UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Nonparametric Bayesian inference for stochastic processes with piecewise constant priors

Belomestny, Denis; van der Meulen, Frank; Spreij, Peter

DOI
10.1007/978-3-031-47417-0_28

Publication date
2024

Document Version
Final published version

Published in
2021-2022 MATRIX Annals

License
Article 25fa Dutch Copyright Act (https://www.openaccess.nl/en/policies/open-access-in-dutch-copyright-law-taverne-amendment)

Link to publication

Citation for published version (APA):
Belomestny, D., van der Meulen, F., & Spreij, P. (2024). Nonparametric Bayesian inference for stochastic processes with piecewise constant priors. In D. R. Wood, J. de Gier, & C. E. Praeger (Eds.), *2021-2022 MATRIX Annals* (Vol. 5, pp. 527-568). (Matrix Book Series; Vol. 5). Springer. https://doi.org/10.1007/978-3-031-47417-0_28

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.
UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:14 Oct 2026

Nonparametric Bayesian inference for stochastic processes with piecewise constant priors

Denis Belomestny, Frank van der Meulen, and Peter Spreij

Abstract We present a survey of some of our recent results on Bayesian nonparametric inference for a multitude of stochastic processes. The common feature is that the prior distribution in the cases considered is on suitable sets of piecewise constant or piecewise linear functions, that differ for the specific situations at hand. Posterior consistency and in most cases contraction rates for the estimators are presented. Numerical studies on simulated and real data accompany the theoretical results.

Key words: Diffusion coefficient; Dispersion coefficient; Gibbs sampler; (Inverse) Gamma Markov chain prior; Nonparametric Bayesian estimation; Stochastic differential equation; Volatility; Point process; Microstructure noise; Lévy density, Gamma-type subordinators.

2000 Mathematics Subject Classification. Primary: 62G20, Secondary: 62M05.

1 Introduction

This paper is a survey of our recent results, obtained with collaborators, in Bayesian nonparametric function estimation. Although the settings and corresponding mod-

Denis Belomestny
Faculty of Mathematics, Duisburg-Essen University, Thea-Leymann-Str. 9, D-45127 Essen, Germany and Faculty of Computer Sciences, HSE University, Moscow, Russian Federation. e-mail: denis.belomestny@uni-due.de

Frank van der Meulen
Department of Mathematics, Vrije Universiteit, De Boelelaan 1111, 1081HV Amsterdam, The Netherlands. e-mail: f.h.van.der.meulen@vu.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

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

D. R. Wood et al. (eds.), *2021-2022 MATRIX Annals*, MATRIX Book Series 5,
https://doi.org/10.1007/978-3-031-47417-0_28

els that we consider exhibit a large variety (volatility estimation in diffusion models, Poisson intensity estimation, estimation for gamma-driven stochastic differential equations, estimation in the presence of microstructure noise, and Lévy density estimation for Gamma-type subordinators), they all share the following common property. The function to be estimated, which can be a function of time or space, is approximated by piecewise constant functions (with one exception that deals with piecewise linear functions). This assumption is one of the possibilities to reduce the estimation of an infinite dimensional parameter to the estimation of only finitely many parameters, this being inevitable if one has to *compute* estimators.

We adopt the Bayesian approach by specifying a prior distribution supported on piecewise constant functions. The goal is then to compute the posterior distribution and show it has good properties, when considered from a frequentist point of view. In this work, the first goal is either accomplished by using conjugate priors or using Markov Chain Monte Carlo to obtain samples from the posterior. The second goal is achieved by showing posterior consistency and quantification of posterior contraction rates. A thorough introduction to these concepts is given in [13]. Informally, posterior consistency means that the posterior mass of a fixed ε-ball around the "true" function tends to 1 as the sample size increases. The contraction rates quantifies the rate at which we can shrink the radius of these balls with increasing sample size while still asymptotically having posterior mass 1.

It turns out that this procedure is feasible in all cases that we consider. It is intuitively clear that rougher (continuous) functions are more difficult to estimate than smoother ones. The degree of smoothness is measured by the Hölder exponent in our approach. We assume in all cases that the functions to be estimated are Hölder continuous. We obtain contraction rates that depend on the Hölder exponent, and the smoother the function, the better the estimation procedure's contraction rate. As we will demonstrate, despite the entirely different settings, the contraction rates are identical in all cases.

For theoretical and numerical implementations, one has to specify priors on the approximating piecewise continuous functions. We will assume a fixed set of bins, reflecting the resolution at which we aim to learn the unknown function. In a first approach, we will a priori take the (finitely many) values on the bins as *independent* random variables with distributions from an appropriate chosen class (mostly gamma or inverse gamma distributions), such that the prior is (partially) conjugate. These specifications lead to the theoretical results on the posterior distributions mentioned above.

