Nonparametric Bayesian volatility estimation

Shota Gugushvili, Frank van der Meulen, Moritz Schauer, and Peter Spreij

Abstract Given discrete time observations over a fixed time interval, we study a nonparametric Bayesian approach to estimation of the volatility coefficient of a stochastic differential equation. We postulate a histogram-type prior on the volatility with piecewise constant realisations on bins forming a partition of the time interval. The values on the bins are assigned an inverse Gamma Markov chain (IGMC) prior. Posterior inference is straightforward to implement via Gibbs sampling, as the full conditional distributions are available explicitly and turn out to be inverse Gamma. We also discuss in detail the hyperparameter selection for our method. Our nonparametric Bayesian approach leads to good practical results in a representative simulation examples. Finally, we apply it on a classical data set in change-point analysis: weekly closings of the Dow-Jones industrial averages.

Shota Gugushvili

Mathematical Institute, Leiden University, P.O. Box 9512, 2300 RA Leiden, The Netherlands. e-mail: `shota.gugushvili@math.leidenuniv.nl`

Frank van der Meulen

Delft Institute of Applied Mathematics, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands. e-mail: `f.h.vandermeulen@tudelft.nl`

Moritz Schauer

Mathematical Institute, Leiden University, P.O. Box 9512, 2300 RA Leiden, The Netherlands. e-mail: `m.r.schauer@math.leidenuniv.nl`

Peter Spreij

Korteweg-de Vries Institute for Mathematics, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands and Institute for Mathematics, Astrophysics and Particle Physics, Radboud University, Nijmegen, The Netherlands. e-mail: `spreij@uva.nl`

1 Introduction

1.1 Problem formulation

Consider a one-dimensional stochastic differential equation (SDE)

$$dX_t = b_0(t, X_t) \, dt + s_0(t) \, dW_t, \quad X_0 = x, \quad t \in [0, T], \tag{1}$$

where b_0 is the drift coefficient, s_0 the deterministic dispersion coefficient or volatility, and x is a deterministic initial condition. Here W is a standard Brownian motion. Assume that standard conditions for existence and uniqueness of a strong solution to (1) are satisfied (see, e.g., [46]), and observations

$$\mathscr{X}_n = \{X_{t_{0,n}}, \dots, X_{t_{n,n}}\}$$

are available, where $t_{i,n} = iT/n$, $i = 0, \dots, n$. Using a nonparametric Bayesian approach, our aim is to estimate the volatility function s_0. In a financial context, knowledge of the volatility is of fundamental importance e.g. in pricing financial derivatives; see [4] and [54]. However, SDEs have applications far beyond the financial context as well, e.g. in physics, biology, life sciences, neuroscience and engineering (see [1], [25], [39] and [74]). Note that by Itô's formula, using a simple transformation of the state variable, also an SDE of the form

$$dX_t = b_0(t, X_t) \, dt + s_0(t) f_0(X_t) \, dW_t, \quad X_0 = x, \quad t \in [0, T],$$

can be reduced to the form (1), provided the function f_0 is known and regular enough; see, e.g., p. 186 in [68]. Some classical examples that fall under our statistical framework are the geometric Brownian motion and the Ornstein-Uhlenbeck process. Note also that as we allow the drift in (1) to be non-linear, marginal distributions of X are not necessarily Gaussian and may thus exhibit heavy tails, which is attractive in financial modelling.

A nonparametric approach guards one against model misspecification and is an excellent tool for a preliminary, exploratory data analysis, see, e.g., [67]. Commonly acknowledged advantages of a Bayesian approach include automatic uncertainty quantification in parameter estimates via Bayesian credible sets, and the fact that it is a fundamentally likelihood-based method. In [53] it has been argued that a nonparametric Bayesian approach is important for honest representation of uncertainties in inferential conclusions. Furthermore, use of a prior allows one to easily incorporate the available external, a priori information into the estimation procedure, which is not straightforward to achieve with frequentist approaches. For instance, this a priori information could be an increasing or decreasing trend in the volatility.

1.2 Literature overview

Literature on nonparametric Bayesian volatility estimation in SDE models is scarce. We can list theoretical contributions [34], [36], [56], and the practically oriented paper [2]. The model in the former two papers is close to the one considered in the present work, but from the methodological point of view different Bayesian priors are used and practical usefulness of the corresponding Bayesian approaches is limited. On the other hand, the models considered in [56] and [2] are rather different from ours, and so are the corresponding Bayesian approaches. The nearest predecessor of the model and the method in our paper is the one studied in [37]. In the sequel we will explain in what aspects the present contribution differs from that one and what the current improvements are. We note in passing that there exists a solid body of literature on nonparametric Bayesian estimation of the drift coefficient, see, e.g., [35], [57], [59], [65], [52], [51] and the review article [76], but Bayesian volatility estimation requires use of substantially different ideas. We also note existence of works dealing with parametric Bayesian estimation in discrete-time stochastic volatility models, see, e.g., [44] and [45], but again, these are not directly related to the problem we study in this paper.

1.3 Approach and results

The main potential difficulties facing a Bayesian approach to inference in SDE models from discrete observations are an intractable likelihood and absence of a closed form expression for the posterior distribution; see, e.g., [21], [25], [64] and [50]. Typically, these difficulties necessitate the use of a data augmentation device (see [69]) and some intricate form of a Markov chain Monte Carlo (MCMC) sampler (see [63]). In [37], these difficulties are circumvented by intentionally setting the drift coefficient to zero, and employing a (conjugate) histogram-type prior on the diffusion coefficient, that has piecewise constant realisations on bins forming a partition of $[0,T]$. Specifically, the (squared) volatility is modelled a priori as a function $s^2 = \sum_{k=1}^{N} \theta_k \mathbf{1}_{B_k}$, with independent and identically distributed inverse gamma coefficients θ_k's, and the prior Π is defined as the law of s^2. Here $B_1, \ldots, B_N$ are bins forming a partition of $[0,T]$. With this independent inverse Gamma (IIG) prior, $\theta_1, \ldots, \theta_N$ are independent, conditional on the data, and of inverse gamma type. Therefore, this approach results in a fast and simple to understand and implement Bayesian procedure. A study of its favourable practical performance, as well as its theoretical validation was recently undertaken in [37]. As shown there under precise regularity conditions, misspecification of the drift is asymptotically, as the sample size $n \to \infty$, harmless for consistent estimation of the volatility coefficient.

Despite a good practical performance of the method in [37], there are some limitations associated with it too. Thus, the method offers limited possibilities for adaptation to the local structure of the volatility coefficient, which may become an issue if the volatility has a wildly varying curvature on the time interval $[0,T]$. A possible

fix to this would be to equip the number of bins N forming a partition of $[0,T]$ with a prior, and choose the endpoints of bins B_k also according to a prior. However, this would force one to go beyond the conjugate Bayesian setting as in [37], and posterior inference in practice would require, for instance, the use of a reversible jump MCMC algorithm (see [32]). Even in the incomparably simpler setting of intensity function estimation for nonhomogeneous Poisson processes with histogram-type priors, this is very challenging, as observed in [75]. Principal difficulties include designing moves between models of differing dimensions that result in MCMC algorithms that *mix well*, and *assessment of convergence* of Markov chains (see [22], p. 204). Thus, e.g., the inferential conclusions in [32] and [33] are different on the same real data example using the same reversible jump method, since it turned out that in the first paper the chain was not run long enough. Cf. also the remarks on Bayesian histograms in [27], p. 546.

Here we propose an alternative approach, inspired by ideas in [7] in the context of audio signal modelling different from the SDE setting that we consider; see also [8], [9], [15], [16] and [71]. Namely, instead of using a prior on the (squared) volatility that has piecewise constant realisations on $[0,T]$ with independent coefficients θ_k's, we will assume that the sequence $\{\theta_k\}$ forms a suitably defined Markov chain. An immediately apparent advantage of using such an approach is that it induces extra smoothing via dependence in prior realisations of the volatility function across different bins. Arguing heuristically, with a large number N of bins B_k it is then possible to closely mimic the local structure of the volatility: in those parts of the interval $[0,T]$, where the volatility has a high curvature or is subject to abrupt changes, a large number of (narrow) bins is required to adequately capture these features. However, the grid used to define the bins B_k's is uniform, and if $\theta_1, \ldots, \theta_N$ are a priori independent, a large N may induce spurious variability in the volatility estimates in those regions of $[0,T]$ where the volatility in fact varies slowly. As we will see in the sequel, this problem may be alleviated using a priori dependent θ_k's.

