log(\xi_k/n) \right\}.$$

**Proof.** We have now  $Y(t, x) = \exp(-\beta x(\frac{n}{\sigma(X_{t-}^n)} - 1))$ . It follows from Proposition 3 and Remark 4 that

$$\begin{aligned} \frac{d\mathbb{P}_T^\sigma}{d\mathbb{P}_T^1} &= \exp \left\{ \int_{(0,T]} \int_{(0,\infty)} \left\{ x\beta \left[ 1 - \frac{n}{\sigma(X_{s-}^n)} \right] \right\} \mu^X(ds, dx) \right. \\ &\quad \left. - \int_0^T \int_{(0,\infty)} \left( e^{x\beta \left[ 1 - \frac{n}{\sigma(X_{s-}^n)} \right]} - 1 \right) \nu(x) dx ds \right\} = \exp(I_1 - I_2), \end{aligned}$$

where

$$\begin{aligned} I_1 &= \int_{(0,T]} \int_{(0,\infty)} \left\{ x\beta \left[ 1 - \frac{n}{\sigma(X_{s-}^n)} \right] \right\} \mu^X(ds, dx) \\ &= \sum_{k=1}^K \int_{(\tau_{k-1}^n, \tau_k^n]} \int_{(0,\infty)} \left\{ x\beta \left[ 1 - \frac{n}{\sigma(X_{s-}^n)} \right] \right\} \mu^X(ds, dx) \\ &= \sum_{k=1}^K \beta \left[ 1 - \frac{n}{\xi_k} \right] \int_{(\tau_{k-1}^n, \tau_k^n]} \int_{(0,\infty)} x \mu^X(ds, dx) = \beta \sum_{k=1}^K \left[ 1 - \frac{n}{\xi_k} \right] (X_{\tau_k^n}^n - X_{\tau_{k-1}^n}^n), \end{aligned}$$

since  $\sigma(X_{s-}^n) = \xi_k$  for  $s \in (\tau_{k-1}^n, \tau_k^n]$  if  $\tau_{k-1}^n < \tau_k^n$ . But the above expression is also valid if  $\tau_{k-1}^n = \tau_k^n$ , since then  $X_{\tau_k^n}^n - X_{\tau_{k-1}^n}^n = 0$ . Furthermore, by similar reasoning,

$$I_2 = \sum_{k=1}^K (\tau_k^n - \tau_{k-1}^n) \int_{\mathbb{R}_+} \frac{\alpha}{x} \left[ e^{-x\beta n \xi_k^{-1}} - e^{-\beta x} \right] dx.$$

We need an intermediate result. For any  $c > 0$  one has  $\int_{\mathbb{R}_+} \frac{1}{x} [e^{-x\beta/c} - e^{-x\beta}] dx = \int_{\mathbb{R}_+} \frac{1}{x} [e^{-x/c} - e^{-x}] dx$  and

$$\begin{aligned} \int_{\mathbb{R}_+} \frac{1}{x} [e^{-x/c} - e^{-x}] dx &= \int_0^\infty \int_1^{1/c} -e^{-ux} du dx \\ &= \int_1^{1/c} \int_0^\infty -e^{-ux} dx du \\ &= \int_1^{1/c} -\frac{1}{u} du = \log c. \end{aligned}$$

Application of this result to  $c = \xi_k/n$  gives  $I_2 = \alpha \sum_k (\tau_k^n - \tau_{k-1}^n) \log(\xi_k/n)$ . □

### 3. Observations with piecewise constant volatility

In this section we will provide contraction rates for the posterior distribution in our Bayesian setup, when the true volatility function  $\sigma$  is piecewise constant. So, we consider the process  $X^n$  whose true distribution results from  $X^n$  being the solution to

$$dX_t^n = \frac{1}{n} \sum_{k=1}^K \sigma_k \mathbf{1}_{B_k}(X_{t-}^n) dL_t, \tag{9}$$

where the constants  $\sigma_k$  are assumed to be strictly positive. So  $\sigma(x) = \sum_k \sigma_k \mathbf{1}_{B_k}(x)$ , in agreement with (4). We assume that the process  $X^n$  is observed until the time where it surpasses the fixed and known level  $b_K$ . We aim at estimating  $\sigma$ , or equivalently, the sequence  $\{\sigma_k\}$ , in a consistent way. In order to accomplish this, the observed process has to spend large times in the bins  $B_k$ , which is effectively the result of the scaling factor  $\frac{1}{n}$  for large  $n$  in (9). In fact, one has that the  $\tau_k^n - \tau_{k-1}^n$  are roughly proportional to  $n$  for  $n \rightarrow \infty$ , see Proposition 8 below for the precise result. In our Bayesian approach we model the  $\sigma_k$  as independent random variables  $\xi_k$  and we take inverse gamma distributions as a prior for each of them, that is  $\xi_k \sim \text{IG}(\alpha_k, \beta_k)$ . Therefore, we have to extend the original probability space to carry the  $\xi_k$  as well, taking into account that  $L$  and the  $\sigma_k$  have to be independent. We can then use Corollary 5, where  $\mathbb{P}_T^\sigma$  is to be interpreted as the conditional law of  $X^n$  on  $[0, T]$  given the  $\xi_k$ .

Let  $x_1, \dots, x_K$  denote realisations of  $\xi_1, \dots, \xi_K$  and let  $X^n$  stand for the path  $X_t^n, t \in [0, T]$ . With  $T = \tau_K^n$ , we then have from Corollary 5 that the posterior joint density of  $(\xi_1, \dots, \xi_K)$  is given by

$$\Pi(x_1, \dots, x_K \mid X^n) \propto \prod_{k=1}^K \exp\left(-n\beta(X_{\tau_k^n}^n - X_{\tau_{k-1}^n}^n) + \beta_k\right) x_k^{-1} x_k^{-\alpha(\tau_k^n - \tau_{k-1}^n) - \alpha_k - 1}.$$

It follows that the  $\xi_k$  are independent under the posterior distribution, and

$$\xi_k \mid X^n \sim \text{IG}(\alpha \Delta \tau_k^n + \alpha_k, n\beta \Delta X_{\tau_k^n} + \beta_k), \tag{10}$$

where  $\Delta \tau_k^n = \tau_k^n - \tau_{k-1}^n, \Delta X_{\tau_k^n} = X_{\tau_k^n}^n - X_{\tau_{k-1}^n}^n$ .

### 3.1. Result on the overshoot

To prepare for our first main result, we need a property of the overshoots of  $X^n$ . The overshoot  $\text{POT}_k^n$  is defined as  $\text{POT}_k^n := X_{\tau_k^n}^n - b_k \geq 0$  for  $k = 0, \dots, K$ . Note that  $\text{POT}_0^n = 0$ .

**Lemma 6.** *Let  $\sigma_K^* = \max\{\sigma_1, \dots, \sigma_K\}$  and  $\delta > 0$ . The probability  $\mathbb{P}(\text{POT}_k^n < \delta)$  satisfies the following bound,*

$$\mathbb{P}(\text{POT}_k^n < \delta) \geq \frac{b_K}{b_K + \delta} (1 - \exp(-n\delta\beta/\sigma_K^*)). \tag{11}$$

If  $\delta_n \rightarrow 0$  such that  $n\delta_n \rightarrow \infty$ , then  $\mathbb{P}(\text{POT}_k^n > \delta_n) \rightarrow 0$ .

**Proof.** Since  $\text{POT}_0^n = 0$ , inequality (11) is trivially true for  $k = 0$ , so we let  $k \in \{1, \dots, K\}$  and  $0 < a < b_k$ . We start by considering the conditional probability  $\mathbb{P}(\text{POT}_k^n < \delta \mid X_{\tau_{k-1}^n}^n = b_k - a)$  and we claim that

$$\mathbb{P}(\text{POT}_k^n < \delta \mid X_{\tau_{k-1}^n}^n = b_k - a) = \frac{\nu_k^n([a, a + \delta])}{\nu_k^n([a, \infty))}, \tag{12}$$

where  $\nu_k^n$  is the Lévy measure of the process  $\frac{\sigma_k}{n}L$ . Note that  $\nu_k^n([a, b]) = \nu([an/\sigma_k, bn/\sigma_k])$  with  $\nu$  the Lévy measure of  $L$ . To see (12), we argue as follows. First we have

$$\begin{aligned} \mathbb{P}(\text{POT}_k^n < \delta \mid X_{\tau_{k-1}^n}^n = b_k - a) &= \mathbb{P}(\Delta X_{\tau_k^n}^n < b_k - X_{\tau_{k-1}^n}^n + \delta \mid X_{\tau_{k-1}^n}^n = b_k - a) \\ &= \mathbb{P}(\Delta X_{\tau_k^n}^n < a + \delta \mid X_{\tau_{k-1}^n}^n = b_k - a). \end{aligned}$$



We will now use (Kyprianou, 2014, Theorem 5.6), that states in terms of densities (which exist here) and in notation adapted to our situation that the random vector  $(X_{\tau_k}^n - b_k, b_k - X_{\tau_k-}^n)$  has joint density (for  $x > 0, 0 < y < b_k$ )

$$f_{X_{\tau_k}^n - b_k, b_k - X_{\tau_k-}^n}(x, y) = f_U(b_k - y)f_V(x + y),$$

where  $f_U$  is the density of the potential function  $U$  (in Kyprianou’s terminology; it will turn out that the precise form of  $f_U$  is not relevant in our context) and  $f_V$  the density of the Lévy measure of the process, in our case  $\nu_k^n$ , as we effectively deal with properties of the process  $\frac{\sigma_k}{n}L$ . It follows that  $b_k - X_{\tau_k-}^n$  has marginal density

$$f_{b_k - X_{\tau_k-}^n}(y) = f_U(b_k - y)\nu_k^n([y, \infty)).$$

Using the change of variables  $(X_{\tau_k}^n - b_k, b_k - X_{\tau_k-}^n) \rightarrow (\Delta X_{\tau_k}^n, b_k - X_{\tau_k-}^n)$  one gets from the transformation formula that  $(\Delta X_{\tau_k}^n, b_k - X_{\tau_k-}^n)$  has density

$$f_{\Delta X_{\tau_k}^n, b_k - X_{\tau_k-}^n}(x, y) = f_{X_{\tau_k}^n - b_k, b_k - X_{\tau_k-}^n}(x - y, y) = f_U(b_k - y)f_V^n(x),$$

for  $x > y, 0 < y < b_k$  and zero elsewhere. It follows that the conditional density of  $\Delta X_{\tau_k}^n$  given  $b_k - X_{\tau_k-}^n = a$  is, for  $x > a$ ,

$$f_{\Delta X_{\tau_k}^n | b_k - X_{\tau_k-}^n = a}(x) = \frac{f_U(b_k - a)f_V^n(x)}{f_U(b_k - a)\nu_k^n([a, \infty))} = \frac{f_V^n(x)}{\nu_k^n([a, \infty))}.$$

Hence,

$$\begin{aligned} \mathbb{P}(\Delta X_{\tau_k}^n < a + \delta | X_{\tau_k-}^n = b_k - a) &= \int_a^{a+\delta} \frac{f_V^n(x)}{\nu_k^n([a, \infty))} dx \\ &= \frac{\nu_k^n(a, a + \delta)}{\nu_k^n([a, \infty))}, \end{aligned}$$

which proves (12). Next we show that  $g(a) := \frac{F(a+\delta) - F(a)}{1 - F(a)}$  is decreasing as a function of  $a$  for  $a \geq \varepsilon$ , where  $\varepsilon > 0$  is sufficiently small. Here  $F$  is the ‘truncated distribution function’ of  $\nu$ ,  $F(x) = \nu((\varepsilon, x])$  for  $x \geq \varepsilon$ . Without loss of generality, we scale  $F$  such that  $F(\infty) = 1$ . Then, denoting by  $f$  the derivative of  $F$ , one has

$$\begin{aligned} \frac{d}{da} \frac{F(a + \delta) - F(a)}{1 - F(a)} &= \frac{(1 - F(a))f(a + \delta) - (1 - F(a + \delta))f(a)}{\underbrace{(1 - F(a))^2}_{=: c > 0}} \\ &= \frac{1}{c} \int_a^\infty f(x)f(a + \delta)dx - \frac{1}{c} \int_{a+\delta}^\infty f(x)f(a)dx \\ &= \frac{1}{c} \int_a^\infty (f(x)f(a + \delta) - f(x + \delta)f(a))dx. \end{aligned}$$