It also appeared that better numerical results, clearly visible in the figures we provide and less sensitive to the chosen set of bins, can be obtained by relaxing the independence assumption on the prior distribution. We accomplish this observation by using appropriate Markov chain priors on the levels of the piecewise constant functions. A principal motivation is that such a prior introduces dependence between the adjacent levels of the piecewise constant approximating functions, which is a rather natural property if one realizes that values of continuous functions tend to be close to each other on adjacent intervals. The theoretical results on the contraction rates we show in several cases coincide with the minimax convergence rates

in frequentist estimation. From that asymptotic point of view, the estimators with the Markov chain prior cannot be better than the ones with the independent priors. Nevertheless, in practical situations, they turn out to be more accurate.

A great variety of numerical examples in all settings are presented. They include results from simulated examples to illustrate the theoretical properties and from real data to derive some conclusions based on our methods.

The paper is organized as follows. Section 2 details the results on volatility estimation for stochastic differential equations driven by a Brownian motion, Section 3 is devoted to estimation for partial observations of diffusion processes in presence of microstructure noise, Section 4 treats intensity estimation for Poisson processes, Section 5 gives results on estimation of the volatility function in gamma-driven stochastic differential equations, Section 6 is on estimation of the Lévy density for Gamma-type subordinators. Finally Section 7 summarizes and concludes.

A remark on the notation throughout the paper: for two sequences (a_n) and (b_n) of real numbers we write $a_n \asymp b_n$ if there are constants c and C in $(0,\infty)$ such that eventually $cb_n \le a_n \le Cb_n$.

2 Volatility estimation

In this section, we consider the nonparametric Bayesian estimation of a volatility function, also synonymously referred to as the dispersion coefficient, that arises in diffusion models. This function may be a smooth function of the time parameter, but it may also happen that it is a realization of another diffusion process, like a Brownian motion. Hence the function to be estimated may have various orders of regularity, which we characterize by their degree of Hölder continuity. This section is based on [15], [17], which also contain many references to related literature.

2.1 Problem formulation

Stochastic differential equations (SDEs) driven by a Brownian motion have been widely used as models in numerous applications ranging from physics to engineering and to finance. Their solutions, weak or strong, are diffusion processes. In this section we consider a one-dimensional diffusion process X satisfying the stochastic differential equation (SDE)

$$\mathrm{d}X_t = a(t,X_t)\,\mathrm{d}t + s(t)\,\mathrm{d}W_t, \quad X_0 = x_0, \quad t \in [0,1]. \tag{1}$$

Here W is a standard Wiener process, and a and s are the drift function and volatility function, respectively. Note that s is a function of time only. It is further assumed that b and s are such that (1) admits a weak solution that is unique in law. We assume that observations

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

from the solution X to (1) are available, where $t_{i,n} = \frac{i}{n}$, $i = 0, \ldots, n$. In the statistical set-up, we denote the true drift and diffusion coefficient by a_0 and s_0, Our aim is to estimate s_0 nonparametrically within the Bayesian setup. We note that nonparametric Bayesian estimation of volatility is very different from the drift coefficient estimation: the latter fundamentally relies on the equivalence of laws of continuously observed diffusion processes that have the same diffusion coefficient. This is not applicable when the diffusion coefficient itself is unknown and is a parameter to be estimated when a continuous record of observations is available. In the current set-up on the contrary, with *discrete time* observations, there is equivalence of laws and it is therefore possible to derive a likelihood, at least theoretically, a fundamental tool for Bayesian inference. An implementable version of such a likelihood is obtained by *intentionally* misspecifying the drift coefficient a by taking it equal to zero, and employing a (conjugate) histogram-type prior on the diffusion coefficient, that has piecewise constant realisations on bins forming a partition of $[0,1]$. Due to this, our nonparametric Bayesian method to estimate the dispersion coefficient s_0 in (1) is easily implemented. Theoretical asymptotic results for estimating s will be obtained in the case when $n \to \infty$ (high frequency observations), keeping the observation interval $[0,1]$ fixed. For simplicity, we take $x_0 = 0$. Because of the fixed interval $[0,1]$, consistent estimation of the drift is impossible, but the high frequency observations make it possible to ignore the drift if the aim is to (consistently) estimate the volatility. We will make this precise in Proposition 2.3.