In the subsequent sections we detail our approach, and study its practical performance via simulation and real data examples. Specifically, we implement our method via a straightforward version of the Gibbs sampler, employing the fact that full conditional distributions of θ_k's are known in closed form (and are in fact inverse gamma). Unlike [37], posterior inference in our new approach requires the use of MCMC. However, this is offset by the advantages of our new approach outlined above, and in fact the additional computational complexity of our new method is modest in comparison to [37]. The prior in our new method depends on hyperparameters, and we will also discuss several ways of their choice in practice.

1.4 Organisation of this paper

In Section 2 we supply a detailed description of our nonparametric Bayesian approach to volatility estimation. In Section 3 we study the performance of our method via extensive simulation examples. In Section 4 we apply the method on a real data

example. Section 5 summarises our findings and provides an outlook on our results. Finally, Section 6 contains some additional technical details of our procedure.

1.5 Notation

We denote the prior distribution on the (squared) volatility function by Π and write the posterior measure given data $\mathscr{X}_n$ as $\Pi(\cdot \mid \mathscr{X}_n)$. We use the notation $\mathrm{IG}(\alpha, \beta)$ for the inverse gamma distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$. This distribution has a density

$$x \mapsto \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\beta/x}, \quad x > 0. \tag{2}$$

For two sequences $\{a_n\}$, $\{b_n\}$, the notation $a_n \asymp b_n$ will be used to denote the fact that the sequences are asymptotically (as $n \to \infty$) of the same order. Finally, for a density f and a function g, the notation $f \propto g$ will mean that f is proportional to g, with proportionality constant on the righthand side recovered as $(\int g)^{-1}$, where the integral is over the domain of definition of g (and of f). The function g can be referred to as an unnormalised probability density.

2 Nonparametric Bayesian approach

2.1 Generalities

Our starting point is the same as in [37]. Namely, we misspecify the drift coefficient b_0 by intentionally setting it to zero (see also [48] for a similar idea of 'misspecification on purpose'). The theoretical justification for this under the 'infill' asymptotics, with the time horizon T staying fixed and the observation times $t_{i,n} = iT/n$, $i = 1, \ldots, n$, filling up the interval $[0,T]$ as $n \to \infty$, is provided in [37], to which we refer for further details (the argument there ultimately relies on Girsanov's theorem). Similar ideas are also encountered in the non-Bayesian setting in the econometrics literature on high-frequency financial data, see, e.g., [55].

Set $Y_{i,n} = X_{t_{i,n}} - X_{t_{i-1,n}}$. With the assumption $b_0 = 0$, the pseudo-likelihood of our observations is tractable, in fact Gaussian,

$$L_n(s^2) = \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi \int_{t_{i-1,n}}^{t_{i,n}} s^2(u)\, du}} \psi\left(\frac{Y_{i,n}}{\sqrt{\int_{t_{i-1,n}}^{t_{i,n}} s^2(u)\, du}} \right) \right\}, \tag{3}$$

where $\psi(u) = \exp(-u^2/2)$. The posterior probability of any measurable set S of volatility functions can be computed via Bayes' theorem as

$$\Pi(S \mid \mathscr{X}_n) = \frac{\int_S L_n(s^2)\Pi(\mathrm{d}s)}{\int L_n(s^2)\Pi(\mathrm{d}s)}.$$

Here the denominator is the normalising constant, the integral over the whole space on which the prior Π is defined, which ensures that the posterior is a *probability* measure (i.e. integrates to one).

2.2 Prior construction

Our prior Π is constructed similarly to [37], with an important difference to be noted below. Fix an integer $m < n$. Then $n = mN + r$ with $0 \le r < m$, where $N = \lfloor \frac{n}{m} \rfloor$. Now define bins $B_k = [t_{m(k-1),n}, t_{mk,n})$, $k = 1,\dots,N-1$, and $B_N = [t_{m(N-1),n}, T]$. Thus the first $N-1$ bins are of length mT/n, whereas the last bin B_N has length $T - t_{m(N-1),n} = n^{-1}(r+m)T < n^{-1}2mT$. The parameter N (equivalently, m) is a hyperparameter of our prior. We model s as piecewise constant on bins B_k, thus $s = \sum_{k=1}^N \xi_k \mathbf{1}_{B_k}$. The prior Π on the volatility s can now be defined by assigning a prior to the coefficients ξ_k's.

Let $\theta_k = \xi_k^2$. Since the bins B_k are disjoint,

$$s^2 = \sum_{k=1}^N \xi_k^2 \mathbf{1}_{B_k} = \sum_{k=1}^N \theta_k \mathbf{1}_{B_k}.$$

As the likelihood depends on s only through its square s^2, it suffices to assign the prior to the coefficients θ_k's of s^2. This is the point where we fundamentally diverge from [37]. Whereas in [37] it is assumed that $\{\theta_k\}$ is an i.i.d. sequence of inverse gamma random variables, here we suppose that $\{\theta_k\}$ forms a Markov chain. This will be referred to as an inverse Gamma Markov chain (IGMC) prior (see [7]), and is defined as follows. Introduce auxiliary variables ζ_k, $k = 2,\dots,N$, and define a Markov chain using the time ordering $\theta_1, \zeta_2, \theta_2, \dots, \zeta_k, \theta_k, \dots, \zeta_N, \theta_N$. Transition distributions of this chain are defined as follows: fix hyperparameters α_1, α_ζ and α, and set

$$\theta_1 \sim \mathrm{IG}(\alpha_1, \alpha_1), \quad \zeta_{k+1} \mid \theta_k \sim \mathrm{IG}(\alpha_\zeta, \alpha_\zeta \theta_k^{-1}), \quad \theta_{k+1} \mid \zeta_{k+1} \sim \mathrm{IG}(\alpha, \alpha \zeta_{k+1}^{-1}). \quad (4)$$

The name of the chain reflects the fact that these distributions are inverse Gamma.

Remark 1. Our definition of the IGMC prior differs from the one in [7] in the choice of the initial distribution of θ_1, which is important to alleviate possible 'edge effects' in volatility estimates in a neighbourhood of $t = 0$. The parameter α_1 determines the initial distribution of the inverse Gamma Markov chain. Letting $\alpha_1 \to 0$ (which corresponds to a vague prior) 'releases' the chain at the time origin. $\square$

Remark 2. As observed in [7], there are various ways of defining an inverse Gamma Markov chain. The point to be kept in mind is that the resulting posterior should

be computationally tractable, and the prior on θ_k's should have a capability of producing realisations with positive correlation structures, as this introduces smoothing among the θ_k's in adjacent bins. This latter property is not possible to attain with arbitrary constructions of inverse Gamma Markov chains, such as e.g. a natural construction $\theta_k | \theta_{k-1} \sim \text{IG}(\alpha, \theta_{k-1}/\alpha)$. On the other hand, positive correlation between realisations θ_k's can be achieved e.g. by setting $\theta_k | \theta_{k-1} \sim \text{IG}(\alpha, (\alpha\theta_{k-1})^{-1})$, but this results in intractable posterior computations. The definition of the IGMC prior in the present work, that employs latent variables ζ_k's, takes care of both these important points. For an additional discussion see [7].   □

Remark 3. Setting the drift coefficient b_0 to zero effectively results in pretending that the process X has independent (Gaussian) increments. In reality, since the drift in practical applications is typically nonzero, increments of the process are dependent, and hence all observations $Y_{i,n}$ contain some indirect information on the value of the volatility s^2 at each time point $t \in [0,T]$. On the other hand, assuming the IGMC prior on s^2 yields a posteriori dependence of coefficients $\{\theta_k\}$, which should be of help in inference with smaller sample sizes n. See Section 4 for an illustration.   □

2.3 Gibbs sampler

It can be verified by direct computations employing (4) that the full conditional distributions of θ_k's and ζ_k's are inverse gamma,

$$\theta_k | \zeta_k, \zeta_{k+1} \sim \text{IG}\left(\alpha + \alpha_\zeta, \frac{\alpha}{\zeta_k} + \frac{\alpha_\zeta}{\zeta_{k+1}}\right), \quad k = 2,\ldots,N-1, \tag{5}$$

$$\theta_1 | \zeta_2 \sim \text{IG}\left(\alpha_1 + \alpha_\zeta, \alpha_1 + \frac{\alpha_\zeta}{\zeta_2}\right), \tag{6}$$

$$\theta_N | \zeta_N \sim \text{IG}\left(\alpha, \frac{\alpha}{\zeta_N}\right), \tag{7}$$

$$\zeta_k | \theta_k, \theta_{k-1} \sim \text{IG}\left(\alpha_\zeta + \alpha, \frac{\alpha_\zeta}{\theta_{k-1}} + \frac{\alpha}{\theta_k}\right), \quad k = 2,\ldots,N. \tag{8}$$

See Section 6 for details. Next, the effective transition kernel of the Markov chain $\{\theta_k\}$ is given by formula (4) in [7], and is a scale mixture of inverse gamma distributions; however, its exact expression is of no direct concern for our purposes. As noted in [7], p. 700, depending on the parameter values α, α_ζ, it is possible for the chain $\{\theta_k\}$ to exhibit either an increasing or decreasing trend. We illustrate this point by plotting realisations of $\{\theta_k\}$ in Figure 1 for different values of α and α_ζ. In the context of volatility estimation this feature is attractive, if prior information on the volatility trend is available.