With  $v(dx) = f(x)dx = \frac{\alpha}{x} \exp(-\beta x)dx$  for every  $\alpha$  and  $\beta$  (which we later replace with  $\frac{n\beta}{\sigma_k}$ ),

$$\begin{aligned} & \frac{d}{da} \mathbb{P}(\text{POT}_k^n < \delta \mid X_{\tau_k^n}^n = b_k - a) \\ &= \frac{1}{c} \int_a^\infty \left( \frac{1}{(a + \delta)x} - \frac{1}{(x + \delta)a} \right) \exp(-\beta(x + \delta + a)) dx \\ &= \frac{1}{c} \int_a^\infty \underbrace{\left( \frac{\delta(a - x)}{ax(x + \delta)(a + \delta)} \right)}_{\leq 0} \exp(-\beta(x + \delta + a)) dx \leq 0. \end{aligned}$$

Noting that  $g(a)$  is defined for any  $a$  away from zero, we now apply the decreasing behaviour of  $g$ ,  $g(a) \geq g(b_K)$  as  $a < b_k < b_K$ , (with  $\beta$  replaced with  $n\beta/\sigma_k$ ) to get the lower bound

$$\mathbb{P}(\text{POT}_k^n < \delta \mid X_{\tau_k^n}^n = b_k - a) = g(a) \geq g(b_K) = \frac{v_k^n([b_K, b_K + \delta])}{v_k^n([b_K, \infty))}. \tag{13}$$

Next we will obtain upper and lower bounds for the fraction  $q(n, a, k, \delta) = \frac{v_k^n([a, a + \delta])}{v_k^n([a, \infty))}$  for any  $a$  and  $k$ . We write the fraction as a fraction of integrals keeping in mind our model with volatility scaled by  $n$  and  $L$  a gamma process. We will provide bounds on  $v_k^n([a, b])$  with  $a < b \leq \infty$ . We use that for  $X_{t-}^n \in B_k$  we have to deal with the Lévy measure  $v_k^n$  with density

$$v_k^n(x) = \frac{\alpha}{x} \exp(-n\beta x/\sigma_k).$$

We compute

$$\begin{aligned} v_k^n([a, b]) &= \int_a^b \frac{\alpha}{x} \exp(-n\beta x/\sigma_k) dx \\ &= \int_{an\beta/\sigma_k}^{bn\beta/\sigma_k} \frac{\alpha}{y} \exp(-y) dy \\ &\leq \int_{an\beta/\sigma_k}^{bn\beta/\sigma_k} \frac{\alpha}{an\beta/\sigma_k} \exp(-y) dy \\ &= \frac{\alpha\sigma_k}{an\beta} (\exp(-an\beta/\sigma_k) - \exp(-bn\beta/\sigma_k)). \end{aligned}$$

Via a similar argument, we have a lower bound

$$v_k^n([a, b]) \geq \frac{\alpha\sigma_k}{bn\beta} (\exp(-an\beta/\sigma_k) - \exp(-bn\beta/\sigma_k)).$$

Hence, using the lower bound with  $b = a + \delta$  and the upper bound with  $b = \infty$ , we obtain

$$q(n, a, k, \delta) \geq \frac{a}{a + \delta} (1 - \exp(-n\delta\beta/\sigma_k)). \tag{14}$$

Applying (14) to (13), we obtain

$$\mathbb{P}(\text{POT}_k^n < \delta \mid X_{\tau_k^n}^n = b_k - a) \geq \frac{b_K}{b_K + \delta} (1 - \exp(-n\delta\beta/\sigma_k)).$$

As the above lower bound is a decreasing function of  $\sigma_k$ , we can make it smaller by replacing  $\sigma_k$  with  $\sigma_K^*$  and obtain

$$\mathbb{P}(\text{POT}_k^n < \delta \mid X_{\tau_k^n}^n = b_k - a) \geq \frac{b_K}{b_K + \delta} (1 - \exp(-n\delta\beta/\sigma_K^*)).$$

Write  $\mu_k^n$  for the distribution of  $X_{\tau_k^n}^n$ . Then the unconditional probability  $\mathbb{P}(\text{POT}_k^n < \delta)$  can be written as

$$\begin{aligned} \mathbb{P}(\text{POT}_k^n < \delta) &= \int_{[b_{k-1}, b_k)} \mathbb{P}(\text{POT}_k^n < \delta \mid X_{\tau_k^n}^n = x) \mu_k^n(dx) \\ &\geq \int_{[b_{k-1}, b_k)} \frac{b_K}{b_K + \delta} (1 - \exp(-n\delta\beta/\sigma_K^*)) \mu_k^n(dx) \\ &= \frac{b_K}{b_K + \delta} (1 - \exp(-n\delta\beta/\sigma_K^*)), \end{aligned}$$

which proves (11). The final assertion on  $\mathbb{P}(\text{POT}_k^n < \delta_n)$  for  $\delta_n \rightarrow 0$  immediately follows from (11).  $\square$

### 3.2. Posterior contraction rate

Statements on convergence in probability refer to the law  $\mathbb{P} = \mathbb{P}^{\sigma, n}$  of the process satisfying the SDE (4). We will denote the posterior distribution of the  $\xi_k$  by  $\Pi_n$  and posterior expectation and variance by  $\mathbb{E}_{\Pi_n}$  and  $\mathbb{V}_{\Pi_n}$ , respectively. Furthermore we write  $\Delta b_k := b_k - b_{k-1}$  for  $k \geq 1$ .

**Lemma 7.** *Let  $\delta_n > 0$  and  $G_k^n = \{\text{POT}_k^n > \delta_n\}$  for  $k \geq 0$ . Then it holds for  $t \geq 0$  and  $k \geq 1$  that*

$$\mathbb{P}\left(L_t < \frac{n(\Delta b_k - \delta_n)}{\sigma_k}\right) \mathbb{P}((G_{k-1}^n)^c) \leq \mathbb{P}(\Delta\tau_k^n > t) \leq \mathbb{P}\left(L_t < \frac{n\Delta b_k}{\sigma_k}\right).$$

**Proof.** Let  $t \geq 0$ . On the event  $\{\Delta\tau_k^n > t\}$  we have  $b_{k-1} \leq X_{\tau_{k-1}^n}^n < b_k$  and for  $s \in (\tau_{k-1}^n, \tau_k^n]$ , one has  $\sigma(X_{s-}^n) = \sigma_k$ . Hence

$$\begin{aligned} \mathbb{P}(\Delta\tau_k^n > t) &= \mathbb{P}(X_{t+\tau_{k-1}^n}^n < b_k) \\ &= \mathbb{P}\left(X_{\tau_{k-1}^n}^n + \frac{1}{n} \int_{\tau_{k-1}^n}^{t+\tau_{k-1}^n} \sigma(X_{s-}^n) dL_s < b_k\right) \\ &= \mathbb{P}\left(\sigma_k(L_{t+\tau_{k-1}^n}^n - L_{\tau_{k-1}^n}^n) < n(b_k - X_{\tau_{k-1}^n}^n)\right) \\ &= \mathbb{P}\left(L_t^{n,k} < \frac{n(\Delta b_k - \text{POT}_{k-1}^n)}{\sigma_k}\right), \end{aligned} \tag{15}$$

where  $L_t^{n,k} = L_{t+\tau_{k-1}^n}^n - L_{\tau_{k-1}^n}^n$ . Trivially, this implies the inequality

$$\mathbb{P}(\Delta\tau_k^n > t) \leq \mathbb{P}\left(L_t^{n,k} < \frac{n\Delta b_k}{\sigma_k}\right) = \mathbb{P}\left(L_t < \frac{n\Delta b_k}{\sigma_k}\right), \tag{16}$$

where the equality follows from the strong Markov property of  $L$  and stationarity of its increments. This gives the upper bound. Next we consider a lower bound for (15),

$$\begin{aligned} \mathbb{P}\left(L_t^{n,k} < \frac{n(\Delta b_k - \text{POT}_{k-1}^n)}{\sigma_k}\right) &\geq \mathbb{P}\left(\{L_t^{n,k} < \frac{n(\Delta b_k - \text{POT}_{k-1}^n)}{\sigma_k}\} \cap (G_{k-1}^n)^c\right) \\ &\geq \mathbb{P}\left(\{L_t^{n,k} < \frac{n(\Delta b_k - \delta_n)}{\sigma_k}\} \cap (G_{k-1}^n)^c\right). \end{aligned}$$

By independence of  $L_{t+\tau_{k-1}^n} - L_{\tau_{k-1}^n}$  and  $X_{\tau_{k-1}^n}^n$ , the strong Markov property of  $L$  and stationarity of its increments, the last probability is equal to

$$\mathbb{P}\left(L_t < \frac{n(\Delta b_k - \delta_n)}{\sigma_k}\right) \mathbb{P}((G_{k-1}^n)^c),$$

which is the desired lower bound. □

**Proposition 8.** Let  $\Delta \bar{\tau}_k^n = \frac{n\Delta b_k \beta}{\alpha \sigma_k}$ . For  $c_n \rightarrow \infty$  such that  $c_n n^{-\frac{1}{2}} \rightarrow 0$  it holds that

$$\mathbb{P}\left(\left(1 - \frac{c_n}{\sqrt{n}}\right) \Delta \bar{\tau}_k^n \leq \Delta \tau_k^n \leq \left(1 + \frac{c_n}{\sqrt{n}}\right) \Delta \bar{\tau}_k^n\right) \rightarrow 1.$$

**Proof.** We consider

$$\mathbb{P}\left(\Delta \tau_k^n \leq \frac{n\Delta b_k \beta}{\alpha \sigma_k} \left(1 + \frac{c_n}{\sqrt{n}}\right)\right) = 1 - \mathbb{P}\left(\Delta \tau_k^n > \frac{n\Delta b_k \beta}{\alpha \sigma_k} \left(1 + \frac{c_n}{\sqrt{n}}\right)\right)$$

and derive an upper bound for it using the lower bound for  $\mathbb{P}(\Delta \tau_k^n < t)$  in Lemma 7 and the fact that  $L_t$  has the  $\Gamma(\alpha t, \beta)$  distribution. One has, with  $t = u_k^n := \frac{n\Delta b_k \beta}{\alpha \sigma_k} \left(1 + \frac{c_n}{\sqrt{n}}\right)$ , and in view of the upper bound in Lemma 7,

$$\mathbb{P}(\Delta \tau_k^n > u_k^n) \leq \mathbb{P}\left(L_{u_k^n} < \frac{n\Delta b_k}{\sigma_k}\right).$$

Note that  $\mathbb{E}L_{u_k^n} = \frac{\alpha}{\beta} u_k^n$  and  $\text{Var} L_{u_k^n} = \frac{\alpha}{\beta^2} u_k^n$ . By the central limit theorem for gamma distributions,  $\hat{L}_{u_k^n} = \frac{L_{u_k^n} - \frac{\alpha}{\beta} u_k^n}{\sqrt{\frac{\alpha}{\beta^2} u_k^n}}$  has an asymptotic standard normal distribution. The probability on the right of the above display can be rewritten as

$$\mathbb{P}\left(\hat{L}_{u_k^n} < \frac{\frac{n\Delta b_k}{\sigma_k} - \frac{\alpha}{\beta} u_k^n}{\sqrt{\frac{\alpha}{\beta^2} u_k^n}}\right).$$

The term on the right-hand side of the inequality in parentheses is seen to be equal to  $-\frac{c_n}{\sigma_k} \sqrt{\frac{\beta \sigma_k \Delta b_k}{1 + \frac{c_n}{\sqrt{n}}}}$ . This term tends to minus infinity and so  $\mathbb{P}(\Delta \tau_k^n > u_k^n) \rightarrow 0$ . Next we consider  $\mathbb{P}(\Delta \tau_k^n > l_k^n)$  with  $l_k^n := \frac{n\Delta b_k \beta}{\alpha \sigma_k} \left(1 - \frac{c_n}{\sqrt{n}}\right)$  and show that this probability tends to one. First we use the lower bound in Lemma 7,

$$\mathbb{P}(\Delta \tau_k^n > l_k^n) \geq \mathbb{P}\left(L_{l_k^n} < \frac{n(\Delta b_k - \delta_n)}{\sigma_k}\right) \mathbb{P}((G_{k-1}^n)^c).$$