2.2 Likelihood, prior and posterior

Under the assumption that the drift in (1) is zero, the (pseudo-)likelihood is that of a Gaussian sample and is given by

$$L_n(s) = \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi \int_{t_{i-1,n}}^{t_{i,n}} s^2(u)\,\mathrm{d}u}} \phi_0 \left(\frac{Y_{i,n}}{\sqrt{\int_{t_{i-1,n}}^{t_{i,n}} s^2(u)\,\mathrm{d}u}} \right) \right\}, \tag{2}$$

where $Y_{i,n} = X_{t_{i,n}} - X_{t_{i-1,n}}$ and $\phi_0(u) = \exp(-u^2/2)$. With Π_n denoting a prior on the dispersion coefficient, provided all the involved quantities are suitably measurable, Bayes' theorem gives that the posterior probability of any measurable subset $S \subset \mathscr{S}$ of dispersion coefficients is given by

$$\Pi_n(S \mid X_{t_{0,n}} \ldots, X_{t_{n,n}}) = \frac{\int_S L_n(s)\Pi_n(\mathrm{d}s)}{\int_{\mathscr{S}} L_n(s)\Pi_n(\mathrm{d}s)},$$

where $\mathscr{S}$ denotes a space on which the prior Π_n is defined.

It follows from (2) that the likelihood depends on the parameter of interest only through the integrals $\int_{t_{i-1,n}}^{t_{i,n}} s^2(u)\,\mathrm{d}u$. Consequently, it appears natural to a priori

model the diffusion coefficient as piecewise constant on intervals $[t_{i-1,n}, t_{i,n}]$. However, some smoothing should also be performed, and this can be achieved by aggregated several neighbouring intervals $[t_{i-1,n}, t_{i,n}]$. Thus, to construct the prior, we proceed as follows: Let m be an integer smaller than n. Then we can uniquely write $n = mN + r$ with $0 \leq r < m$, and in fact $N = \lfloor \frac{n}{m} \rfloor$. Both m and N will depend on n (and we also write m_n and N_n to emphasize this when appropriate). With this assumption we have bins $B_k = [t_{m(k-1),n}, t_{mk,n})$, $k = 1, \ldots, N-1$ and $B_N = [t_{m(N-1),n}, 1]$.

Let $\mathscr{S}_n$ be the set of functions s that can be represented as $s = \sum_{k=1}^{N_n} \xi_k \mathbf{1}_{B_k}$. The prior Π_n on the dispersion coefficient $s \in \mathscr{S}_n$ is defined by putting a prior on the coefficients ξ_k's. Since

$$s^2 = \sum_{k=1}^{N_n} \xi_k^2 \mathbf{1}_{B_k} = \sum_{k=1}^{N_n} \theta_k \mathbf{1}_{B_k}, \tag{3}$$

where we have put $\theta_k = \xi_k^2$, equivalently one can place the prior on the coefficients θ_k's of the diffusion coefficient s^2. We call the prior Π_n a histogram-type prior.

As a result of choosing Π_n such that the θ_i are independent having an inverse gamma distribution, we can explicitly compute the conjugate posterior of the θ_i, again of inverse gamma type.

Lemma 2.1. *Assume $\theta_1, \ldots, \theta_N$ are independent with the inverse gamma $\mathrm{IG}(\alpha, \beta)$ distribution. Then $\theta_1, \ldots, \theta_N$ are a posteriori independent and, for $k = 1, \ldots, N-1$,*

$$\theta_k \mid \mathscr{X}_n \sim \mathrm{IG}\left(\alpha + m/2, \beta + nZ_k/2\right).$$

with

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

whereas

$$\theta_N \mid \mathscr{X}_n \sim \mathrm{IG}\left(\alpha + (m+r)/2, \beta + nZ_N/2\right),$$

with

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

2.3 Asymptotics

We let $\mathbb{P}_{a_0,s_0}$ denote the law of the path $(X_t : t \in [0,1])$ from (1) under the true parameter values (a_0, s_0). In particular the notation $\mathbb{P}_{0,s_0}$ is used for such a law when the drift coefficient is equal to zero. For expectation we use the same subscripts. We denote the prior distribution on the dispersion coefficient by Π_n (with n the number of observations) and write the posterior as $\Pi_n(\cdot \mid \mathscr{X}_n)$. By $\|\cdot\|_2$ we denote the L_2-norm with respect to the Lebesgue measure on the Borel sets of $[0,1]$.

The standing assumption on the volatility in this section is the following.

Assumption 2.2 Assume that the true volatility function s_0 is Hölder continuous on $[0,1]$ with Hölder constant L and Hölder exponent $\lambda \in (0,1]$, $|s_0(u) - s_0(v)| \leq L|u-v|^{\lambda}$ for all $u, v \in [0,1]$, and is bounded away from zero.

The first result formalises that under our asymptotic regime, we can safely ignore the drift by, erroneously in general, putting it equal to zero.