Inference in [7] is performed using a mean-field variational Bayes approach, see, e.g., [5]. Here we describe instead a fully Bayesian approach relying on Gibbs sampling (see, e.g., [26] and [29]), cf. [9].
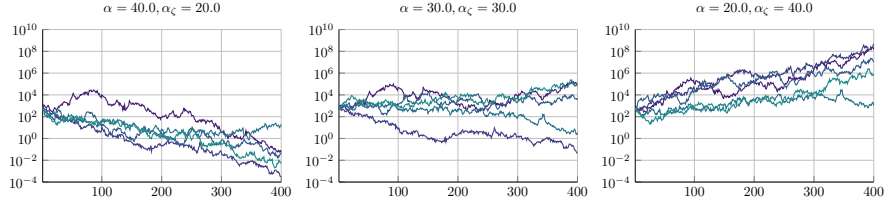
Fig. 1 Realisations of the Markov chain $\{\theta_k\}$ with $\alpha = 40$, $\alpha_\zeta = 20$ (left panel) and $\alpha = 30$, $\alpha_\zeta = 30$ (center panel) and $\alpha = 20$, $\alpha_\zeta = 40$ (right panel). In all cases, θ_1 is fixed to 500 .

The algorithm is initialised at values $\zeta_2, \ldots, \zeta_N$, e.g. generated from the prior specification (4). In order to derive update formulae for the full conditionals of the θ_k's, define

$$Z_k = \sum_{i=(k-1)m+1}^{km} Y_{i,n}^2, \quad k = 1, \ldots, N-1,$$

$$Z_N = \sum_{i=(N-1)m+1}^{n} Y_{i,n}^2.$$

With this notation, the likelihood from (3) satisfies

$$L_n(\theta) \propto \theta_N^{-(m+r)/2} \exp\left(-\frac{nZ_N}{2T\theta_N}\right) \prod_{k=1}^{N-1} \theta_k^{-m/2} \exp\left(-\frac{nZ_k}{2T\theta_k}\right).$$

Using this formula and equation (5), and recalling the form of the inverse gamma density (2), it is seen that the update distribution for θ_k, $k = 2, \ldots, N-1$, is

$$\mathrm{IG}\left(\alpha + \alpha_\zeta + \frac{m}{2}, \frac{\alpha}{\zeta_k} + \frac{\alpha_\zeta}{\zeta_{k+1}} + \frac{nZ_k}{2T}\right),$$

whereas by (7) the ones for θ_1 and θ_N are

$$\mathrm{IG}\left(\alpha_1 + \alpha_\zeta + \frac{m}{2}, \alpha_1 + \frac{\alpha_\zeta}{\zeta_2} + \frac{nZ_1}{2T}\right), \quad \mathrm{IG}\left(\alpha + \frac{m+r}{2}, \frac{\alpha}{\zeta_N} + \frac{nZ_N}{2T}\right),$$

respectively.

Next, the latent variables ζ_k's will be updated using formula (8). This update step for ζ_k's does not directly involve the data $\mathscr{X}_n$, except through the previous values of θ_k's.

Finally, one iterates these two Gibbs steps for θ_k's and ζ_k's a large number of times (until chains can be assessed as reasonably converged), which gives posterior samples of the θ_k's. Using the latter, the posterior inference can proceed in the usual way, e.g. by computing the sample posterior mean of θ_k's, as well as sample quantiles, that provide, respectively, a point estimate and uncertainty quantification via

marginal Bayesian credible bands for the squared volatility s^2. Similar calculations on the square roots of the posterior samples can be used to obtain point estimates and credible bands for the volatility function s itself.

2.4 Hyperparameter choice

Like [37], our Bayesian procedure depends on the hyperparameter N. As argued in [37], in practice it is recommended to use the theoretical results in [37] (that suggest to take $N \asymp n^{\lambda/(2\lambda+1)}$, if the true volatility function s_0 is λ-Hölder smooth) and try several values of N simultaneously. Different N's all provide information on the unknown volatility, but at different resolution levels; see Section 5 in [37] for an additional discussion.

Additionally, we have hyperparameters α, α_ζ and α_1, that govern properties of the Markov chain prior. In [7], where an IGMC prior was introduced, guidance on the hyperparameter selection is not discussed. In [8], the hyperparameters are fine-tuned by hand in specific problems studied there (audio denoising and single channel audio source separation). Another practical solution is to try several different fixed combinations of the hyperparameters α, α_ζ and α_1, if only to verify sensitivity of inferential conclusions with respect to variations in the hyperparameters. Some further methods for hyperparameter optimisation are discussed in [16]. In [8] optimisation of the hyperparameters via the maximum likelihood method is suggested; practical implementation relies on the EM algorithm (see [13]), and some additional details are given in [15]. Put in other terms, the proposal in [15] amounts to using an empirical Bayes method (see, e.g., [20], [61] and [62]). The use of the latter is widespread and often leads to good practical results, but the method is still insufficiently understood theoretically, except in toy models like the white noise model (see, however, [17] and [58] for some results in other contexts). On the practical side, in our case, given that the dimension of the sequences $\{\zeta_k\}$ and $\{\theta_k\}$ is rather high, namely $2N-1$ with N large, and the marginal likelihood is not available in closed form, this approach is expected to be computationally intensive. Therefore, a priori there is no reason not to try instead a fully Bayesian approach by equipping the hyperparameters with a prior, and this is in fact our default approach in the present work. However, the corresponding full conditional distribution turns out to be nonstandard, which necessitates the use of a Metropolis-Hastings step within the Gibbs sampler (see, e.g., [38], [49] and [70]). We provide the necessary details in Section 6.

3 Synthetic data examples

Computations in this section have been done in the programming language Julia, see [3]. In order to test the practical performance of our estimation method, we use

a challenging example with the blocks function from [18]. As a second example, we consider the case of the Cox-Ross-Ingersoll model. Precise details are given in the subsections below.

We used the Euler scheme on a grid with 800 001 equidistant points on the interval $[0,1]$ to obtain realisations of a solution to (1) for different combinations of the drift and dispersion coefficients. These were then subsampled to obtain $n = 4000$ observations in each example.

The hyperparameter α_1 was set to 0.1, whereas for the other two hyperparameters we assumed that $\alpha = \alpha_\zeta$ and used a diffuse $\mathrm{IG}(0.3, 0.3)$ prior, except in specially noted cases below. Inference was performed using the Gibbs sampler from Section 2, with a Metropolis-Hastings step to update the hyperparameter α. The latter used an independent Gaussian random walk proposal with a scaling to ensure the acceptance rate of ca. 50%; see Section 6. The Gibbs sampler was run for 200 000 iterations and we used a burn-in of 1000 samples. In each example we plotted 95% marginal credible bands obtained from the central posterior intervals for the coefficients $\xi_k = \sqrt{\theta_k}$.

3.1 Blocks function

As our first example, we considered the case when the volatility function was given by the blocks function from [18]. With a vertical shift for positivity, this is defined as follows:

$$s(t) = 10 + 3.655606 \times \sum_{j=1}^{11} h_j K(t - t_j), \quad t \in [0,1], \tag{9}$$

where $K(t) = (1 + \mathrm{sgn}(t))/2$, and

$$\{t_j\} = (0.1, 0.13, 0.15, 0.23, 0.25, 0.4, 0.44, 0.65, 0.76, 0.78, 0.81),$$
$$\{h_j\} = (4, -5, 3, -4, 5, -4.2, 2.1, 4.3, -3.1, 2.1, -4.2).$$

The function serves as a challenging benchmark example in nonparametric regression: it is mostly very smooth, but spatially inhomogeneous and characterised by abrupt changes (cf. Chapter 9 in [72]). Unlike nonparametric regression, the noise (Wiener process) in our setting should be thought of as multiplicative and proportional to s rather than additive, which combined with the fact that s takes rather large values further complicates the inference problem. Our main goal here was to compare the performance of the IGMC prior-based approach to the IIG prior-based one from [37]. To complete the SDE specification, our drift coefficient was chosen to be a rather strong linear drift $b_0(x) = -10x + 20$.