Note that  $\mathbb{P}((G_{k-1}^n)^c) \rightarrow 1$  by Lemma 6. As for the previous case, we look at the standardisation  $\hat{L}_{l_k^n}$  of  $L_{l_k^n}$  and consider

$$\mathbb{P}\left(L_{l_k^n} < \frac{n(\Delta b_k - \delta_n)}{\sigma_k}\right) \geq \mathbb{P}\left(\hat{L}_{l_k^n} < \frac{\frac{n(\Delta b_k - \delta_n)}{\sigma_k} - \frac{\alpha}{\beta} l_k^n}{\sqrt{\frac{\alpha}{\beta^2} l_k^n}}\right).$$

The right hand side of the inequality in parentheses is seen to be equal to

$$\frac{1}{\sigma_k} \sqrt{\frac{\beta \sigma_k}{1 - \frac{c_n}{\sqrt{n}}}} \left( c_n \sqrt{\Delta b_k} - \delta_n \sqrt{\frac{n}{\Delta b_k}} \right).$$

This term tends to plus infinity since  $\frac{c_n}{\sqrt{n}} \rightarrow 0$  and if we choose, as we do,  $\delta_n$  such that  $\delta_n \sqrt{n}$  is bounded, then  $\mathbb{P}(\Delta \tau_k^n > l_k^n) \rightarrow 0$ . □

**Lemma 9.** *Let  $c_n \rightarrow \infty$  such that  $c_n n^{-\frac{1}{2}} \rightarrow 0$ . Then the posterior mean squared error  $\mathbb{E}_{\Pi_n} (\xi_k - \sigma_k)^2 = O(\frac{c_n^2}{n})$  on a set of probability tending to one for all  $k = 1, \dots, K$ .*

**Proof.** Recall the inverse gamma posterior distribution of the  $\xi_k$  as given in (10). We will first consider the posterior bias

$$\text{bias}_k = \mathbb{E}_{\Pi_n} \xi_k - \sigma_k = \frac{n\beta \Delta X_{\tau_k^n}^n + \beta_k}{\alpha \Delta \tau_k^n + \alpha_k - 1} - \sigma_k.$$

and will provide upper and lower bounds for it. Let  $\gamma_k = \frac{\beta \Delta b_k}{\alpha \sigma_k}$  and  $\delta_n \rightarrow 0$  with  $n\delta_n \rightarrow \infty$ . In addition to the events  $G_k^n$  we need the sets

$$F_k^n = \{\Delta \tau_k^n > \gamma_k n(1 + \frac{c_n}{\sqrt{n}})\} \tag{17}$$

$$H_k^n = \{\Delta \tau_k^n < \gamma_k n(1 - \frac{c_n}{\sqrt{n}})\}. \tag{18}$$

It follows from Proposition 8 that  $\mathbb{P}(F_k^n) \rightarrow 0$  and  $\mathbb{P}(H_k^n) \rightarrow 0$ . Furthermore, we need  $\delta_n$  such that  $n\delta_n \rightarrow \infty$ . From Lemma 6 we obtain  $\mathbb{P}(G_k^n) \rightarrow 0$ . A first lower bound is given by

$$\text{bias}_k \geq \frac{n\beta(\Delta b_k - \text{POT}_{k-1}^n) + \beta_k}{\alpha \Delta \tau_k^n + \alpha_k - 1} - \sigma_k,$$

which we split on the events  $(F_k^n)^c \cap (G_{k-1}^n)^c$  and  $F_k^n \cup G_{k-1}^n$ . As  $\mathbb{P}(F_k^n \cup G_{k-1}^n) \rightarrow 0$ , we ignore the behaviour on  $F_k^n \cup G_{k-1}^n$ . On the event  $(F_k^n)^c \cap (G_{k-1}^n)^c$  we can further bound  $\text{bias}_k$  from below by

$$\frac{n\beta(\Delta b_k - \delta_n) + \beta_k}{n\beta \Delta b_k(1 + \frac{c_n}{\sqrt{n}}) / \sigma_k + \alpha_k - 1} - \sigma_k,$$

which we compute further as

$$\sigma_k \frac{-\delta_n - \frac{c_n \Delta b_k}{\sqrt{n}} + \frac{\beta_k - (\alpha_k - 1)\sigma_k}{n\beta}}{\Delta b_k(1 + \frac{c_n}{\sqrt{n}}) + \frac{(\alpha_k - 1)\sigma_k}{n\beta}}.$$

This is seen to be of order  $O(\frac{c_n}{\sqrt{n}})$  provided we choose  $\delta_n \sqrt{n}$  bounded, as we can do, still keeping  $n\delta_n \rightarrow \infty$ . Next we derive an upper bound for the bias. We trivially have

$$\text{bias}_k \leq \frac{n\beta(\Delta b_k + \text{POT}_k^n) + \beta_k}{\alpha\Delta\tau_k^n + \alpha_k - 1} - \sigma_k.$$

On the set  $(H_k^n)^c \cap (G_k^n)^c$ , which has probability tending to one, we can further upper bound this by

$$\sigma_k \frac{\delta_n + \frac{c_n \Delta b_k}{\sqrt{n}} + \frac{\beta_k - (\alpha_k - 1)\sigma_k}{n\beta}}{\Delta b_k(1 - \frac{c_n}{\sqrt{n}}) + \frac{(\alpha_k - 1)\sigma_k}{n\beta}},$$

which is again of order  $O(\frac{c_n}{\sqrt{n}})$  for  $\delta_n \sqrt{n}$  bounded. We conclude that  $\text{bias}_k$  is of order  $O(\frac{c_n}{\sqrt{n}})$  on an event with probability tending to one.

We move on to the posterior variance and derive upper and lower bounds on it. Recall

$$\text{Var}_{\Pi_n} \xi_k = \frac{(n\beta\Delta X_{\tau_k^n}^n + \beta_k)^2}{(\alpha\Delta\tau_k^n + \alpha_k - 1)^2(\alpha\Delta\tau_k^n + \alpha_k - 2)}.$$

On the set  $(F_k^n)^c \cap (G_{k-1}^n)^c$  this is larger than

$$\frac{(n\beta(\Delta b_k - \delta_n) + \beta_k)^2}{(n\beta\Delta b_k(1 + \frac{c_n}{\sqrt{n}})/\sigma_k + \alpha_k - 1)^2(n\beta\Delta b_k(1 + \frac{c_n}{\sqrt{n}})/\sigma_k + \alpha_k - 2)},$$

which is of order  $O(\frac{1}{n})$ , since  $\delta_n \rightarrow 0$ .

Next we give an upper bound for the posterior variance on the set  $(H_k^n)^c \cap (G_k^n)^c$ , which is

$$\frac{(n\beta(\Delta b_k + \delta_n) + \beta_k)^2}{(n\beta\Delta b_k(1 - \frac{c_n}{\sqrt{n}})/\sigma_k + \alpha_k - 1)^2(n\beta\Delta b_k(1 - \frac{c_n}{\sqrt{n}})/\sigma_k + \alpha_k - 2)}.$$

As for the lower bound, also this bound is of order  $O(\frac{1}{n})$ . Combining the properties of posterior bias and variance, we obtain the posterior mean squared error  $\mathbb{E}_{\Pi_n}(\xi_k - \sigma_k)^2$  is of order  $O(\frac{c_n^2}{n})$  with probability tending to one. □

**Remark 10.** The assertion of Lemma 9 can alternatively be formulated as  $\mathbb{E}_{\Pi_n}(\xi_k - \sigma_k)^2 = O_{\mathbb{P}}(\frac{c_n^2}{n})$ .

Our main result of this section is the following theorem, which says that the posterior contraction rate for estimating  $\sigma_k$  is  $n^{-1/2}$ .

**Theorem 11.** *Let  $(m_n)$  be any sequence of positive real numbers converging to infinity. Then, for  $n \rightarrow \infty$ ,*

$$\Pi_n\left(|\xi_k - \sigma_k| > \frac{m_n}{\sqrt{n}}\right) \rightarrow 0 \text{ in probability.}$$

**Proof.** It is sufficient to prove the assertion for  $m_n$  increasing to infinity slow enough. For such  $m_n$ , let  $c_n \rightarrow \infty$  such that  $\frac{c_n}{m_n} \rightarrow 0$ , for instance  $c_n = \sqrt{m_n}$ . Then also  $c_n n^{-\frac{1}{2}} \rightarrow 0$  and by Chebychev's

inequality and Lemma 9 we have for all  $k = 1, \dots, K$ , that

$$\Pi_n(|\xi_k - \sigma_k| > \frac{m_n}{\sqrt{n}}) \leq \frac{n}{m_n^2} \mathbb{E}_{\Pi_n} (\xi_k - \sigma_k)^2 = O(\frac{c_n^2}{m_n^2}),$$

with probability tending to one. □

### 4. Hölder continuous volatility

We consider again the process  $X^n$  satisfying (4), but the standing assumption in this section is that  $\sigma$  is Hölder continuous, that is, there are constants  $H \geq 0$  and  $0 < \lambda \leq 1$  such that for all  $x, y > 0$  it holds that  $|\sigma(x) - \sigma(y)| \leq H|x - y|^\lambda$ . Moreover,  $\sigma$  is assumed to be bounded from below by a positive constant  $\underline{\sigma}$ .

Here is some further notation for the present section.

- The number of bins and their width depend on  $n$ . So we write  $B_k^n = (b_{k-1}^n, b_k^n]$ ,  $k = 1, \dots, K$ ,  $K = K_n$ . We assume equidistant bins. Let  $b_K$  be the endpoint of the last bin, assumed to be fixed. We take the other bin boundaries  $b_k^n$  as  $b_k^n = \frac{b_K k}{K}$ ,  $k = 1, \dots, K$ . A given  $x \in (0, b_K]$  then belongs to bin  $B_k^n$  with  $k = k_n(x) = \lceil \frac{Kx}{b_K} \rceil$ .
- $\Delta b_k^n = b_k^n - b_{k-1}^n$ . Note that for  $x \in B_k^n$  it holds that  $|\sigma(x) - \sigma_k^n| \leq H(\Delta b_k^n)^\lambda$  for  $\sigma_k^n \in \{\sigma(b_{k-1}^n), \sigma(b_k^n)\}$ .
- If  $x \in B_k^n$ , we write  $\Delta \bar{\tau}_k^n(x) = \frac{n \Delta b_k^n \beta}{\alpha \sigma(x)}$ .
- Furthermore, we assume the number of bins  $K = K_n \asymp n^\kappa$  for  $0 < \kappa < 1$ . Then, given also the above assumption on the  $b_k^n$  and the definition of  $\bar{\tau}_k^n(x)$ , one has  $\Delta b_k^n \asymp n^{-\kappa}$  and  $\bar{\tau}_k^n(x) \asymp n^{1-\kappa}$  for all  $k$  and  $x \in B_k$ .

We observe the process  $X^n$  until it crosses the last bin. It follows from Proposition 14 below that the time this happens,  $\tau_K^n$ , is with high probability of order  $cn$  with  $c$  upperbounded by  $\frac{\beta b_K}{\alpha \underline{\sigma}}$ .

Although  $\sigma$  is continuous, we model it in our Bayesian approach as a piecewise constant, that is, as

$$\xi^n(x) = \sum_{k=0}^K \xi_k \mathbf{1}_{B_k^n}(x), \tag{19}$$

where the  $\xi_k$  are assigned the inverse gamma prior distributions as in Section 3.

#### 4.1. Behaviour of $\Delta \tau_k^n$

We need a variation on Lemma 6. Let  $\text{POT}_k^n = X_{\tau_k^n}^n - b_k$  and define for  $\delta_n > 0$  and  $k = 0, \dots, K$  the events  $G_k^n = \{\text{POT}_k^n > \delta_n\}$ . Note that  $G_0^n = \emptyset$ .

**Lemma 12.** *Let  $\sigma^* = \max\{\sigma(x) : 0 \leq x \leq b_K\}$  and  $\delta > 0$ . For all  $n$ ,  $k = 0, \dots, K$ , the probability  $\mathbb{P}(\text{POT}_k^n < \delta)$  satisfies the following lower bound*

$$\mathbb{P}(\text{POT}_k^n < \delta) \geq \frac{b_K}{b_K + \delta} (1 - \exp(-n\delta\beta/\sigma^*)). \tag{20}$$

If  $\delta_n \rightarrow 0$  such that  $n\delta_n \rightarrow \infty$ , then  $\mathbb{P}(G_k^n) = \mathbb{P}(\text{POT}_k^n > \delta_n) \rightarrow 0$ .