Proposition 2.3. *Let, for $\varepsilon > 0$,*

$$U_{s_0,\varepsilon} = \{s \in \mathscr{S}_n \colon \|s - s_0\|_2 < \varepsilon\}$$

be the L_2-neighbourhood of s_0 of radius ε. Let Assumption 2.2 hold and assume that there exists a sequence $\varepsilon_n \to 0$

$$\mathbb{E}_{0,s_0}[\Pi_n(U^c_{s_0,\varepsilon_n} \mid \mathscr{X}_n)] \to 0$$

as $n \to \infty$. Then also

$$\mathbb{E}_{a_0,s_0}[\Pi_n(U^c_{s_0,\varepsilon_n} \mid \mathscr{X}_n)] \to 0.$$

As we consider the asymptotics $n \to \infty$, we take the number of bins N to depend on the sample size n, and indicate this in our notation by writing N_n. The next result gives the posterior contraction rate $n^{-\lambda/(2\lambda+1)}$ for the L_2-norm. Establishing posterior contraction in the L_2-metric is rather natural, as $\int_0^1 s_0^2(t)\mathrm{d}t$ is the quadratic variation of the process $(X_t : t \in [0,1])$ over the interval $[0,1]$.

Theorem 2.4. *Assume that Assumption 2.2 holds. Assume $N_n \asymp n^{1/(2\lambda+1)}$. If we let $\varepsilon_n \asymp n^{-\lambda/(2\lambda+1)}$, then for any sequence h_n tending to infinity (as $n \to \infty$) we have*

$$\mathbb{E}_{a_0,s_0}\left[\Pi_n(\|s^2 - s_0^2\|_2 \geq h_n\varepsilon_n \mid \mathscr{X}_n)\right] \to 0$$

as $n \to \infty$.

A comparison with the frequentist minimax convergence rate in [20] shows that the posterior for the diffusion coefficient contracts at the optimal rate in the L_2-metric. With histogram-type priors considered in this work no further improvement in the posterior contraction rate, when $\lambda = 1$, is possible beyond $n^{-1/3}$, even if the function s_0 is smoother than a Lipschitz function. An intuitive reason for this is that realisations of our histogram-type priors are too rough for this; cf. p. 629 in [37].

2.4 Numerical results

In this section we illustrate the theoretical results of Section 2.3 with a number of numerical examples. For an implementation of our Bayesian analysis a judicious choice of the number of bins N is crucial. For obtaining such a choice there are various methodologies. For instance one can use the Deviance Information Criterion (DIC) of [38] to select N, or choose N as the maximizer of the marginal likelihood

$n \mapsto \int_{\mathscr{S}} L_n(s)\Pi_n(\mathrm{d}s)$. We refer to [15] for details. We use simulations of diffusion processes with known drift and diffusion coefficient to gain insight into the numerical performance of our method. We are particularly interested in the practical consequence of using a pseudo-likelihood ignoring the drift.

First we simulate realisations from the model for different dispersion and drift coefficients. Given subsamples of those realisations sampled at different rates we compute the posterior distribution of the dispersion function using the piecewise constant prior (histogram-type prior) of Lemma 2.1 with varying numbers of bins. As an illustration, plots of marginal posterior bands are compared with the true dispersion function. The marginal posterior bands are obtained by computing $1-\alpha$ central posterior intervals (see [11], p. 33) separately for the coefficients θ_k's using Lemma 2.1. This way of computing the posterior intervals will also be used in later sections. By Proposition 2.3, assuming that there is no drift still leads to consistent Bayesian estimation of the dispersion coefficient, even if the data are from a diffusion process with nonzero drift. This is illustrated by our simulation results of this section.

We simulated sample paths of the diffusion $(X_t : t \in [0,1])$ where the true dispersion coefficient is given by one of

$$s_1(t) = 3/2 + \sin(2(4t-2)) + 2\exp(-16(4t-2)^2),$$
$$s_2(t) = B_t(\omega_0) + 1,$$

and the drift is given by one of

$$a_1(x) = 0,$$
$$a_2(x) = -10x + 20.$$

Note the change of notation in these examples, true dispersion and drift are not denoted s_0 and a_0 here, but differently to distinguish these special choices from the general case. The function s_1 is a benchmark function used in [10] in the context of nonparametric regression, up to a vertical shift to ensure positivity. To define $s_2(t)$, we took a fixed realisation of a Brownian motion B, indicated by ω_0, starting at 1 with $B_t(\omega_0) > -1$ for $t \in [0,1]$. The function s_1 is Lipschitz continuous, while s_2 is Hölder continuous with exponent essentially $\frac{1}{2}$. As the prior on the coefficients on the individual bins we used $\mathrm{IG}(0.1, 0.1)$ distributions and the coefficients were assumed to be independent.