In Figure 2 we plot the blocks function (9) and the corresponding realisation of the process X used in this simulation run.

The left and right panels of Figure 3 contrast the results obtained using the IGMC prior with $N = 160$ and $\alpha = \alpha_\zeta = 20$ versus $N = 320$ and $\alpha = \alpha_\zeta = 40$. These

plots illustrate the fact that increasing N has the effect of undersmoothing prior realisations, that can be balanced by increasing the values of α_ζ, α, which has the opposite smoothing effect. Because of this, in fact, both plots look quite similar.
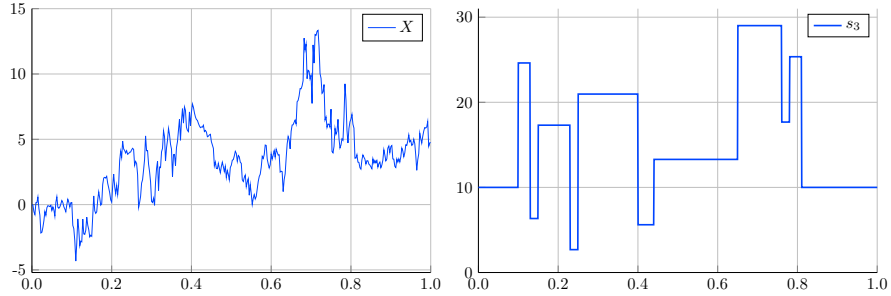

Fig. 2 The sample path of the process X from (9) (left panel) and the corresponding volatility function s (right panel).
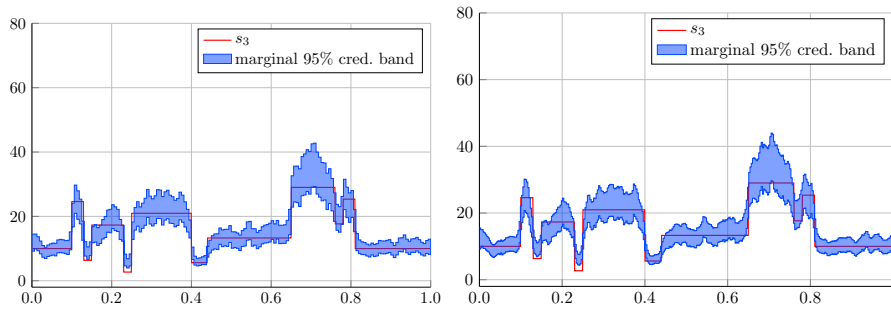

Fig. 3 Volatility function s from (9) with superimposed 95% marginal credible band for the IGMC prior, using $N = 160$, $\alpha = \alpha_\zeta = 20$ (left panel) and $N = 320$, $\alpha = \alpha_\zeta = 40$ (right panel); in both cases, $\alpha_1 = 0.1$.

The top left and top right panels of Figure 4 give estimation results obtained with the IIG prior-based approach from [37]. The number of bins was again $N = 160$ and $N = 320$, and in both these cases we used diffuse independent $IG(0.1, 0.1)$ priors on the coefficients of the (squared) volatility function (see [37] for details). These results have to be contrasted to those obtained with the IGMC prior, plotted in the bottom left and bottom right panels of Figure 4, where we assumed $\alpha_1 = 0.1$ and $\alpha = \alpha_\zeta \sim IG(0.3, 0.3)$. The following conclusions emerge from Figure 4:

- Although both the IGMC and IIG approaches recover globally the shape of the volatility function, the IIG approach results in much greater uncertainty in inferential conclusions, as reflected in wider marginal confidence bands. The effect is

especially pronounced in the case $N = 320$, where the width of the band for the IIG prior renders it almost useless for inference.

- The bands based on the IGMC prior look more 'regular' than the ones for the IIG prior.
- Comparing the results to Figure 3, we see the benefits of equipping the hyperparameters α, α_ζ with a prior: credible bands in Figure 3 do not adequately capture two dips of the function s right before and after the point $t = 0.2$, since s completely falls outside the credible bands there. Thus, an incorrect amount of smoothing is used in Figure 3.
- The method based on the IIG prior is sensitive to the bin number selection: compare the top left panel of Figure 4 using $N = 160$ bins to the top right panel using $N = 320$ bins, where the credible band is much wider. On the other hand, the method based on the IGMC prior automatically rebalances the amount of smoothing it uses with different numbers of bins N, thanks to the hyperprior on the parameters α, α_ζ; in fact, the bottom two plots in Figure 4 look similar to each other.
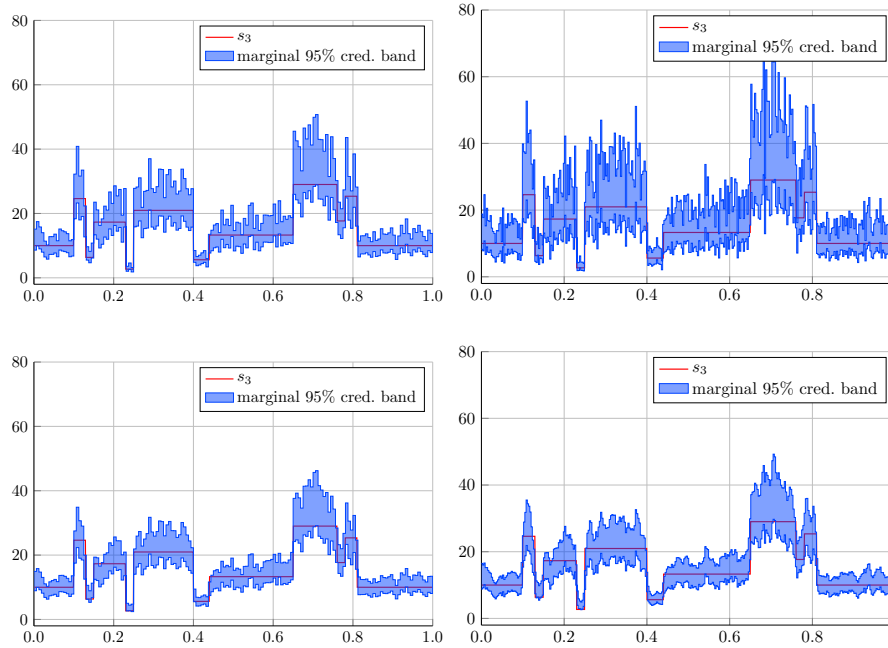

Fig. 4 Volatility function s with superimposed 95% marginal credible band for the IIG prior $\mathrm{IG}(0.1, 0.1)$, using $N = 160$ (top left panel) and $N = 320$ bins (top right panel). Volatility function s from (9) with superimposed 95% marginal credible band for the IGMC prior, using $N = 160$ (bottom left panel) and $N = 320$ bins (bottom right panel); in both cases, $\alpha_1 = 0.1$ and $\alpha = \alpha_\zeta \sim \mathrm{IG}(0.3, 0.3)$.

3.2 CIR model

Our core estimation procedure, as described in the previous sections, assumes that the volatility function is deterministic. In this subsection, however, in order to test the limits of applicability of our method and possibilities for future extensions, we applied it to a case where the volatility function was stochastic. The study in [55] lends support to this approach, but here we concentrate on practical aspects and defer the corresponding theoretical investigation until another occasion.

Specifically, we considered the Cox-Ross-Ingersoll (CIR) model or the square root process,

$$\mathrm{d}X_t = (\eta_1 - \eta_2 X_t)\mathrm{d}t + \eta_3\sqrt{X_t}\mathrm{d}W_t, \quad X_0 = x > 0, \quad t \in [0,T]. \tag{10}$$

Here $\eta_1, \eta_2, \eta_3 > 0$ are parameters of the model. This diffusion process was introduced in [23] and [24], and gained popularity in finance as a model for short-term interest rates, see [11]. The condition $2\eta_1 > \eta_3^2$ ensures strict positivity and ergodicity of X. The volatility function s_0 from (1) now corresponds to a realisation of a stochastic process $t \mapsto \eta_3\sqrt{X_t}$, where X solves the CIR equation (10).