**Proof.** As in the proof of Lemma 6 we first look at  $\mathbb{P}(\text{POT}_k^n < \delta \mid X_{\tau_{k-1}^n}^n = b_k^n - a)$ . This probability depends on the values of  $\sigma$  on the bin  $B_{k-1}^n$ , but for all  $x$  (in  $B_{k-1}^n$ ) one has  $\sigma(x) \leq \sigma^*$ . Hence

$$\mathbb{P}(\text{POT}_k^n < \delta \mid X_{\tau_{k-1}^n}^n = b_k^n - a) \geq \mathbb{P}((\text{POT}_k^n)^* < \delta \mid X_{\tau_{k-1}^n}^n = b_k^n - a),$$

where  $(\text{POT}_k^n)^*$  is the overshoot belonging to the process that has constant volatility  $\sigma^*$  on  $[\tau_{k-1}^n, \tau_k^n)$ . We can therefore use the lower bounds of Lemma 6 to obtain (20).  $\square$

Next we derive bounds on  $\mathbb{P}(\Delta\tau_k^n > t)$ .

**Lemma 13.** *Let  $0 < x < b_K$  and let  $k = k_n(x)$  such that  $x \in B_k^n$ . Let  $\delta_n$  be a sequence of positive numbers. Then, for  $\Delta b_k^n$  small enough, it holds that*

$$\mathbb{P}\left(L_t < \frac{n(\Delta b_k^n - \delta_n)}{\sigma(x) + 2H(\Delta b_k^n)\lambda}\right) \mathbb{P}((G_{k-1}^n)^c) \leq \mathbb{P}(\Delta\tau_k^n > t) \leq \mathbb{P}\left(L_t < \frac{n\Delta b_k^n}{\sigma(x) - 2H(\Delta b_k^n)\lambda}\right).$$

**Proof.** Let  $t \geq 0$ . On the event  $\{\Delta\tau_k^n > t\}$  we have  $b_{k-1}^n \leq X_{\tau_{k-1}^n}^n < b_k^n$ . Hence

$$\begin{aligned} \mathbb{P}(\Delta\tau_k^n > t) &= \mathbb{P}(X_{t+\tau_{k-1}^n}^n < b_k^n) \\ &= \mathbb{P}\left(X_{\tau_{k-1}^n}^n + \frac{1}{n} \int_{\tau_{k-1}^n}^{t+\tau_{k-1}^n} \sigma(X_{s-}^n) dL_s < b_k^n\right) \\ &= \mathbb{P}\left(\int_{\tau_{k-1}^n}^{t+\tau_{k-1}^n} \sigma(X_{s-}^n) dL_s < n(b_k^n - X_{\tau_{k-1}^n}^n)\right) \\ &= \mathbb{P}\left(\int_{\tau_{k-1}^n}^{t+\tau_{k-1}^n} \sigma(X_{s-}^n) dL_s < n(\Delta b_k^n - \text{POT}_{k-1}^n)\right). \end{aligned}$$

Note that for  $s \in (\tau_{k-1}^n, \tau_k^n]$ , one has  $X_{s-}^n \in (b_{k-1}^n, b_k^n]$ . Let  $\sigma_k^n \in \{\sigma(b_{k-1}^n), \sigma(b_k^n)\}$ , then for  $x \in [b_{k-1}^n, b_k^n]$ , one has  $|\sigma(x) - \sigma_k^n| \leq H(\Delta b_k^n)\lambda$ . Hence, with  $L_t^{n,k} = L_{t+\tau_{k-1}^n}^n - L_{\tau_{k-1}^n}^n$ , one has

$$(\sigma_k^n - H(\Delta b_k^n)\lambda)L_t^{n,k} \leq \int_{\tau_{k-1}^n}^{t+\tau_{k-1}^n} \sigma(X_{s-}^n) dL_s \leq (\sigma_k^n + H(\Delta b_k^n)\lambda)L_t^{n,k}.$$

It follows that we have the double inequality

$$\mathbb{P}\left(L_t^{n,k} < \frac{n(\Delta b_k^n - \text{POT}_{k-1}^n)}{\sigma_k^n + H(\Delta b_k^n)\lambda}\right) \leq \mathbb{P}(\Delta\tau_k^n > t) \leq \mathbb{P}\left(L_t^{n,k} < \frac{n(\Delta b_k^n - \text{POT}_{k-1}^n)}{\sigma_k^n - H(\Delta b_k^n)\lambda}\right). \tag{21}$$

Trivially, this implies the inequality

$$\mathbb{P}(\Delta\tau_k^n > t) \leq \mathbb{P}(L_t^{n,k} < \frac{n\Delta b_k^n}{\sigma_k^n - H(\Delta b_k^n)\lambda}).$$

The latter probability is by stationarity of increments and the strong Markov property of  $L$  equal to

$$\mathbb{P}\left(L_t < \frac{n\Delta b_k^n}{\sigma_k^n - H(\Delta b_k^n)\lambda}\right).$$



If we take  $x \in B_k^n$ , we have for the latter probability the upper bound

$$\mathbb{P}\left(L_t < \frac{n\Delta b_k^n}{\sigma(x) - 2H(\Delta b_k^n)^\lambda}\right) \tag{22}$$

as desired. Next we consider the lower bound in (21),

$$p_k^n = \mathbb{P}\left(L_t^{n,k} < \frac{n(\Delta b_k^n - \text{POT}_{k-1}^n)}{\sigma_k^n + H(\Delta b_k^n)^\lambda}\right), \tag{23}$$

and proceed to give a further lower bound for it. With  $G_{k-1}^n = \{\text{POT}_{k-1}^n > \delta_n\}$  one has

$$p_k^n \geq \mathbb{P}\left(\left\{L_t^{n,k} < \frac{n(\Delta b_k^n - \text{POT}_{k-1}^n)}{\sigma_k^n + H(\Delta b_k^n)^\lambda}\right\} \cap (G_{k-1}^n)^c\right).$$

The latter probability is bounded from below by

$$\mathbb{P}\left(\left\{L_t^{n,k} < \frac{n(\Delta b_k^n - \delta_n)}{\sigma_k^n + H(\Delta b_k^n)^\lambda}\right\} \cap (G_{k-1}^n)^c\right).$$

By independence of  $L_{t+\tau_{k-1}^n} - L_{\tau_{k-1}^n}$  and  $X_{\tau_{k-1}^n}^n$ , stationarity and the strong Markov property of  $L$ , this is equal to

$$\mathbb{P}\left(L_t < \frac{n(\Delta b_k^n - \delta_n)}{\sigma_k^n + H(\Delta b_k^n)^\lambda}\right) \mathbb{P}((G_{k-1}^n)^c).$$

If we take  $x \in B_k^n$ , we have for the first probability in the display the lower bound

$$\mathbb{P}\left(L_t < \frac{n(\Delta b_k^n - \delta_n)}{\sigma(x) + 2H(\Delta b_k^n)^\lambda}\right). \tag{24}$$

This concludes the proof. □

Let  $0 < x < b_K$  and  $k = k_n(x)$  such that  $x \in B_k^n$ . We next present a result, Proposition 14, on the asymptotic behaviour of  $\Delta\tau_k^n(x)$ , which shows that, with high probability, it is concentrated near  $\Delta\bar{\tau}_k^n(x)$  as introduced above. For the result we need the condition on the bin width,  $\Delta b_k^n \asymp n^{-\kappa}$ , and require

$$\kappa \geq \frac{1}{2\lambda + 1}. \tag{25}$$

Along with this condition we let  $\delta_n \asymp n^{-\delta}$ , and require

$$\frac{1 + \kappa}{2} \leq \delta < 1. \tag{26}$$

**Proposition 14.** *Let  $x \in (0, b_K)$  and  $x \in B_k^n$ , for  $k = k_n(x)$ . Let  $\Delta\bar{\tau}_k^n(x) = \frac{n\Delta b_k^n \beta}{\alpha\sigma(x)}$ . Under conditions (25) and (26) and for  $c_n \rightarrow \infty$  such that  $c_n n^{-\frac{1}{2}(1-\kappa)} \rightarrow 0$  it holds that*

$$\mathbb{P}\left(\left(1 - \frac{c_n}{\sqrt{n\Delta b_k^n}}\right) \Delta\bar{\tau}_k^n(x) \leq \Delta\tau_k^n \leq \left(1 + \frac{c_n}{\sqrt{n\Delta b_k^n}}\right) \Delta\bar{\tau}_k^n(x)\right) \rightarrow 1.$$

**Proof.** We consider for  $u_k^n(x) := \frac{n\Delta b_k^n \beta}{\alpha \sigma(x)} \left(1 + \frac{c_n}{\sqrt{n\Delta b_k^n}}\right)$  the probability

$$\mathbb{P}(\Delta\tau_k^n \leq u_k^n(x)) = 1 - \mathbb{P}(\Delta\tau_k^n > u_k^n(x))$$

and derive a lower bound for it using the upper bound for  $\mathbb{P}(\Delta\tau_k^n < t)$  for  $t = u_k^n(x)$  as in Lemma 13. One has

$$\mathbb{P}(\Delta\tau_k^n > u_k^n(x)) \leq \mathbb{P}\left(L_{u_k^n(x)} < \frac{n\Delta b_k^n}{\sigma_k^n - H(\Delta b_k^n)^\lambda}\right).$$

Note that  $\mathbb{E}L_{u_k^n(x)} = \frac{\alpha}{\beta} u_k^n(x)$  and  $\text{Var} L_{u_k^n(x)} = \frac{\alpha}{\beta^2} u_k^n(x)$ . Hence, by the central limit theorem applied to gamma distributed random variables,  $\hat{L}_{u_k^n(x)} = \frac{L_{u_k^n(x)} - \frac{\alpha}{\beta} u_k^n(x)}{\sqrt{\frac{\alpha}{\beta^2} u_k^n(x)}}$  asymptotically has the standard normal distribution. The probability on the right of the above display is, for large enough  $n$  less than  $\mathbb{P}(L_{u_k^n(x)} < \frac{n\Delta b_k^n}{\sigma(x) - 2H(\Delta b_k^n)^\lambda})$ , which can be rewritten as

$$\mathbb{P}\left(\hat{L}_{u_k^n(x)} < \frac{\frac{n\Delta b_k^n}{\sigma(x) - 2H(\Delta b_k^n)^\lambda} - \frac{\alpha}{\beta} u_k^n(x)}{\sqrt{\frac{\alpha}{\beta^2} u_k^n(x)}}\right).$$

The right hand side of the inequality in parentheses can be rewritten as

$$\frac{1}{\sigma(x) - 2H(\Delta b_k^n)^\lambda} \sqrt{\frac{\beta\sigma(x)}{1 + \frac{c_n}{\sqrt{n\Delta b_k^n}}}} \left(-c_n + \frac{2H}{\sigma(x)} (\Delta b_k^n)^\lambda \sqrt{n\Delta b_k^n} \left(1 + \frac{c_n}{\sqrt{n\Delta b_k^n}}\right)\right).$$

This term tends to minus infinity if  $\frac{c_n}{\sqrt{n\Delta b_k^n}} \rightarrow 0$ , which is assumed, and if  $(\Delta b_k^n)^{\lambda+\frac{1}{2}} \sqrt{n}$  is bounded, the latter happens under condition (25). Under this condition it follows from the central limit theorem that  $\mathbb{P}(\Delta\tau_k^n > u_k^n(x)) \rightarrow 0$ .