Figures 1 and 3 show marginal 98 % posterior bands for different combinations of bin number and observation regime for both dispersion coefficients when the drift is zero. Figure 2 shows the marginal posterior bands for s_1 which are obtained if an affine drift term $a_2(x) = -10x + 20$ is present, but neglected in the estimation procedure. Comparison with Figure 1 shows that presence of a strong nonzero drift hardly affects the obtained credible bands. Note that credible bands successfully recover the overall shape of the functions s_i, although the recovery is not too refined. The recovery is somewhat less accurate for function s_2, see Figure 3, than for func-

tion s_1, see Figure 1. This is in perfect agreement with our theoretical results from Theorem 2.4, that give a slower posterior contraction rate for less smooth functions.

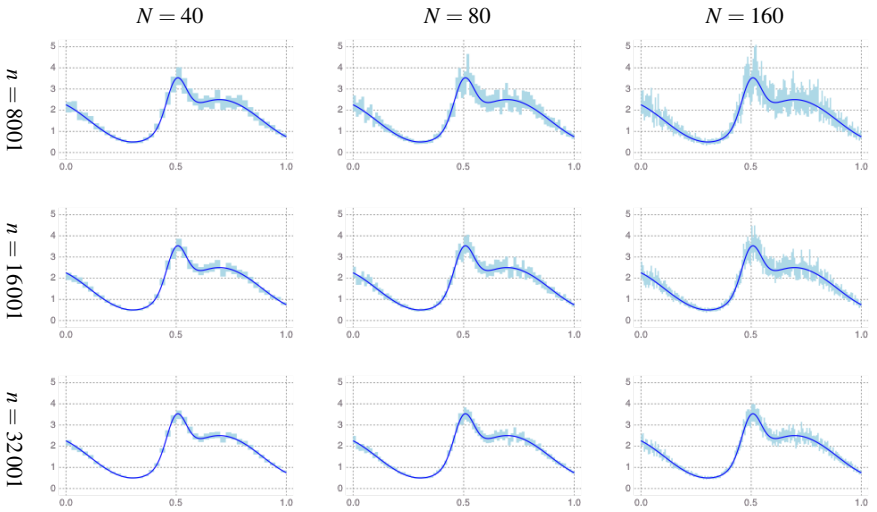

Fig. 1 Estimation results for s_1 with varying number of observations and bins, no drift. The solid curve is the true dispersion coefficient while the light blue areas are 98 % marginal posterior bands. n is the number of observations, N is the number of bins.

2.5 Inverse Gamma Chain Markov prior

In this section we postulate another prior on the values of the bins. Rather than assuming independent inverse Gamma distributions as in Section 2.2, we take an *inverse Gamma Markov chain (IGMC)* prior. Posterior inference is still straightforward to implement via Gibbs sampling, as the full conditional distributions are available explicitly and turn out to be inverse Gamma again. We will compare the performance of under both priors. The results of this section are based on [17].

We still consider the diffusion process that solves the SDE (1), with the observations $\mathscr{X}_n = \{X_{t_{0,n}}, \ldots, X_{t_{n,n}}\}$. We also keep on taking the drift equal to zero in our statistical analysis. A main reason to depart from the independent prior setting is that this has only limited possibility 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 $[0,1]$.

The alternative, fundamentally different, approach, we undertake here, is inspired by ideas in [4] in the context of audio signal modelling, see also [7] and [30], different from the SDE setting that we consider. Instead of using a prior on the (squared)

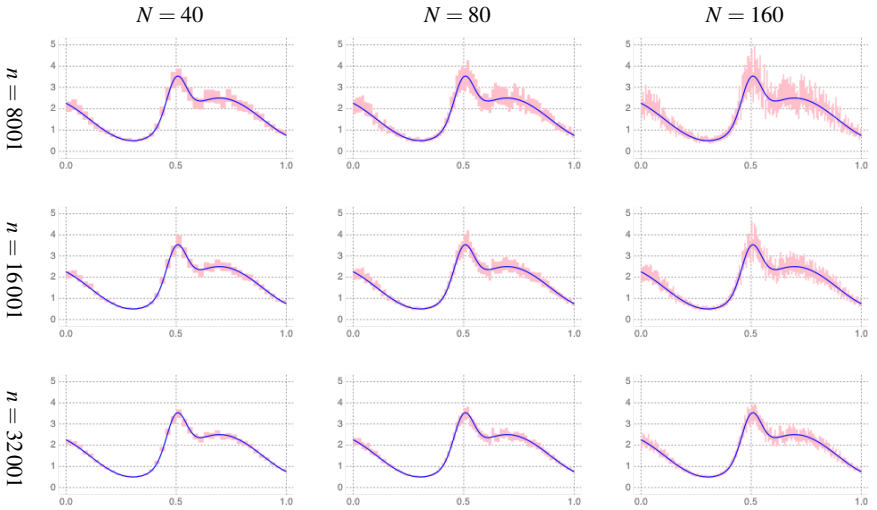

Fig. 2 Estimation for s_1 with varying number of observations and bins and drift $a_2(x) = -10x + 20$. The solid curve is the true dispersion coefficient while the light red areas are 98 % marginal posterior bands. n is the number of observations, N is the number of bins.