We took arbitrary parameter values

$$\eta_1 = 6, \quad \eta_2 = 3, \quad \eta_1 = 2, \quad x = 1. \tag{11}$$

A sample path of X is plotted in the left panel of Figure 5, whereas the corresponding volatility is given in the right panel of the same figure. In Figure 6 we display estimation results obtained with the IGMC prior, using $N = 160$ and $N = 320$ bins and hyperparameter specifications $\alpha_1 = 0.1$ and $\alpha = \alpha_\zeta \sim \mathrm{IG}(0.3, 0.3)$. A conclusion that emerges from this figure is that our Bayesian method captures the overall shape of the realised volatility in a rather satisfactory manner.
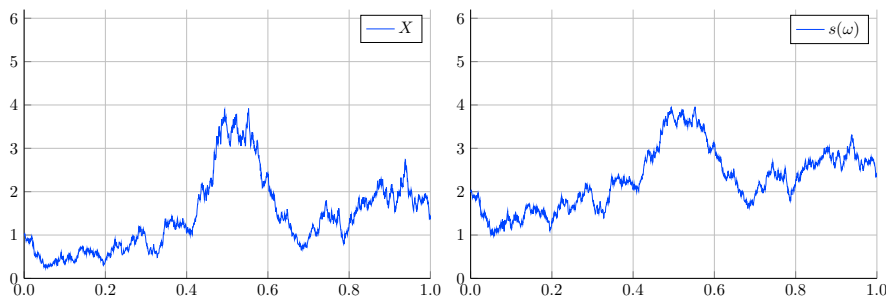

Fig. 5 The sample path of the process X from (10) (left panel) and the corresponding realised volatility function $s(\omega)$ (right panel). The parameter values are given in (11).
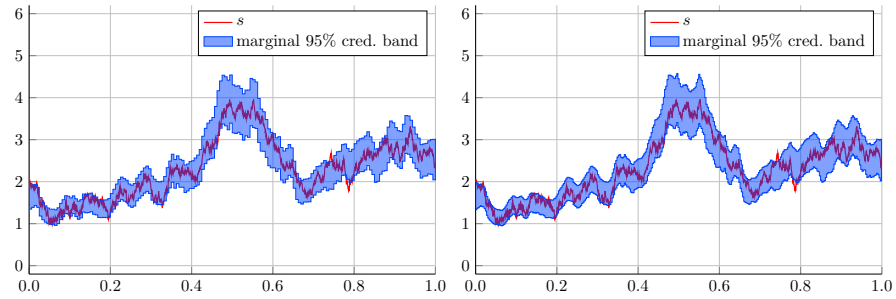
Fig. 6 Volatility function s from (10) with superimposed 95% marginal credible band for the IGMC prior, using $N = 160$ (left panel) and $N = 320$ bins (right panel); in both cases, $\alpha_1 = 0.1$ and $\alpha = \alpha_\zeta \sim \mathrm{IG}(0.3, 0.3)$.

4 Dow-Jones industrial averages

In this section we provide a reanalysis of a classical dataset in change-point detection in time series; see, e.g., [10], [14], [40], [41] and [42]. Specifically, we consider weekly closing values of the Dow-Jones industrial averages in the period 2 July 1971 – 2 August 1974. In total there are 162 observations available, which constitute a relatively small sample, and thus the inference problem is rather nontrivial. The data can be accessed as the dataset DWJ in the sde package (see [43]) in **R** (see [60]). See the left panel of Figure 7 for a visualisation. In [42] the weekly data X_{t_i}, $i = 1, \ldots, n$, are transformed into returns $Y_{t_i} = (X_{t_i} - X_{t_{i-1}})/X_{t_{i-1}}$, and the least squares change-point estimation procedure from [12] has been performed. Reproducing the corresponding computer code in **R** results in a change-point estimate of 16 March 1973. That author speculates that this change-point is related to the Watergate scandal.
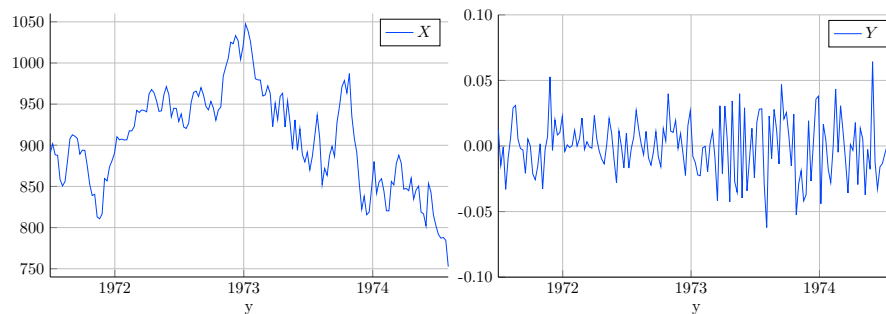

Fig. 7 Dow-Jones weekly closings of industrial averages over the period 2 July 1971 – 2 August 1974 (left panel) and the corresponding returns (right panel).

Similar to [42], parametric change-point analyses in [10], [14] and [41] give a change-point in the third week of March 1973. However, as noted in [42], examination of the plot of the time series Y_{t_i} (see Figure 7, the right panel) indicates that another change-point may be present in the data. Then dropping observations after 16 March 1973 and analysing the data for existence of a change-point using only the initial segment of the time series, the author discovers another change-point on 17 December 1971, which he associates with suspending the convertibility of the US dollar into gold under President Richard Nixon's administration.

From the above discussion it should be clear that nonparametric modelling of the volatility may provide additional insights for this dataset. We first informally investigated the fact whether an SDE driven by the Wiener process is a suitable model for the data at hand. Many of such models, e.g. the geometric Brownian motion, a classical model for evolution of asset prices over time (also referred to as the Samuelson or Black-Scholes model), rely on an old tenet that returns of asset prices follow a normal distribution. Although the assumption has been empirically disproved for high-frequency financial data (daily or intraday data; see, e.g., [6], [19] and [47]), its violation is less severe for widely spaced data in time (e.g. weekly data, as in our case). In fact, the Shapiro-Wilk test that we performed in **R** on the returns past the change-point 16 March 1973 did not reject the null hypothesis of normality (*p*-value 0.4). On the other hand, the quantile-quantile (QQ) plot of the same data does perhaps give an indication of a certain mild deviation from normality, see Figure 8, where we also plotted a kernel density estimate of the data (obtained via the command density in **R**, with bandwidth determined automatically through cross-validation).
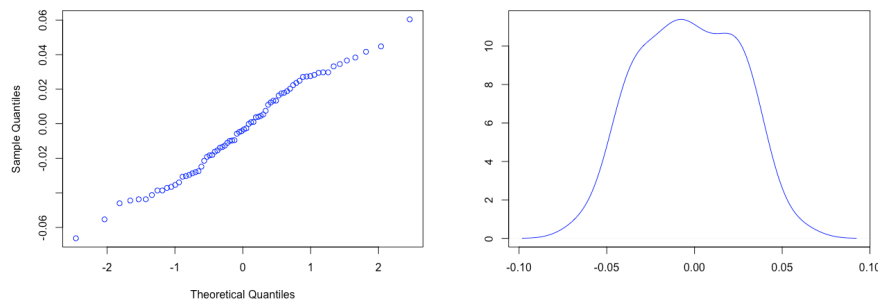

Fig. 8 QQ plot of the returns of Dow-Jones weekly closings of industrial averages over the period 16 March 1973 – 2 August 1974 (left panel) and a kernel density estimate of the same data (right panel).

In Figure 9 we plot the sample autocorrelation and partial autocorrelation functions based on returns Y_{t_i}'s past the change-point 16 March 1973. These do not give decisive evidence against the assumption of independence of Y_{t_i}'s. Neither does the Ljung-Box test (the test is implemented in **R** via the command Box.test), which

yields a *p*-value 0.057 when applied with 10 lags (the *p*-value is certainly small, but not overwhelmingly so).
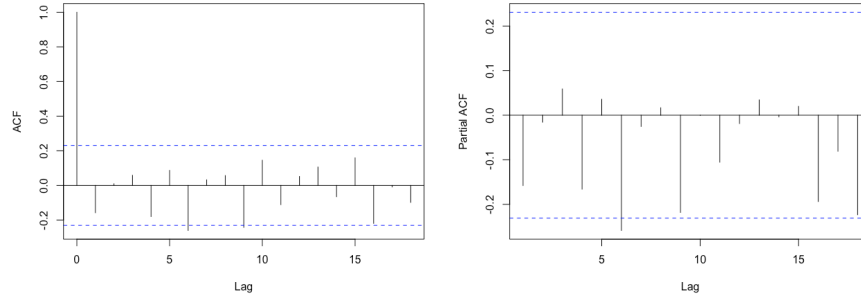

Fig. 9 Sample autocorrelation (left panel) and partial autocorrelation functions of the returns of Dow-Jones weekly closings of industrial averages over the period 16 March 1973 – 2 August 1974.