Next we consider  $\mathbb{P}(\Delta\tau_k^n > l_k^n(x))$  with  $l_k^n(x) := \frac{n\Delta b_k^n \beta}{\alpha \sigma(x)} \left(1 - \frac{c_n}{\sqrt{n\Delta b_k^n}}\right)$  and show that this probability tends to one. First we use the lower bound, taken from Lemma 13,

$$\mathbb{P}(\Delta\tau_k^n > l_k^n(x)) \geq \mathbb{P}\left(L_{l_k^n(x)} < \frac{n(\Delta b_k^n - \delta_n)}{\sigma_k^n + H(\Delta b_k^n)^\lambda}\right) \mathbb{P}((G_{k-1}^n)^c).$$

Lemma 12 says that  $\mathbb{P}((G_{k-1}^n)^c) \rightarrow 1$  and

$$\mathbb{P}\left(L_{l_k^n(x)} < \frac{n(\Delta b_k^n - \delta_n)}{\sigma_k^n + H(\Delta b_k^n)^\lambda}\right) \geq \mathbb{P}\left(L_{l_k^n(x)} < \frac{n(\Delta b_k^n - \delta_n)}{\sigma(x) + 2H(\Delta b_k^n)^\lambda}\right).$$

As for the previous case, we look at the standardisation  $\hat{L}_{l_k^n(x)}$  of  $L_{l_k^n(x)}$  and consider

$$\mathbb{P}\left(L_{l_k^n(x)} < \frac{n(\Delta b_k^n - \delta_n)}{\sigma_k^n + H(\Delta b_k^n)^\lambda}\right) \geq \mathbb{P}\left(\hat{L}_{l_k^n(x)} < \frac{\frac{n(\Delta b_k^n - \delta_n)}{\sigma(x) + 2H(\Delta b_k^n)^\lambda} - \frac{\alpha}{\beta} l_k^n(x)}{\sqrt{\frac{\alpha}{\beta^2} l_k^n(x)}}\right).$$

The right hand side of the inequality in parentheses is seen to be equal to

$$\frac{1}{\sigma(x) + 2H(\Delta b_k^n)^\lambda} \sqrt{\frac{\beta\sigma(x)}{1 - \frac{c_n}{\sqrt{n\Delta b_k^n}}}} \left( c_n - \delta_n \sqrt{\frac{n}{\Delta b_k^n}} - \frac{2H}{\sigma(x)} (\Delta b_k^n)^\lambda \sqrt{n\Delta b_k^n} \left(1 - \frac{c_n}{\sqrt{n\Delta b_k^n}}\right) \right).$$

This term tends to plus infinity under Condition (25) if  $\delta_n \sqrt{\frac{n}{\Delta b_k^n}}$  is bounded, which happens under condition (26), and  $\frac{c_n}{\sqrt{n\Delta b_k^n}} \rightarrow 0$ , which is assumed. Consequently, by the central limit theorem,

$$\mathbb{P}(L_k^n(x) < \frac{n(\Delta b_k^n - \delta_n)}{\sigma_k^n + H(\Delta b_k^n)^\lambda}) \rightarrow 1. \quad \square$$

### 4.2. Posterior contraction rate

Let  $x \in (0, b_K)$  and  $k = k_n(x)$  such that  $x \in B_k^n$ . The  $\xi_k$  corresponding to  $x$  is, as it depends on  $n$ , also denoted  $\xi^n(x)$  instead of  $\xi_k$ .

As in Section 3.2, we consider the posterior mean squared error  $\mathbb{E}_{\Pi_n}(\xi_k - \sigma_k)^2$ , which we analyse through the corresponding posterior bias and variance. Then the bias of the posterior mean for  $x \in B_k^n$  is

$$\text{bias}(x) = \mathbb{E}_{\Pi_n} \xi_k - \sigma(x) = \frac{n\beta(X_{\tau_k^n} - X_{\tau_{k-1}^n}) + \beta_k}{\alpha\Delta\tau_k^n + \alpha_k - 1} - \sigma(x). \tag{27}$$

The posterior variance is

$$\text{Var}_{\Pi_n} \xi_k = \frac{(n\beta\Delta X_{\tau_k^n} + \beta_k)^2}{(\alpha\Delta\tau_k^n + \alpha_k - 1)^2(\alpha\Delta\tau_k^n + \alpha_k - 2)}.$$

As before we give upper and lower bounds for posterior bias and variance. To do so we need, along with the already introduced events  $G_k^n = \{\text{POT}_k^n > \delta_n\}$ , the events  $F_k^n = \{\Delta\tau_k^n > \Delta\bar{\tau}_k^n(x)(1 + \frac{c_n}{\sqrt{n\Delta b_k^n}})\}$  and  $H_k^n = \{\Delta\tau_k^n < \Delta\bar{\tau}_k^n(x)(1 - \frac{c_n}{\sqrt{n\Delta b_k^n}})\}$  for  $c_n \rightarrow \infty$  (arbitrarily slowly). We know from Proposition 14 that  $\mathbb{P}(F_k^n) \rightarrow 0$  and  $\mathbb{P}(H_k^n) \rightarrow 0$ , and from Lemma 12 that  $\mathbb{P}(G_k^n) \rightarrow 0$ .

**Lemma 15.** *Assume the model with piecewise constant volatility  $\xi(x)$  as given by (3.2) whereas the true volatility function  $x \mapsto \sigma(x)$  is Hölder continuous and bounded from below by a strictly positive constant  $\underline{\sigma}$ . Assume  $\Delta b_k^n \asymp n^{-K}$  and condition (25). Let  $c_n \rightarrow \infty$ . Then the posterior mean squared error*

$\mathbb{E}_{\Pi_n}(\xi^n(x) - \sigma(x))^2 = O(\frac{c_n^2}{n\Delta b_k^n})$  for all  $k = 1, \dots, K$ , uniformly in  $x \in [0, b_K]$ , with probability tending to one. That is,

$$\sup_{x \in [0, b_K]} \mathbb{E}_{\Pi_n}(\xi^n(x) - \sigma(x))^2 = O\left(\max_k \frac{c_n^2}{n\Delta b_k^n}\right)$$

with probability tending to one.

**Proof.** Let  $x \in B_k^n$ . We consider the bias (27) first. The bias can be split into its behaviour on the sets  $(F_k^n)^c \cap (G_{k-1}^n)^c$  and  $F_k^n \cup G_{k-1}^n$ . As  $\mathbb{P}(F_k^n \cup G_{k-1}^n) \rightarrow 0$ , we only have to analyze what happens on  $(F_k^n)^c \cap (G_{k-1}^n)^c$ .

We give upper and lower bounds for this bias. We start with a first lower bound. As  $\Delta X_{\tau_k^n}^n = \Delta b_k^n + \text{POT}_k^n - \text{POT}_{k-1}^n \geq \Delta b_k^n - \text{POT}_{k-1}^n$ , and recalling that  $\text{POT}_{k-1}^n < \delta_n$  on  $(G_{k-1}^n)^c$ , we obtain that on  $(F_k^n)^c \cap (G_{k-1}^n)^c$  one has

$$\text{bias}(x) \geq \frac{n\beta(\Delta b_k^n - \delta_n)\mathbf{1}_{\{\Delta\tau_k^n > 0\}} + \beta_k}{n\beta\Delta b_k^n(1 + \frac{c_n}{\sqrt{n\Delta b_k^n}})/\sigma(x) + \alpha_k - 1} - \sigma(x),$$

of which the right hand side can be rewritten as

$$\sigma(x) \frac{-\frac{\delta_n}{\Delta b_k^n} - \frac{c_n}{\sqrt{n\Delta b_k^n}} + \frac{\beta_k - (\alpha_k - 1)\sigma(x)}{\beta n\Delta b_k^n}}{1 + \frac{c_n}{\sqrt{n\Delta b_k^n}} + (\alpha_k - 1)\frac{\sigma(x)}{\beta n\Delta b_k^n}}.$$

This term is of order  $\frac{c_n}{\sqrt{n\Delta b_k^n}}$  if  $\frac{\delta_n}{\Delta b_k^n} \sqrt{n\Delta b_k^n}$  stays bounded, which happens for  $\delta_n = n^{-\delta}$  under condition (26). By continuity of  $\sigma$  on  $[0, b_K]$  this bound is uniform in  $x$ . Next we turn to an upper bound for the bias. Now we consider the bias on the events  $(H_k^n)^c \cap (G_k^n)^c$  and  $H_k^n \cup G_k^n$ . As  $\mathbb{P}(H_k^n \cup G_k^n) \rightarrow 0$ , we can ignore the bias on the latter event. Using  $\Delta X_{\tau_k^n}^n = \Delta b_k^n + \text{POT}_k^n - \text{POT}_{k-1}^n \leq \Delta b_k^n + \text{POT}_k^n$ , we have on the set  $(H_k^n)^c \cap (G_k^n)^c$

$$\text{bias}(x) \leq \frac{n\beta(\delta_n + \Delta b_k^n) + \beta_k}{n\beta\Delta b_k^n(1 - \frac{c_n}{\sqrt{n\Delta b_k^n}})/\sigma(x) + \alpha_k - 1} - \sigma(x),$$

whose right hand side becomes

$$\sigma(x) \left( \frac{1 + \frac{\delta_n}{\Delta b_k^n} + \frac{\beta_k}{n\beta\Delta b_k^n}}{1 - \frac{c_n}{\sqrt{n\Delta b_k^n}} + \frac{(\alpha_k - 1)\sigma(x)}{n\beta\Delta b_k^n}} - 1 \right),$$

and that is equal to

$$\sigma(x) \frac{\frac{\delta_n}{\Delta b_k^n} + \frac{c_n}{\sqrt{n\Delta b_k^n}} + \frac{\beta_k - (\alpha_k - 1)\sigma(x)}{n\beta\Delta b_k^n}}{1 - \frac{c_n}{\sqrt{n\Delta b_k^n}} + \frac{(\alpha_k - 1)\sigma(x)}{n\beta\Delta b_k^n}}.$$

Similar to what we have seen for the lower bound of the bias, also the upper bound is of order  $\frac{c_n}{\sqrt{n\Delta b_k^n}}$  under condition (26), and uniform in  $x$ . Summarising, under the stipulated conditions, we obtain that  $\text{bias}(x)$  is of order  $\frac{c_n}{\sqrt{n\Delta b_k^n}}$  on a set with probability tending to one.

We move on to the posterior variance of  $\xi_k$  for  $x \in B_k^n$ ,

$$\text{Var}_{\Pi_n} \xi_k = \frac{(n\beta\Delta X_{\tau_k^n}^n + \beta_k)^2}{(\alpha\Delta\tau_k^n + \alpha_k - 1)^2(\alpha\Delta\tau_k^n + \alpha_k - 2)},$$

for which we will derive upper and lower bounds as well. Paralleling the computations for the bias, we have on the event  $(F_k^n)^c \cap (G_{k-1}^n)^c$  the immediate lower bound

$$\text{Var}_{\Pi_n} \xi_k \geq \frac{(n\beta(\Delta b_k^n - \delta_n) + \beta_k)^2}{(n\beta\Delta b_k^n(1 + \frac{c_n}{\sqrt{n\Delta b_k^n}})/\sigma(x) + \alpha_k - 1)^2(n\beta\Delta b_k^n(1 + \frac{c_n}{\sqrt{n\Delta b_k^n}})/\sigma(x) + \alpha_k - 2)},$$

where the right hand side equals

$$\frac{\sigma(x)^3}{\beta n \Delta b_k^n} \frac{(1 - \frac{\delta_n}{\Delta b_k^n} + \frac{\beta_k}{\beta n \Delta b_k^n})^2}{(1 + \frac{c_n}{\sqrt{n\Delta b_k^n}} + \frac{\sigma(x)(\alpha_k - 1)}{\beta n \Delta b_k^n})^2 (1 + \frac{c_n}{\sqrt{n\Delta b_k^n}} + \frac{\sigma(x)(\alpha_k - 2)}{\beta n \Delta b_k^n})}.$$

This is obviously of order  $O(\frac{1}{n\Delta b_k^n})$ , as  $\frac{\delta_n}{\Delta b_k^n} \rightarrow 0$ . Next we give an upper bound for the posterior variance, for which we only consider what happens on  $(H_k^n)^c \cap (G_k^n)^c$ . On that event one has

$$\text{Var}_{\Pi_n} \xi_k \leq \frac{(n\beta(\Delta b_k^n + \delta_n) + \beta_k)^2}{(n\beta\Delta b_k^n(1 - \frac{c_n}{\sqrt{n\Delta b_k^n}})/\sigma(x) + \alpha_k - 1)^2(n\beta\Delta b_k^n(1 - \frac{c_n}{\sqrt{n\Delta b_k^n}})/\sigma(x) + \alpha_k - 2)}.$$

One sees that this quantity is of order  $O(\frac{1}{n\Delta b_k^n})$ , if  $\frac{\delta_n}{\Delta b_k^n}$  tends to zero, which happens under the condition (26) for  $\delta_n = n^{-\delta}$ . Combining the two results on the bounds, we conclude that the posterior variance is of order  $O(\frac{1}{n\Delta b_k^n}) = O(n^{-1+\kappa})$  with probability tending to one. As for the bias, also this order bound is uniform in  $x$ . As a last step, by the above established properties of posterior bias and variance, we obtain the posterior mean squared error  $\mathbb{E}_{\Pi_n}(\xi^n(x) - \sigma(x))^2$  is of order  $O(\frac{c_n^2}{n\Delta b_k^n}) = O(\frac{c_n^2}{n^{1-\kappa}})$  with probability tending to one. Again, this order bound is uniform in  $x$  (and  $k$ ). □

Here is the main result of this section, which says that the contraction rate of the posterior distribution is (at least)  $n^{-\lambda/(2\lambda+1)}$ ,  $\lambda$  being the Hölder exponent of  $\sigma$ .