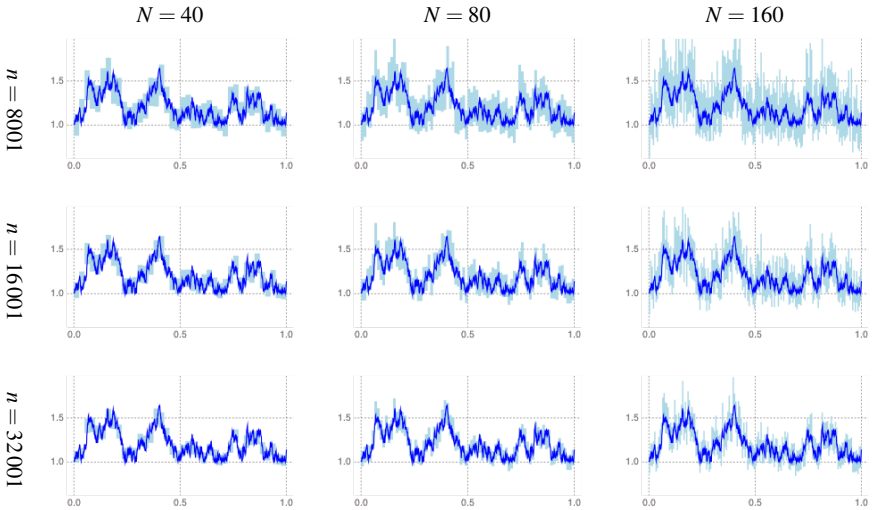

Fig. 3 Estimation results for s_2 with varying number of observations and bins, no drift. The solid curve is the true dispersion coefficient while the light blue areas are 98 % marginal posterior bands. n is the number of observations, N is the number of bins.

volatility that has piecewise constant realisations on $[0,1]$ with independent coefficients θ_k's, as in Section 2.2, 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. Specifically, we proceed as follows, first specifying the piecewise constant prior on the levels on the bins, now being a Markov chain. The bins B_k are the same as before. We again model s as piecewise constant on bins B_k, thus $s = \sum_{k=1}^N \xi_k \mathbf{1}_{B_k}$, and with $\theta_k = \xi_k^2$ it follows that $s^2 = \sum_{k=1}^N \theta_k \mathbf{1}_{B_k}$.

Introduce auxiliary variables ζ_k, $k = 2,\ldots,N$, and define a Markov chain using the time ordering $\theta_1, \zeta_2, \theta_2, \ldots, \zeta_k, \theta_k, \ldots, \zeta_N, \theta_N$. Transition distributions of this chain are defined as follows: fix hyperparameters α_1, α_ζ and α, and set

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

The prior is referred to as an inverse Gamma Markov chain (IGMC) prior, which reflects the fact that the distributions in (4) are inverse Gamma. In fact, it can even be verified by direct computations employing (4) that the full conditional distributions of θ_k's and ζ_k's are inverse gamma too, see [17] for details, which facilitates Gibbs sampling. As observed in [4], 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. The definition of the IGMC prior in the present work, that employs the latent variables ζ_k's, takes care of both these important points. For an additional discussion see [4].

In the practical examples presented below, both the number of bins, N, and the parametes $\alpha, \alpha_1, \alpha_\zeta$ are considered as hyperparameters. Assigning values to them can be done in multiple ways, see [17], but it has been observed in our simulation examples that inferential conclusions with the IGMC prior are quite robust with respect to the choice of N. This is because through the hyperparameters α and α_ζ, the IGMC prior has an additional layer for controlling the amount of applied smoothness; when α and α_ζ are equipped with a prior (as we do), they can in fact be learned from the data.

The test function to be estimated is the blocks function from [8] visualized in Figure 4. With a vertical shift for positivity, this is defined as follows:

$$s(t) = 10 + 3.655606 \times \sum_{j=1}^{11} h_j K(t - t_j), \quad t \in [0,1], \quad (5)$$

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

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

Our main goal here was to compare the performance of the IGMC prior-based approach to the IIG prior-based one from Section 2.2. To complete the SDE specification, the true drift coefficient was chosen to be a rather strong linear drift $a_2(x) = -10x + 20$.