Summarising our findings, we detected only a mild evidence against the assumption that the returns of the Dow-Jones weekly closings of industrial averages (over the period 16 March 1973 – 2 August 1974, but similar conclusions can be reached also over the other subperiods covered by the DWJ dataset) are approximately independent and follow a normal distribution. Thus there is no strong a priori reason to believe that a geometric Brownian motion is an outright unsuitable model in this setting: it can be used as a first approximation. To account for time-variability of volatility (as suggested by the change-point analysis), we incorporate a time-dependent volatility function in the model, and for additional modelling flexibility we also allow a state-dependent drift. Setting $Z_t = \log(X_t/X_0)$, our model is thus given by

$$\mathrm{d}Z_t = b_0(t, Z_t)\mathrm{d}t + s_0(t)\mathrm{d}W_t, \quad Z_0 = 0. \tag{12}$$

An alternative here is to directly (i.e. without any preliminary transformation) model the Dow-Jones data using equation (1). We consider both possibilities, starting with the model (12).

As in our simulation examples in Section 3, we used a vague prior on θ_1 corresponding to the limit $\alpha_1 \to 0$, whereas for the other two hyperparameters we assumed $\alpha = \alpha_\zeta \sim \mathrm{IG}(0.3, 0.3)$. The scaling in the independent Gaussian random walk proposal in the Metropolis-Hastings step was chosen in such a way so as to yield an acceptance rate of ca. 50%. The Gibbs sampler was run for 200 000 iterations, and the first 1000 samples were dropped as a burn-in. We present the estimation results we obtained using $N = 13$ and $N = 26$ bins, see Figure 10. Although the sample size *n* is quite small in this example, the data are informative enough to yield nontrivial inferential conclusions even with diffuse priors. Both plots in Figure 10 are qualitatively similar and suggest:

- A decrease in volatility at the end of 1971, which can be taken as corresponding to the change-point in December 1971 identified in [42]. Unlike that author, we do not directly associate it with suspending the convertibility of the US dollar into gold (that took place in August 1971 rather than December 1971).
- A gradual increase in volatility over the subsequent period stretching until the end of 1973. Rather than only the Watergate scandal (and a change-point in March 1973 as in [42]), there could be further economic causes for that, such as the 1973 oil crisis and the 1973–1974 stock market crash.
- A decrease in volatility starting in early 1974, compared to the immediately preceding period.

In general, in this work we do not aim at identifying causes for changes in volatility regimes, but prefer to present our inference results, that may subsequently be used in econometric analyses.
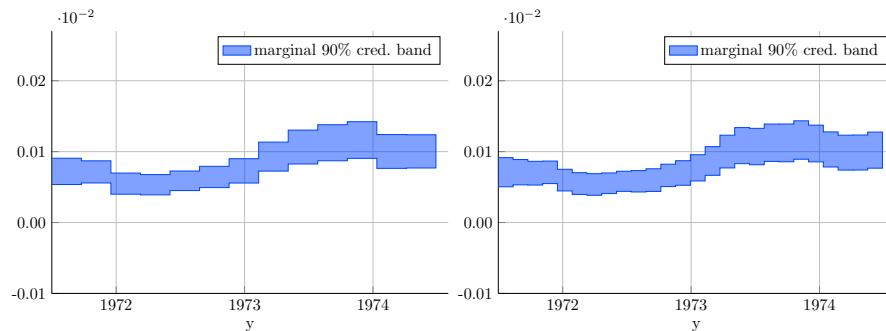

Fig. 10 Marginal 90% credible bands for the volatility function of the log Dow-Jones industrial averages data. The left panel corresponds to $N = 13$ bins, while the right panel to $N = 26$ bins.

Now we move to the Bayesian analysis of the data using model (1). The prior settings were as in the previous case, and we display the results in Figure 11. The overall shapes of the inferred volatility functions are the same in both Figure 10 and Figure 11, and hence similar conclusions apply.

Finally, we stress the fact that our nonparametric Bayesian approach and change-point estimation are different in their scope: whereas our method aims at estimation of the entire volatility function, change-point estimation (as its name actually suggests) concentrates on identifying change-points in the variance of the observed time series, which is a particular feature of the volatility. To that end it assumes the (true) volatility function is piecewise constant, which on the other hand is not an assumption required in our method. Both techniques are useful, and each can provide insights that may be difficult to obtain from another.
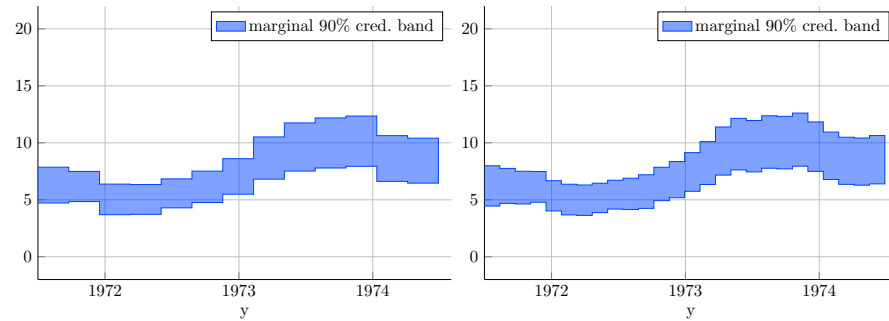
Fig. 11 Marginal 90% credible bands for the volatility function of the Dow-Jones industrial averages data. The left panel corresponds to $N = 13$ bins, while the right panel to $N = 26$ bins.

5 Conclusions

Bayesian inference for SDEs from discrete-time observations is a difficult task, owing to intractability of the likelihood and the fact that the posterior is not available in closed form. Posterior inference therefore typically requires the use of intricate MCMC samplers. Designing algorithms that result in Markov chains that mix well and explore efficiently the posterior surface is a highly nontrivial problem. Inspired by some ideas from the audio signal processing literature and our earlier work [37], in this paper we introduced a novel nonparametric Bayesian approach to estimation of the volatility coefficient of an SDE. Our method is easy to understand and straightforward to implement via Gibbs sampling, and performs well in practice. Thereby our hope is that our work will contribute to further dissemination and popularisation of a nonparametric Bayesian approach to inference in SDEs, specifically with financial applications in mind. Our work can also be viewed as a partial fulfillment of anticipation in [31] that some ideas developed originally in the context of audio and music processing "will also find use in other areas of science and engineering, such as financial or biomedical data analysis".

As a final remark, we do not attempt to provide a theoretical, i.e. asymptotic frequentist analysis of our new approach here (see, e.g., the recent monograph [30], and specifically [37] for such an analysis in the SDE setting), but leave this as a topic of future research.

6 Formulae for parameter updates

In this section we present additional details on the derivation of the update formulae for the Gibbs sampler from Section 2. The starting point is to employ the Markov property from (4), and using the standard Bayesian notation, to write the joint density of $\{\zeta_k\}$ and $\{\theta_k\}$ as

$$p(\theta_1)\prod_{k=2}^{N}p(\zeta_k|\theta_{k-1})p(\theta_k|\zeta_k). \tag{13}$$

6.1 Full conditional distributions

We first indicate how (5) was derived. Insert expressions for the individual terms in (13) from (4) and collect separately terms that depend on θ_k only, to see that the density of the full conditional distribution of θ_k, $k = 2, \ldots, N-1$, is proportional to

$$\theta_k^{-\alpha-1}e^{-\alpha/(\theta_k\zeta_k)}\theta_k^{-\alpha_\zeta}e^{-\alpha_\zeta/(\theta_k\zeta_{k+1})}.$$

Upon normalisation, this expression is the density of the $\mathrm{IG}(\alpha + \alpha_\zeta, \alpha\zeta_k^{-1} + \alpha_\zeta\zeta_{k+1}^{-1})$ distribution, which proves fromula (5). Formula (7) follows directly from the last expression in (4). Formula (8) is proved analogously to (5). Finally, (6) follows from (4) and Bayes' formula. Cf. also [15], Appendix B.6.