**Theorem 16.** *Assume the model with volatility (19) whereas the true volatility function  $x \mapsto \sigma(x)$  is Hölder continuous of order  $\lambda \leq 1$  and bounded from below. Let the bin sizes shrink proportional to  $n^{-\frac{1}{2\lambda+1}}$ , and let  $(m_n)$  be any sequence of real numbers (arbitrarily slowly) diverging to infinity. Then, for  $n \rightarrow \infty$*

$$\sup_{x \in [0, b_K]} \Pi_n \left( |\xi^n(x) - \sigma(x)| > m_n n^{-\frac{\lambda}{2\lambda+1}} \right) \rightarrow 0 \text{ in probability.}$$

**Proof.** Let  $x \in B_k^n$  and  $c_n \rightarrow \infty$  such that  $\frac{c_n}{m_n} \rightarrow 0$ , for instance  $c_n = \sqrt{m_n}$ . By Chebychev’s inequality and Lemma 15, for  $\Delta b_k^n \asymp n^{-\kappa}$  with  $\kappa$  such that (25) is satisfied, we have uniformly in  $x$

$$\Pi_n \left( |\xi^n(x) - \sigma(x)| > m_n n^{-\frac{\lambda}{2\lambda+1}} \right) \leq \frac{n^{\frac{2\lambda}{2\lambda+1}}}{m_n^2} \mathbb{E}_{\Pi_n}(\xi^n(x) - \sigma(x))^2 = O\left( \frac{n^{\frac{2\lambda}{2\lambda+1}}}{m_n^2} \frac{c_n^2}{n\Delta b_k^n} \right),$$

with probability tending to one. The choice  $\kappa = \frac{1}{2\lambda+1}$  satisfies (25) and  $\Delta b_k^n \asymp n^{-\frac{1}{2\lambda+1}}$  is assumed. Hence, we see that the order bound becomes  $O(\frac{c_n^2}{m_n^2})$ , which tends to zero. □

### 4.3. Lower bounds

In this section we show that the contraction rates in Theorem 16 are minimax optimal. Let  $\Sigma(\lambda, L)$  denote the Hölder class of functions  $f$  on  $[0, 1]$  satisfying

$$|f(x) - f(y)| \leq L|x - y|^\lambda, \quad x, y \in [0, 1].$$

We endow  $\Sigma(\lambda, L)$  with the supnorm, denoted  $\|\cdot\|_\infty$ . Denote by  $\mathbb{P}_T^{\sigma, n}$  the law of the process  $(X_t^n)_{t \in [0, T]}$  solving the Levy-driven SDE, similar to (5),

$$dX_t^n = \sigma(X_{t-}^n) dL_t^n, \quad X_0^n = 0,$$

where  $L^n$  is a gamma process with a Lévy density

$$v_n(x) = \frac{n\alpha}{x} \exp(-n\beta x).$$

We shall prove the following statement.

**Proposition 17.** *There are constants  $c_0, c_1 > 0$  not depending on  $n$  such that*

$$\liminf_{n \rightarrow \infty} \inf_{\hat{\sigma}_n} \sup_{\sigma \in \Sigma(\lambda, L)} \mathbb{P}_T^{\sigma, n} \left( \|\sigma - \hat{\sigma}_n\|_\infty \geq c_0 n^{-\lambda/(1+2\lambda)} \right) \geq c_1, \tag{28}$$

where the infimum is taken over all estimators  $\hat{\sigma}_n$ , that is, all measurable functions of the path  $X_t^n$ ,  $t \in [0, T]$ .

It is well known that the posterior cannot converge at a rate faster than the optimal rate of convergence for point estimators, see Ghosal, Ghosh and Van Der Vaart (2000), page 507. Since we have a lower bound and it gives the rate matching our posterior contraction rate, our Bayesian approach is optimal from the frequentist point of view.

**Proof.** Our strategy is to follow the approach as in Chapter 2 of Tsybakov (2008). Therefore, our goal is to establish

$$\inf_{\hat{\sigma}_n} \sup_{\sigma \in \Sigma(\lambda, L)} \mathbb{P}_T^{\sigma, n} (\|\sigma - \hat{\sigma}_n\|_\infty \geq s_n) \geq \frac{1}{2} (1 - V(\mathbb{P}_T^{\sigma_0, n}, \mathbb{P}_T^{\sigma_1, n})), \tag{29}$$

where  $\sigma_0, \sigma_1$  are two distinct elements of  $\Sigma(\lambda, L)$  such that

$$\|\sigma_0 - \sigma_1\|_\infty \geq 2s_n, \tag{30}$$

and  $V(\mathbb{P}_T^{\sigma_0, n}, \mathbb{P}_T^{\sigma_1, n})$  is the total variation distance between  $\mathbb{P}_T^{\sigma_0, n}$  and  $\mathbb{P}_T^{\sigma_1, n}$ . The  $s_n$  denotes the desired convergence rate, in our case we aim at  $s_n \asymp n^{-\lambda/(2\lambda+1)}$ . The result in (29) results from the exposition in Section 2.2 and Theorem 2.2(i) together with its proof in Tsybakov (2008).

We first select  $\sigma_0$  and  $\sigma_1$ . Fix  $h \in (0, 1]$  and set

$$\sigma_0(x) \equiv 1, \quad \sigma_1(x) = 1 - h^\lambda \psi(x/h),$$

where  $\psi \in \Sigma(\lambda, 1/2)$  is a nonnegative monotone decreasing function supported on  $[0, 1]$  satisfying  $\|\psi\|_\infty \leq 1, \psi(x) > 0$  for  $x \in [0, 1/2]$ . One can take, for example,

$$\psi(u) = a \exp\left(-\frac{1}{1-u^2}\right) \mathbf{1}_{[0,1]}(u)$$

for  $a > 0$  small enough. In fact, taking  $a = 1/e$  for this choice of  $\psi$  we have  $\|\psi\|_\infty = \psi(0) = 1$ . Note furthermore  $\psi \in \Sigma(1, 1/2)$ , implying  $\psi \in \Sigma(\lambda, 1/2)$  for any  $\lambda \leq 1$  and  $\sigma_1 \in \Sigma(\lambda, 1/2)$ . Moreover, now  $\|\sigma_0 - \sigma_1\|_\infty = h^\lambda$ . Later we will choose  $h$  and  $s_n$  such that  $h^\lambda \geq 2s_n$  in order that (30) holds.

We will next show that  $V(\mathbb{P}_T^{\sigma_0, n}, \mathbb{P}_T^{\sigma_1, n})$  is eventually less than some constant  $\nu < 1$ , after which we can choose  $c_1 = \frac{1}{2}(1 - \nu)$  in (29) to obtain (28). The total variation distance  $V$  between the laws  $\mathbb{P}_T^{\sigma_0, n}$  and  $\mathbb{P}_T^{\sigma_1, n}$  satisfies (see [Kabanov, Liptser and Shiryaev \(1986\)](#))

$$V^2(\mathbb{P}_T^{\sigma_0, n}, \mathbb{P}_T^{\sigma_1, n}) \leq 16 \mathbb{E}_{\mathbb{P}_T^{\sigma_0, n}} \left[ \int_0^T \int_0^\infty (\sqrt{Y(t, x)} - 1)^2 v_n(x) dx dt \right],$$

$$Y(t, x) = \frac{1}{\sigma_1(X_{t-})} v_n\left(\frac{x}{\sigma_1(X_{t-})}\right) / v_n(x).$$

Using the inequality  $1 - e^{-x} \leq x$  holding for all  $x > 0$ , we get

$$\begin{aligned} V^2(\mathbb{P}_T^{\sigma_0, n}, \mathbb{P}_T^{\sigma_1, n}) &\leq 16 \mathbb{E}_{\mathbb{P}_T^{\sigma_0, n}} \left[ \int_0^T \int_0^\infty \left( \frac{xn\beta(1 - \sigma_1^{-1}(X_{t-}))}{2} \right)^2 v_n(x) dx dt \right] \\ &\lesssim \mathbb{E}_{\mathbb{P}_T^{\sigma_0, n}} \left[ \int_0^T \int_0^\infty n^2 h^{2\lambda} \psi^2(X_{t-}/h) x^2 v_n(x) dx dt \right] \\ &\lesssim n^2 h^{2\lambda} \mathbb{E}_{\mathbb{P}_T^{\sigma_0, n}} \left[ \int_0^T \psi^2(X_{t-}/h) dt \right] \int_0^\infty x^2 v_n(x) dx \\ &\lesssim nh^{2\lambda} \mathbb{E}_{\mathbb{P}_T^{\sigma_0, n}} \left[ \int_0^T \psi^2(X_{t-}/h) dt \right], \end{aligned} \tag{31}$$

where here and in the sequel  $\lesssim$  for inequality up to a constant depending on  $\alpha$  and  $\beta$ . Furthermore

$$\mathbb{E}_{\mathbb{P}_T^{\sigma_0, n}} \left[ \int_0^T \psi^2(X_{t-}/h) dt \right] = \int_0^T \int_0^\infty \left[ \psi^2(z/h) p_{L_t^n}(z) \right] dz dt,$$

where  $p_{L_t^n}$  is the density of  $L_t^n$ . Using well-known results for the Gamma function, see e.g. the sharp version of the Stirling formula of [Theorem 1.6 in Batir \(2008\)](#), we have for  $t \geq 2/(n\alpha)$

$$p_{L_t^n}(z) = \frac{(n\beta)^{n\alpha t}}{\Gamma(n\alpha t)} z^{n\alpha t - 1} e^{-n\beta z} \lesssim \frac{1}{z} \sqrt{\frac{n\alpha t}{2\pi}} \exp\left\{n\alpha t \left(\log\left(\frac{\beta z}{\alpha t}\right) + 1 - \frac{\beta z}{\alpha t}\right)\right\}.$$

Fix some  $0 < \delta < 1/2$  and consider the integral

$$\begin{aligned} \int_0^\infty \left[ \psi^2(z/h) p_{L_t^n}(z) \right] dz &\lesssim \int_0^\infty \frac{1}{y} \sqrt{\frac{n\alpha t}{2\pi}} \psi^2(\alpha t y / \beta h) e^{n\alpha t(1-y+\log(y))} dy \\ &= \int_{|y-1| \leq \delta} \frac{1}{y} \sqrt{\frac{n\alpha t}{2\pi}} \psi^2(\alpha t y / \beta h) e^{n\alpha t(1-y+\log(y))} dy \end{aligned}$$

$$\begin{aligned}
 &+ \int_{y>1+\delta} \frac{1}{y} \sqrt{\frac{n\alpha t}{2\pi}} \psi^2(\alpha t y / \beta h) e^{n\alpha t(1-y+\log(y))} dy \\
 &+ \int_{y<1-\delta} \frac{1}{y} \sqrt{\frac{n\alpha t}{2\pi}} \psi^2(\alpha t y / \beta h) e^{n\alpha t(1-y+\log(y))} dy \\
 &=: I_1 + I_2 + I_3.
 \end{aligned}$$

Since  $1 - y + \log(y) \leq -c(1 - y)^2$  for some  $c = c(\delta) > 0$  if  $|1 - y| \leq \delta < 1/2$ , we have with  $z = \sqrt{n\alpha t}(y - 1)$ ,

$$\begin{aligned}
 I_1 &\lesssim \int_{-\delta}^{\delta} \psi^2(\alpha t(1 + z/\sqrt{n\alpha t})/\beta h) e^{-cz^2} dz, \\
 I_2 &\lesssim \sqrt{n\alpha t} e^{n\alpha t(-\delta+\log(1+\delta))}, \\
 I_3 &\lesssim \sqrt{n\alpha t} e^{n\alpha t(\delta+\log(1-\delta))}.
 \end{aligned}$$