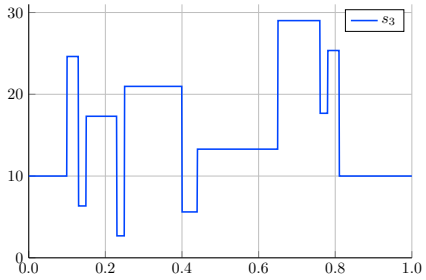

Fig. 4 The blocks function s from Equation (5).

The top left and top right panels of Figure 5 give estimation results obtained with the IIG prior-based approach from Section 2.2. The number of bins was $N = 160$ or $N = 320$, and in both these cases we used diffuse independent $\mathrm{IG}(0.1, 0.1)$ priors on the coefficients of the (squared) volatility function (see [15] 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 5, where we assumed $\alpha_1 = 0.1$ and $\alpha = \alpha_\zeta \sim \mathrm{IG}(0.3, 0.3)$. The following conclusions emerge from Figure 5:

- 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.
- The method based on the IIG prior is sensitive to the bin number selection: compare the top left panel of Figure 5 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 5 look similar to each other.

We have shown more pronounced results when using the IGMC prior than by using the IIG prior, although the latter already shows optimal asymptotic properties, as reported in Section 2.3. We will see similar behaviour in Section 4.6.

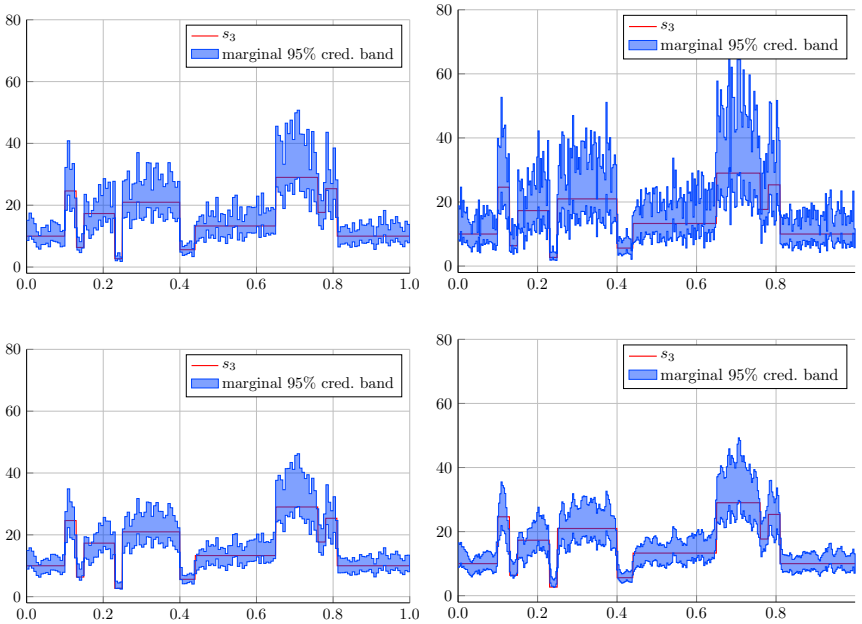

Fig. 5 Volatility function s with superimposed 95% marginal credible band for the IIG prior IG$(0.1, 0.1)$, using $N = 160$ (top left panel) and $N = 320$ bins (top right panel). Volatility function s from (5) with superimposed 95% marginal credible band for the IGMC prior, using $N = 160$ (bottom left panel) and $N = 320$ bins (bottom right panel); in both cases, $\alpha_1 = 0.1$ and $\alpha = \alpha_\zeta \sim \text{IG}(0.3, 0.3)$.

2.6 Closing remarks

We have studied nonparametric Bayesian estimation of a volatility function, and presented posterior contraction results when the prior was of independent inverse gamma type. We have given numerical results and have also illustrated that the performance of the statistical procedure had increased when switching from independent priors on the coefficients to a Markov chain prior. The theoretical results can also be applied to real data, see [15] for an example with exchange rate data and Section 3 for a follow up with partial observations.

3 Volatility estimation in presence of microstructure noise

The goal of this section is nonparametric Bayesian estimation of the volatility in presence of market microstructure noise. The exposition is based on [16], which in turn uses some ideas from [17], whose findings have also been used in Section 2.5.