6.2 Metropolis-within-Gibbs step

Now we describe the Metropolis-Hastings step within the Gibbs sampler, that is used to update the hyperparameters of our algorithm, in case the latter are equipped with a prior. For simplicity, assume $\alpha = \alpha_\zeta$ (we note that such a choice is used in practical examples in [8]), and suppose α is equipped with a prior, $\alpha \sim \pi$. Let the hyperparameter α_1 be fixed. Obviously, α_1 could have been equipped with a prior as well, but this would have further slowed down our estimation procedure, whereas the practical results in Sections 3 and 4 we obtained are already satisfactory with α_1 fixed. Using (4) and (13), one sees that the joint density of $\{\zeta_k\}$, $\{\theta_k\}$ and α is proportional to

$$\pi(\alpha) \times \theta_1^{-\alpha_1-1} \times e^{-\alpha_1\theta_1^{-1}}$$
$$\times \prod_{k=2}^{N}\left\{\frac{\alpha^\alpha}{\Gamma(\alpha)\theta_{k-1}^\alpha}\zeta_k^{-\alpha-1}e^{-\alpha/(\theta_{k-1}\zeta_k)}\frac{\alpha^\alpha}{\Gamma(\alpha)\zeta_k^\alpha}\theta_k^{-\alpha-1}e^{-\alpha/(\theta_k\zeta_k)}\right\}.$$

This in turn is proportional to

$$q(\alpha) = \pi(\alpha) \times \left(\frac{\alpha^\alpha}{\Gamma(\alpha)}\right)^{2(N-1)} \times \prod_{k=2}^{N}(\theta_{k-1}\theta_k\zeta_k^2)^{-\alpha}$$
$$\times \exp\left(-\alpha\sum_{k=2}^{N}\frac{1}{\zeta_k}\left(\frac{1}{\theta_{k-1}}+\frac{1}{\theta_k}\right)\right).$$

The latter expression is an unnormalised full conditional density of α, and can be used in the Metropolis-within-Gibbs step to update α.

The rest of the Metropolis-Hastings step is standard, and the following approach was used in our practical examples: pick a proposal kernel $g(\alpha' \mid \alpha)$, for instance a Gaussian random walk proposal $g(\alpha' \mid \alpha) = \phi_\sigma(\alpha' - \alpha)$, where ϕ_σ is the density of a normal random variable with mean zero and variance σ^2. Note that this specific choice may result in proposing a negative value α', which needs to be rejected straightaway as invalid. Then, for computational efficiency, instead of moving to another step within the Gibbs sampler, one keeps on proposing new values α' until a positive one is proposed. This is then accepted with probability

$$A = \min \left(1, \frac{q(\alpha')}{q(\alpha)} \frac{\Phi_\sigma(\alpha)}{\Phi_\sigma(\alpha')} \right),$$

where $\Phi_\sigma(\cdot)$ is the cumulative distribution function of a normal random variable with mean zero and variance σ^2; otherwise the current value α is retained. See [73] for additional details and derivations. Finally, one moves to other steps in the Gibbs sampler, namely to updating ζ_k's and θ_k's, and iterates the procedure. The acceptance rate in the Metropolis-Hastings step can be controlled through the scale parameter σ of the proposal density ϕ_σ. Some practical rules for determination of an optimal acceptance rate in the Metropolis-Hastings algorithm are given in [28], and those for the Metropolis-within-Gibbs algorithm in [66].

References

1. Allen, E.: Modeling with Itô stochastic differential equations, *Mathematical Modelling: Theory and Applications*, vol. 22. Springer, Dordrecht (2007)
2. Batz, P., Ruttor, A., Opper, M.: Approximate Bayes learning of stochastic differential equations. ArXiv e-prints (2017)
3. Bezanson, J., Edelman, A., Karpinski, S., Shah, V.B.: Julia: a fresh approach to numerical computing. SIAM Rev. **59**(1), 65–98 (2017)
4. Björk, T.: Arbitrage Theory in Continuous Time, 3 edn. Oxford University Press (2009)
5. Blei, D.M., Kucukelbir, A., McAuliffe, J.D.: Variational inference: a review for statisticians. J. Amer. Stat. Assoc. **112**(518), 859–877 (2017)
6. Carr, P., Geman, H., Madan, D.B., Yor, M.: The fine structure of asset returns: an empirical investigation. J. Bus. **75**(2), 305–332 (2002)
7. Cemgil, A.T., Dikmen, O.: Conjugate gamma Markov random fields for modelling nonstationary sources. In: M.E. Davies, C.J. James, S.A. Abdallah, M.D. Plumbley (eds.) Independent Component Analysis and Signal Separation: 7th International Conference, ICA 2007, London, UK, September 9–12, 2007. Proceedings, pp. 697–705. Springer Berlin Heidelberg, Berlin, Heidelberg (2007)
8. Cemgil, A.T., Dikmen, O.: Inference and learning in gamma chains for Bayesian audio processing. In: Proceedings of Acoustics '08, pp. 4055–4060 (2008)

9. Cemgil, A.T., Févotte, C., Godsill, S.J.: Variational and stochastic inference for Bayesian source separation. Digit. Signal Process. **17**(5), 891–913 (2007). Special Issue on Bayesian Source Separation
10. Chen, J., Gupta, A.K.: Parametric statistical change point analysis, 2 edn. Birkhäuser/Springer, New York (2012). With applications to genetics, medicine, and finance
11. Cox, J.C., Ingersoll Jr., J.E., Ross, S.A.: A theory of the term structure of interest rates. Econometrica **53**(2), 385–407 (1985)
12. De Gregorio, A., Iacus, S.M.: Least squares volatility change point estimation for partially observed diffusion processes. Comm. Statist. Theory Methods **37**(13-15), 2342–2357 (2008)
13. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Stat. Soc. B. Met. **39**(1), 1–38 (1977)
14. Díaz, J.: Bayesian detection of a change of scale parameter in sequences of independent gamma random variables. J. Econometrics **19**(1), 23–29 (1982)
15. Dikmen, O., Cemgil, A.T.: Inference and parameter estimation in Gamma chains. Tech. Rep. CUED/F-INFENG/TR.596, University of Cambridge (2008)
16. Dikmen, O., Cemgil, A.T.: Gamma Markov random fields for audio source modeling. IEEE Trans. Audio, Speech, Language Process. **18**(3), 589–601 (2010)
17. Donnet, S., Rivoirard, V., Rousseau, J., Scricciolo, C.: Posterior concentration rates for empirical Bayes procedures with applications to Dirichlet process mixtures. Bernoulli **24**(1), 231–256 (2018)
18. Donoho, D.L., Johnstone, I.M.: Adapting to unknown smoothness via wavelet shrinkage. J. Amer. Statist. Assoc. **90**(432), 1200–1224 (1995)
19. Eberlein, E., Keller, U.: Hyperbolic distributions in finance. Bernoulli **1**(3), 281–299 (1995)
20. Efron, B.: Large-scale inference, *Institute of Mathematical Statistics (IMS) Monographs*, vol. 1. Cambridge University Press, Cambridge (2010). Empirical Bayes methods for estimation, testing, and prediction
21. Elerian, O., Chib, S., Shephard, N.: Likelihood inference for discretely observed nonlinear diffusions. Econometrica **69**(4), 959–993 (2001)
22. Fearnhead, P.: Exact and efficient Bayesian inference for multiple changepoint problems. Stat. Comput. **16**(2), 203–213 (2006)
23. Feller, W.: Diffusion processes in genetics. In: Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950, pp. 227–246. University of California Press, Berkeley and Los Angeles (1951)
24. Feller, W.: Two singular diffusion problems. Ann. of Math. (2) **54**, 173–182 (1951)
25. Fuchs, C.: Inference for diffusion processes. Springer, Heidelberg (2013). With applications in life sciences, With a foreword by Ludwig Fahrmeir
26. Gelfand, A.E., Smith, A.F.M.: Sampling-based approaches to calculating marginal densities. J. Amer. Stat. Assoc. **85**(410), 398–409 (1990)
27. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B.: Bayesian data analysis, 3 edn. Texts in Statistical Science Series. CRC Press, Boca Raton, FL (2014)
28. Gelman, A., Roberts, G.O., Gilks, W.R.: Efficient Metropolis jumping rules. In: J.M. Bernardo, J.O. Berger, A.P. Dawid, A.F.M. Smith (eds.) Bayesian statistics, vol. 5, pp. 599–607. Oxford University Press (1996)
29. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. **6**(6), 721–741 (1984)
30. Ghosal, S., van der Vaart, A.: Fundamentals of nonparametric Bayesian inference, *Cambridge Series in Statistical and Probabilistic Mathematics*, vol. 44. Cambridge University Press, Cambridge (2017)
31. Godsill, S.J., Cemgil, A.T., Févotte, C., Wolfe, P.J.: Bayesian computational methods for sparse audio and music processing. In: 15th European Signal Processing Conference (EURASIP), pp. 345–349 (2007)
32. Green, P.J.: Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika **82**(4), 711–732 (1995)