Hence for any  $t_0 \geq 2/(n\alpha)$  we derive, using the above estimates of  $I_1, I_2, I_3$ ,

$$\begin{aligned}
 \int_0^T \int_0^\infty \left[ \psi^2(z/h) p_{L_t^n}(z) \right] dz dt &= \int_0^{t_0} \int_0^\infty \left[ \psi^2(z/h) p_{L_t^n}(z) \right] dz dt \\
 &+ \int_{t_0}^T \int_0^\infty \left[ \psi^2(z/h) p_{L_t^n}(z) \right] dz dt \\
 &\lesssim t_0 + h \int_{-\delta}^{\delta} \int_0^{T/h} \psi^2\left(\frac{\alpha s}{\beta} + \frac{z}{\beta} \sqrt{\frac{\alpha s}{nh}}\right) ds dz \\
 &+ \sqrt{n\alpha t_0} \left( e^{n\alpha t_0(-\delta+\log(1+\delta))} + e^{n\alpha t_0(\delta+\log(1-\delta))} \right), \tag{32}
 \end{aligned}$$

where for the last term it is used that the function  $x \mapsto \sqrt{x}e^{-xa}$  with  $a > 0$  is monotone decreasing for  $x \geq 1/(2a)$ . The double integral in (32) is bounded in  $n, h$  as follows from

$$\int_0^\infty \psi^2\left(\frac{\alpha s}{\beta} + \frac{z}{\beta} \sqrt{\frac{\alpha s}{nh}}\right) ds \leq 1/\alpha + \int_{1/\alpha}^\infty \psi^2\left(\frac{\alpha s}{\beta}(1 - \delta)\right) ds,$$

where we used that  $\phi$  is bounded by 1, increasing and that in this integral  $|z| \leq \delta$ . Choosing  $t_0 = o(h)$  and such that  $nt_0 \rightarrow \infty$ , one sees that the term with  $h$  in (32) is the dominating term. Recalling (31), we subsequently take  $h = cn^{-1/(1+2\lambda)}$  for a small enough constant  $c > 0$ , to derive for some  $0 < \nu < 1$

$$V^2(\mathbb{P}_T^{\sigma_0, n}, \mathbb{P}_T^{\sigma_1, n}) \leq \nu^2$$

for all  $n$  large enough. With this choice of  $h$  we take  $s_n = \frac{1}{2}c^\lambda n^{-\lambda/(1+2\lambda)}$ . Then (30) is satisfied, and from (29) we arrive at (28) with  $c_0 = \frac{1}{2}c^\lambda$  and  $c_1 = \frac{1}{2}(1 - \nu)$ , both strictly positive.  $\square$

### 5. Real data example

The North Greenland Ice Core Project (NGRIP) obtained from drilling through arctic ice an oxygen isotope record reaching 120 000 years into the past beyond the last glacial (North Greenland Ice Core



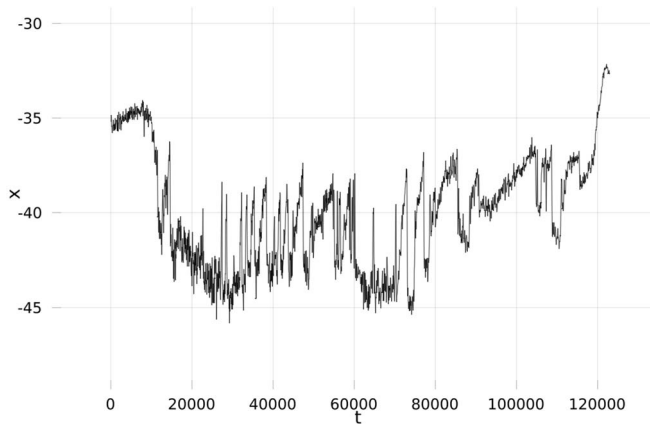


Figure 3. NGRIP oxygen isotope  $\delta_{18}\text{O}$  measurements against time (years BP 2000).

Project Members (2007).) Figure 3 shows measurements of the indicator  $\delta_{18}\text{O}$  derived from oxygen isotope measurements at times  $t = 0, 50, \dots$  in  $\Delta t = 50$  year intervals, with in total  $n = 2459$  observations. In geological scales such a  $\Delta t$  can be considered small. The oxygen isotope record is a proxy for past temperature, and the data shows characteristic sudden changes in global temperature, a topic of urgent relevance. It has been suggested to model the NGRIP data as a realisation of a stochastic differential equation with solution  $Y$ , also as a Lévy-driven SDE to account for the heavier tailed noise, see Ditlevsen (1999). An in-depth study is given in Carson et al. (2019).

Here we are interested in estimating the volatility  $\zeta$  of the process  $Y_t$ , observed on the equidistant time grid  $0, \Delta t, \dots, n\Delta t$  with  $n$  observations. Estimating the volatility is an important step in data assimilation and inference tasks related to rapid temperature transitions, for example during so called Dansgaard-Oeschger events.

For a particular realisation  $y$  of  $Y$  and a fixed  $\Delta t$ , the realised quadratic variation process over a time grid with step size  $\Delta t$  can be defined as  $q_{t+\Delta t} - q_t = \Delta q_t$  with  $\Delta q_t := (\Delta Y_t)^2 = (y_{t+\Delta t} - y_t)^2$ ,  $q_0 = 0$ . See Figure 4 for a visualisation and note the resemblance (up to scaling) of this figure with Figure 1 in Example 1. We remark that the realised quadratic variation process of a diffusion process can be considered as a measure of intrinsic progress of time, also referred to as internal clock.

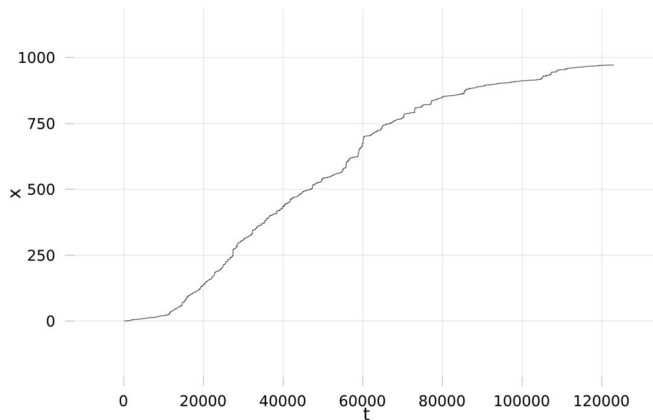
For motivation of the model proposed below, suppose that  $Y$  is a diffusion process satisfying  $dY_t = \zeta_t dW_t$ , with  $W$  a Brownian motion. For small  $\Delta t$  one has  $(\Delta Y_t)^2 \approx \zeta_t^2 (W_{t+\Delta t} - W_t)^2$ , which has (given the past up to time  $t$ ) a  $\text{Gamma}(\frac{1}{2}, \frac{1}{2\Delta t \zeta_t^2})$  distribution.

Consider next the Lévy-driven SDE

$$dX_t = c\zeta_t^2 dL_t, \tag{33}$$

where  $L$  is a gamma process with parameters  $\alpha, \beta$  to be specified shortly. For  $X$  solving (33) we have that  $\Delta X_t \approx c\zeta_t^2 \Delta L_t$ , which is, conditional on the past up to time  $t$ , a  $\text{Gamma}(\alpha\Delta t, \beta/(c\zeta_t^2))$  random variable. Then with  $\alpha = \frac{1}{2\Delta t}$  and  $\beta = \frac{c}{2\Delta t}$ , the conditional distributions of  $\Delta X_t$  and  $(\Delta Y_t)^2$  are approximately gamma with the same parameters, for any choice of  $c > 0$ . We used  $c = n\Delta t/q_{n\Delta t}$ , which implies  $\mathbb{E}L_{n\Delta t} = q_{n\Delta t}$ . With this in mind, we model  $q_{i\Delta t}$  as observation of a realisation of the continuous time process

$$dX_t = \sigma(X_{t-})dL_t. \tag{34}$$



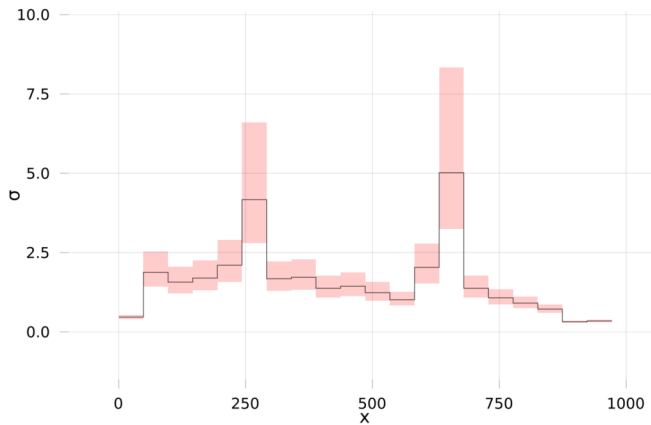
**Figure 4.** Realized quadratic variation process for NGRIP oxygen isotope  $\delta_{18}\text{O}$  measurements against time (years BP 2000).

with unknown  $\sigma$ . As both  $\alpha$  and  $\beta$  are proportional to  $1/\Delta t \asymp n$ , this corresponds to the regime detailed after (5) with parameters of the driving gamma process proportional to  $n$ , which suggests that our asymptotic results are practically relevant for this problem. Here, we choose to model the slope  $\sigma$  of the curve of realized quadratic variation, see in Figure 4, as a function of  $Y_t = \int_0^t cS_s^2 dL_s$ , using the monotonicity of the realisation.

In estimating  $\sigma$  with (34), using terminology common in the financial literature, we estimate volatility of  $\delta_{18}\text{O}$  as function of the time measured by the intrinsic (or business) clock. To perform the statistical analysis with our approach we set  $K = 20$ , which appears to be a good compromise in terms of bias-variance trade-off. We take equidistant bins over the range of observations of  $q_t$ . Note that a fixed resolution in space implies a variable resolution in time, with larger bins at times where the increments of  $X$  are small. Then the posterior is determined by choosing the prior parameters; here we took weakly informative parameters  $\alpha_k = \beta_k \equiv 0.1$ . Figure 5 gives the marginal posterior band for  $\sigma$ . The figure shows that in general with higher intrinsic age  $X_t$ , the volatility of the measurements decreases. This phenomenon has been noted before, and a relation to aging processes in the ice has been suggested, highlighting two periods of unusual activity.

## 6. Brief discussion of some extensions

In the treatment of our statistical problem the standing assumption so far was that observations were rooted in a model derived from (1), and that the process has been observed continuously in time. In an asymptotic setting we have derived contraction rates for the Bayesian estimator in Sections 3 and 4. There are (at least) two variations that immediately come to mind, an extended model that includes a drift component and a different observation regime, i.e. discrete time observations. We address these in the next two short sections and discuss how to adapt our previous statistical procedure to these situations.



**Figure 5.** Shaded: marginal 90%-posterior credible band for the piecewise constant posterior for the scale of the quadratic variation process of the ice core time series. Black: marginal posterior median.

### 6.1. Inclusion of drift

In principle, one could also consider a modification of the original SDE model by adding a drift,

$$X_t = \int_0^t a(X_s) ds + \int_0^t \sigma(X_{s-}) dL_s. \tag{35}$$

The property that  $X$  satisfying (35) has increasing paths is guaranteed by imposing that  $a$  is nonnegative. Inclusion of the drift has the consequence that the laws of  $X$  on  $[0, T]$  under presence and absence of a drift, but keeping the  $\int_0^t \sigma(X_{s-}) dL_s$  part the same for both cases, are automatically mutually singular. This happens because an absolutely continuous change of measure in this setting only changes the third characteristic of a semimartingale, which excludes a change of drift. This implies that the drift part,  $\int_0^t a(X_s) ds$ , can be identified with probability one.

Indeed, if  $X$  is observed in continuous time, one also observes the process  $\sum_{s \leq \cdot} \Delta X_s$ , and therefore also the difference  $X - \sum_{s \leq \cdot} \Delta X_s$ , which is  $\int_0^t a(X_s) ds$ . For the statistical problem of estimating  $\sigma$ , one can apply a modification of our procedure, by incorporating the values  $a(X_s)$ . For instance, if next to  $\sigma$  also  $a$  is piecewise constant on bins, with known (or estimated) values  $a_k$ , the assertion of Proposition 8 has to be modified by taking  $\Delta \bar{\tau}_k^n$  as  $\frac{n \Delta b_k \beta}{\alpha \sigma_k + \beta a_k}$ . Hence we derive analogously to Theorem 11 that for any sequence  $m_n \rightarrow \infty$ ,

$$\Pi_n \left( |\xi_k - \sigma_k| > \frac{m_n}{\sqrt{n}} \right) \rightarrow 0 \text{ in probability,}$$

for  $k = 1, \dots, K$ , where we model the  $(\sigma_k)$  as independent random variables  $(\xi_k)$  and  $\Pi_n$  stands for the posterior distribution of  $(\xi_k)$ .

### 6.2. Discrete time observations

In previous sections we have assumed that  $X^n$  given by (4) is observed continuously on a long time interval  $[0, T^n]$ . In this section the index  $n$  plays no role (in the situation we will describe shortly, we even take  $n = 1$  without loss of generality), it is fixed and therefore omitted below from our notation.

The current setting is the practically relevant situation where  $X$  is observed only discretely in time at timepoints  $t_0, \dots, t_m$ . We use the abbreviated notation  $X^d = (X_{t_i})_{i=1}^m$ . It is assumed that realisations  $X_{t_i} = x_{t_i}$ ,  $i = 0, \dots, m$  have been observed. In this discrete time setting, the level hitting times  $\tau_k$ , are only known approximately: if  $b_k \in [X_{t_i}, X_{t_{i+1}})$ , then  $\tau_k \in [t_i, t_{i+1})$ . Our previous inference procedure for piecewise constant  $\sigma$ , assuming observation of the  $\tau_k$ , can then still be used in a data augmentation type Gibbs sampler, targeting the joint posterior distribution of  $X$  and  $\sigma$  given the observations. The sampler alternates an interpolation step, where the random times  $\tau_k$  are sampled conditional on  $\sigma$  and the observations  $X_{t_i}$ , and an inference step where  $\sigma$  is sampled conditional on the  $\tau_k$  as described before to obtain the posterior distribution. In this section we propose a procedure how to sample the  $\tau_k$  conditional on  $\sigma$  and observations  $X_{t_i}$  to complement the continuous time inference procedure developed before, e.g. as in Section 3. For that extension, it is enough to assume that  $\sigma$  is a piecewise constant function, fixed and known. So we only consider the problem of sampling the posterior conditional on the observation of the process  $X_{t_i}^n$  for piecewise constant  $\sigma$  given by (9). By the Markov property, the conditional distribution of  $X$  given observations  $X^d$  factorises into independent distributions of the path of  $X$  on time intervals  $[t_i, t_{i+1})$ . As the conditional distribution of  $\sigma$  only depends on the  $\tau_k$ , it is enough to sample the trajectory of  $X$  on intervals  $[t_i, t_{i+1})$  that are such that  $b_k \in [X_{t_i}, X_{t_{i+1}})$  for some  $k$ . We therefore concentrate on the task of sampling  $X$  (for this purpose we can take  $n = 1$  without loss of generality) conditional on  $X_S = u$ ,  $X_T = v$ ,  $S < T$  with  $b = b_k \in [u, v)$ .

Simplifying, we assume that each bin contains at least one observation. Observe that if  $\sigma = 1$  then  $X_t$ , conditional on  $X_{t_i} = x_{t_i}$ ,  $X_{t_{i+1}} = x_{t_{i+1}}$  for  $t_i < t < t_{i+1}$ , is just a gamma process bridge with known distribution. By Corollary 5 the law of the conditional process is absolutely continuous with respect to the law of a conditional gamma process,

$$\frac{d\mathbb{P}_T^\sigma}{d\mathbb{P}_T^1} = \exp \left\{ \beta \sum_{k=1}^K \left[ 1 - n\xi_k^{-1} \right] (X_{\tau_k} - X_{\tau_{k-1}}) - \alpha \sum_{k=1}^K (\tau_k - \tau_{k-1}) \log(\xi_k/n) \right\}.$$

We denote the law of  $\tau = (\tau_0, \dots, \tau_K)$  conditional on  $X^d = x^d$  by  $Q_\tau^\sigma$ , emphasising the dependence on  $\sigma$ . By the abstract Bayes formula, it holds that, with the approximation below obtained by replacing  $X_{\tau_k}$  with  $b_k$ ,

$$\frac{dQ_\tau^\sigma}{dQ_\tau^1}(\tau_0, \dots, \tau_K) \approx \prod_{i=1}^m \frac{p^\sigma(t_{i-1}, x_{t_{i-1}}; t_i, x_{t_i})}{p^1(t_{i-1}, x_{t_{i-1}}; t_i, x_{t_i})} R(\tau_0, \dots, \tau_K)$$

where  $p^\sigma(t_{i-1}, x_{t_{i-1}}; t_i, x_{t_i})$  denotes the transition density of the process  $X$  with volatility  $\sigma$  given  $X_{t_{i-1}} = x_{t_{i-1}}$  to the value  $X_{t_i} = x_{t_i}$ , and

$$R(\tau_0, \dots, \tau_K) = \exp \left\{ \beta \sum_{k=1}^K \left[ 1 - n\xi_k^{-1} \right] (b_k - b_{k-1}) - \alpha \sum_{k=1}^K (\tau_k - \tau_{k-1}) \log(\xi_k/n) \right\}.$$

A random sample of  $\tau$  under  $Q_\tau^1$  can be obtained, observing that the unnormalised density of  $\tau_k$  given  $X_{t_{i-1}} = x_{t_{i-1}}$  and  $X_{t_i} = x_{t_i}$ ,  $b_k \in [x_{t_i}, x_{t_{i+1}})$  is approximately given by ( $\tau_k$  is now also used for a realisation of  $\tau_k$ )

$$f(\tau_k) = p^1(t_{i-1}, x_{t_{i-1}}; \tau_k, b_k) p^1(\tau_k, b_k; t_i, x_{t_i})$$

with the approximation obtained by again replacing  $X_{\tau_k}$  by  $b_k$ . So the posterior distribution of  $\xi_k$  given  $X^d = x^d$  can be computed using a Gibbs type Metropolis-Hastings approach where in one step every  $\tau_k$  is sampled conditional on  $\xi_k$  given  $X^d = x^d$  using  $Q_\tau^1$  as proposal distribution with  $R$  as (approximate) likelihood, and then in a next step,  $\xi_k$  are sampled given  $\tau_k$  using (10).

### 6.3. More general driving processes

In this paper, we focus on SDEs with driving gamma processes. In principle, one can consider other driving Lévy processes, but then we will lose some nice features of the current model. The situation here is analogous to the diffusion case with Brownian noise. The Brownian noise enables one to write down a likelihood ratio ‘à la Girsanov. If one replaces the Brownian motion with an arbitrary continuous local martingale, the likelihood ratio will no longer be explicit. If one changes the gamma process to another increasing process in our setting, a similar phenomenon will arise. The exact expression of the likelihood ratio in Corollary 9 will change, and one has to use another prior instead of the inverse gamma one. Conceptually one can follow a similar strategy as we have proposed now, but a concrete realization will be different, and it is doubtful whether the explicit expressions for the likelihood and the posteriori distribution exist in this case.

## Acknowledgments

The research leading to these results has received funding from the European Research Council under ERC Grant Agreement 320637. The research of the first author was supported by the HSE University Basic Research Program and the German Science Foundation research grant (DFG Sachbeihilfe) 406700014.

## References

- Batir, N. (2008). Inequalities for the gamma function. *Arch. Math. (Basel)* **91** 554–563.
- Batz, P., Ruttor, A. and Opper, M. (2018). Approximate Bayes learning of stochastic differential equations. *Phys. Rev. E* **98** 022109.
- Belomestny, D., Gugushvili, S., Schauer, M. and Spreij, P. (2019). Nonparametric Bayesian inference for Gamma-type Lévy subordinators. *Commun. Math. Sci.* **17** 781–816.
- Belomestny, D., Gugushvili, S., Schauer, M. and Spreij, P. (2021). Weak solutions to gamma-driven stochastic differential equations. Available at [arXiv:2108.11891](https://arxiv.org/abs/2108.11891).
- Carson, J., Crucifix, M., Preston, S.P. and Wilkinson, R.D. (2019). Quantifying age and model uncertainties in palaeoclimate data and dynamical climate models with a joint inferential analysis. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **475** 20180854.
- Chance, D.M., Hillebrand, E. and Hilliard, J.E. (2008). Pricing an option on revenue from an innovation: An application to movie box office revenue. *Manage. Sci.* **54** 1015–1028.
- Ditlevsen, P.D. (1999). Observation of  $\alpha$ -stable noise induced millennial climate changes from an ice-core record. *Geophys. Res. Lett.* **26** 1441–1444.
- Dufresne, F., Gerber, H.U. and Shiu, E.S.W. (1991). Risk theory with the gamma process. *Astin Bull.* **21** 177–192.
- Eguchi, S. and Uehara, Y. (2020). Schwartz type model selection for ergodic stochastic differential equation models. Available at [arXiv:1904.12398](https://arxiv.org/abs/1904.12398).
- Ghosal, S., Ghosh, J.K. and Van Der Vaart, A.W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531.
- Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference. Cambridge Series in Statistical and Probabilistic Mathematics* **44**. Cambridge: Cambridge Univ. Press.
- Gugushvili, S., Mariucci, E. and van der Meulen, F. (2020). Decomposing discrete distributions: A non-parametric Bayesian approach. *Scand. J. Stat.* **47** 464–492.
- Gugushvili, S., van der Meulen, F. and Spreij, P. (2015). Nonparametric Bayesian inference for multidimensional compound Poisson processes. *Mod. Stoch. Theory Appl.* **2** 1–15.
- Gugushvili, S., van der Meulen, F. and Spreij, P. (2018). A non-parametric Bayesian approach to decomposing from high frequency data. *Stat. Inference Stoch. Process.* **21** 53–79.

- Gugushvili, S., van der Meulen, F., Schauer, M. and Spreij, P. (2019). Nonparametric Bayesian volatility estimation. In *2017 MATRIX Annals* (J. de Gier, C.E. Praeger and T. Tao, eds.) 279–302. Cham: Springer.
- Gushchin, A., Pavlyukevich, I. and Ritsch, M. (2019). Drift estimation for a Lévy-driven Ornstein-Uhlenbeck process with heavy tails. Available at [arXiv:1911.11202](https://arxiv.org/abs/1911.11202).
- Jasra, A., Kamatani, K. and Masuda, H. (2019). Bayesian inference for stable Lévy-driven stochastic differential equations with high-frequency data. *Scand. J. Stat.* **46** 545–574.
- Jasra, A., Stephens, D.A., Doucet, A. and Tsagaris, T. (2011). Inference for Lévy-driven stochastic volatility models via adaptive sequential Monte Carlo. *Scand. J. Stat.* **38** 1–22.
- Kabanov, Y.M., Liptser, R.S. and Shiryaev, A. (1986). On the variation distance for probability measures defined on a filtered space. *Probab. Theory Related Fields* **71** 19–35.
- Koskela, J., Spanò, D. and Jenkins, P.A. (2019). Consistency of Bayesian nonparametric inference for discretely observed jump diffusions. *Bernoulli* **25** 2183–2205.
- Kyprianou, A.E. (2014). *Fluctuations of Lévy Processes with Applications. Introductory Lectures*, 2nd ed. *Universitext*. Heidelberg: Springer.
- North Greenland Ice Core Project Members (2007). 50 year means of oxygen isotope data from ice core NGRIP. Supplement to: North Greenland Ice Core Project Members (2004): High-resolution record of Northern Hemisphere climate extending into the last interglacial period. *Nature* **431** 147–151. <https://doi.org/10.1038/nature02805>
- Müller, P. and Mitra, R. (2013). Bayesian nonparametric inference—why and how. *Bayesian Anal.* **8** 269–302.
- Müller, P., Quintana, F.A., Jara, A. and Hanson, T. (2015). *Bayesian Nonparametric Data Analysis. Springer Series in Statistics*. Cham: Springer.
- Nickl, R. and Söhl, J. (2017). Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions. *Ann. Statist.* **45** 1664–1693.
- Protter, P.E. (2004). *Stochastic Integration and Differential Equations*, 2nd ed. *Applications of Mathematics (New York)* **21**. Berlin: Springer. Stochastic Modelling and Applied Probability.
- Reed, W.J. and McKelvey, K.S. (2002). Power-law behaviour and parametric models for the size-distribution of forest fires. *Ecol. Model.* **150** 239–254.
- Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis. Monographs on Statistics and Applied Probability*. London: CRC Press.
- Todorov, V. (2011). Econometric analysis of jump-driven stochastic volatility models. *J. Econometrics* **160** 12–21.
- Tsybakov, A.B. (2008). *Introduction to Nonparametric Estimation*. Berlin: Springer.
- Uehara, Y. (2019b). Statistical inference for misspecified ergodic Lévy driven stochastic differential equation models. *Stochastic Process. Appl.* **129** 4051–4081. [MR3997671 https://doi.org/10.1016/j.spa.2018.11.007](https://doi.org/10.1016/j.spa.2018.11.007)
- van Noordwijk, J. (2009). A survey of the application of gamma processes in maintenance. *Reliab. Eng. Syst. Saf.* **94** 2–21.
- Wenocur, M.L. (1989b). A reliability model based on the gamma process and its analytic theory. *Adv. in Appl. Probab.* **21** 899–918.

*Received December 2020 and revised August 2021*