33. Green, P.J.: Trans-dimensional Markov chain Monte Carlo. In: Highly structured stochastic systems, *Oxford Statist. Sci. Ser.*, vol. 27, pp. 179–206. Oxford Univ. Press, Oxford (2003). With part A by Simon J. Godsill and part B by Juha Heikkinen

34. Gugushvili, S., Spreij, P.: Consistent non-parametric Bayesian estimation for a time-inhomogeneous Brownian motion. ESAIM Probab. Stat. **18**, 332–341 (2014)

35. Gugushvili, S., Spreij, P.: Nonparametric Bayesian drift estimation for multidimensional stochastic differential equations. Lith. Math. J. **54**(2), 127–141 (2014)

36. Gugushvili, S., Spreij, P.: Posterior contraction rate for non-parametric Bayesian estimation of the dispersion coefficient of a stochastic differential equation. ESAIM Probab. Stat. **20**, 143–153 (2016)

37. Gugushvili, S., van der Meulen, F., Schauer, M., Spreij, P.: Nonparametric Bayesian estimation of a Hölder continuous diffusion coefficient. ArXiv e-prints (2017)

38. Hastings, W.K.: Monte Carlo sampling methods using Markov chains and their applications. Biometrika **57**(1), 97–109 (1970)

39. Hindriks, R.: Empirical dynamics of neuronal rhythms: data-driven modeling of spontaneous magnetoencephalographic and local field potential recordings. Ph.D. thesis, Vrije Universiteit Amsterdam (2011)

40. Hsu, D.A.: Tests for variance shift at an unknown time point. J. R. Stat. Soc. C-App. **26**(3), 279–284 (1977)

41. Hsu, D.A.: Detecting shifts of parameter in gamma sequences with applications to stock price and air traffic flow analysis. J. Amer. Stat. Assoc. **74**(365), 31–40 (1979)

42. Iacus, S.M.: Simulation and inference for stochastic differential equations. Springer Series in Statistics. Springer, New York (2008). With **R** examples

43. Iacus, S.M.: sde: Simulation and Inference for Stochastic Differential Equations (2016). URL `https://CRAN.R-project.org/package=sde`. R package version 2.0.15

44. Jacquier, E., Polson, N.G., Rossi, P.E.: Bayesian analysis of stochastic volatility models. J. Bus. Econ. Stat. **12**(4), 371–389 (1994)

45. Jacquier, E., Polson, N.G., Rossi, P.E.: Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. J. Econometrics **122**(1), 185–212 (2004)

46. Karatzas, I., Shreve, S.E.: Brownian motion and stochastic calculus, *Graduate Texts in Mathematics*, vol. 113, 2 edn. Springer-Verlag, New York (1991)

47. Küchler, U., Neumann, K., Sørensen, M., Streller, A.: Stock returns and hyperbolic distributions. Math. Comput. Model. **29**(10), 1–15 (1999)

48. Martin, R., Ouyang, C., Domagni, F.: 'Purposely misspecified' posterior inference on the volatility of a jump diffusion process. Statist. Probab. Lett. **134**(Supplement C), 106–113 (2018)

49. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equation of state calculations by fast computing machines. J. Chem. Phys. **21**(6), 1087–1092 (1953)

50. van der Meulen, F., Schauer, M.: Bayesian estimation of discretely observed multidimensional diffusion processes using guided proposals. Electron. J. Stat. **11**(1), 2358–2396 (2017)

51. van der Meulen, F., Schauer, M., van Zanten, H.: Reversible jump MCMC for nonparametric drift estimation for diffusion processes. Comput. Statist. Data Anal. **71**, 615–632 (2014)

52. van der Meulen, F., van Zanten, H.: Consistent nonparametric Bayesian inference for discretely observed scalar diffusions. Bernoulli **19**(1), 44–63 (2013)

53. Müller, P., Mitra, R.: Bayesian nonparametric inference—why and how. Bayesian Anal. **8**(2), 269–302 (2013)

54. Musiela, M., Rutkowski, M.: Martingale methods in financial modelling, *Stochastic Modelling and Applied Probability*, vol. 36, 2 edn. Springer-Verlag, Berlin (2005)

55. Mykland, P.A.: A Gaussian calculus for inference from high frequency data. Ann. Finance **8**(2-3), 235–258 (2012)

56. Nickl, R., Söhl, J.: Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions. Ann. Statist. **45**(4), 1664–1693 (2017)

57. Papaspiliopoulos, O., Pokern, Y., Roberts, G.O., Stuart, A.M.: Nonparametric estimation of diffusions: a differential equations approach. Biometrika **99**(3), 511–531 (2012)

58. Petrone, S., Rousseau, J., Scricciolo, C.: Bayes and empirical Bayes: do they merge? Biometrika **101**(2), 285–302 (2014)
59. Pokern, Y., Stuart, A.M., van Zanten, J.H.: Posterior consistency via precision operators for Bayesian nonparametric drift estimation in SDEs. Stochastic Process. Appl. **123**(2), 603–628 (2013)
60. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2017)
61. Robbins, H.: An empirical Bayes approach to statistics. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. I, pp. 157–163. University of California Press, Berkeley and Los Angeles (1956)
62. Robbins, H.: The empirical Bayes approach to statistical decision problems. Ann. Math. Statist. **35**, 1–20 (1964)
63. Robert, C.P., Casella, G.: Monte Carlo statistical methods, 2 edn. Springer Texts in Statistics. Springer-Verlag, New York (2004)
64. Roberts, G.O., Stramer, O.: On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. Biometrika **88**(3), 603–621 (2001)
65. Ruttor, A., Batz, P., Opper, M.: Approximate Gaussian process inference for the drift function in stochastic differential equations. In: C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, K.Q. Weinberger (eds.) Advances in Neural Information Processing Systems 26, pp. 2040–2048. Curran Associates, Inc. (2013)
66. Sherlock, C., Fearnhead, P., Roberts, G.O.: The random walk Metropolis: linking theory and practice through a case study. Statist. Sci. **25**(2), 172–190 (2010)
67. Silverman, B.W.: Density estimation for statistics and data analysis. Monographs on Statistics and Applied Probability. Chapman & Hall, London (1986)
68. Soulier, P.: Nonparametric estimation of the diffusion coefficient of a diffusion process. Stochastic Anal. Appl. **16**(1), 185–200 (1998)
69. Tanner, M.A., Wong, W.H.: The calculation of posterior distributions by data augmentation. J. Amer. Statist. Assoc. **82**(398), 528–550 (1987). With discussion and with a reply by the authors
70. Tierney, L.: Markov chains for exploring posterior distributions. Ann. Statist. **22**(4), 1701–1762 (1994). With discussion and a rejoinder by the author
71. Virtanen, T., Cemgil, A.T., Godsill, S.: Bayesian extensions to non-negative matrix factorisation for audio signal modelling. In: 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1825–1828 (2008)
72. Wasserman, L.: All of nonparametric statistics. Springer Texts in Statistics. Springer, New York (2006)
73. Wilkinson, D.J.: Metropolis Hastings MCMC when the proposal and target have differing support. https://darrenjw.wordpress.com/2012/06/04/metropolis-hastings-mcmc-when-the-proposal-and-target-have-differing-support/ (2012). Online; accessed on 23 December 2017
74. Wong, E., Hajek, B.: Stochastic processes in engineering systems. Springer Texts in Electrical Engineering. Springer-Verlag, New York (1985)
75. Yang, T.Y., Kuo, L.: Bayesian binary segmentation procedure for a Poisson process with multiple changepoints. J. Comput. Graph. Statist. **10**(4), 772–785 (2001)
76. van Zanten, H.: Nonparametric Bayesian methods for one-dimensional diffusion models. Math. Biosci. **243**(2), 215–222 (2013)