3.1 Problem formulation

As in Section 2, let X solve the one-dimensional stochastic differential equation (SDE) (1). Assume (not necessarily uniformly spaced) observation times $\{t_i : 1 \leq i \leq n\}$ and suppose our observations are $\mathscr{Y}_n = \{Y_1, \ldots, Y_n\}$. Here

$$Y_i = X_{t_i} + V_i, \quad 0 < t_1 < \cdots < t_n = 1, \tag{6}$$

and $\{V_i\}$ is a sequence of independent and identically distributed random variables, independent of the Wiener process W driving (1). The measurement errors $\{V_i\}$ are referred to as microstructure noise. They model such features of financial time series as their discreteness or approximations due to market friction. For widely spaced data in time, microstructure noise can be ignored without great consequences for inferring the volatility function. However, for densely spaced data in time (the so-called high-frequency data regime), this ceases to be true and even small amount of the microstructure noise adversely affects the volatility estimates.

As referenced in [16], there exists a large body of literature on nonparametric volatility estimation under microstructure noise. The existing work mainly focusses on estimation of the integrated volatility and is frequentist in its nature. Here we take a Bayesian approach instead. General advantages of the latter have been sufficiently highlighted in the preceding sections.

The key steps of our approach are as follows: we first reduce our model to the Gaussian linear state space model, which is computationally convenient. Next we specify an inverse Gamma Markov chain prior on the volatility function s. This allows us to employ for posterior inference the Gibbs sampler. The latter incorporates a classic Forward Filtering Backward Simulation (FFBS) step.

After detailing the algorithm, we demonstrate the performance of our method on simulated and real data examples, and also provide a theoretical result on the limiting behaviour of the Gamma Markov chain prior.

3.2 Linear state space model

We rigorously derive our Bayesian algorithm in the case when the "true", data-generating volatility s is a deterministic function of time t. Next, if the "true" s is in fact a stochastic process, we simply apply the algorithm without further changes, as if s were deterministic. The intuition is as follows: first the stochastic volatility is generated, and given a realisation of it, the observations are generated by an independent Brownian motion W.

Denote $t_0 = 0$. From equation (1),

$$X_{t_i} = X_{t_{i-1}} + \int_{t_{i-1}}^{t_i} a(t, X_t)\mathrm{d}t + \int_{t_{i-1}}^{t_i} s(t)\mathrm{d}W_t. \tag{7}$$

Due to the properties of the Wiener process paths, over short time intervals $[t_{i-1},t_i]$ the term $\int_{t_{i-1}}^{t_i} s(t)\mathrm{d}W_t$ in (7) dominates the term $\int_{t_{i-1}}^{t_i} a(t,X_t)\mathrm{d}t$. Indeed, the latter scales as Δt_i, whereas the former as $\sqrt{\Delta t_i}$. Following [15], as in Section 2, we act as if the process X had a zero drift, $a \equiv 0$.

We thus *assume* $X_{t_i} = X_{t_{i-1}} + U_i$, where $U_i = \int_{t_{i-1}}^{t_i} s(t)\mathrm{d}W_t$. This implies that

$$U_i \sim N(0,w_i) \quad \text{with} \quad w_i = \int_{t_{i-1}}^{t_i} s^2(t)\mathrm{d}t. \tag{8}$$

Furthermore, $\{U_i\}$ is a sequence of independent random variables, for so are the increments of a Wiener process over disjoint time intervals. To lighten the notation, in the sequel we will write $x_i = X_{t_i}$, $y_i = Y_i$, $u_i = U_i$, $v_i = V_i$. Our model reduces to the *linear state space model*

$$\begin{aligned} x_i &= x_{i-1} + u_i, \\ y_i &= x_i + v_i, \end{aligned} \tag{9}$$

where $i = 1,\ldots,n$. In the time series literature, the first equation in (9) is called the *state equation*, whereas the second equation is referred to as the *observation equation*. We assume that $\{v_i\}$ is a sequence of independent $N(0,\eta_v)$ distributed random variables, also independent of the Wiener process W in (1). As a result, $\{v_i\}$ is independent of $\{u_i\}$. We equip the *initial state* x_0 with the $N(\mu_0,C_0)$ prior distribution. Then (9) turns into a *Gaussian* linear state space model. The latter is most convenient computationally, as one does not have to deal with an *intractable likelihood*. The latter constitutes the main *computational bottleneck* for Bayesian inference in SDE models.

3.3 Prior

For the measurement error variance η_v, we assume a priori an inverse Gamma distribution, $\eta_v \sim \mathrm{IG}(\alpha_v,\beta_v)$. This choice is dictated by conjugacy considerations. In general, the inverse gamma prior on the variance parameter in Gaussian models is common practice in Bayesian statistics.

The construction of the prior for s is more involved. Fix an integer $m < n$. Then we can write $n = mN + r$ with $0 \le r < m$, where $N = \lfloor n/m \rfloor$. Define bins $B_k = [t_{m(k-1)}, t_{mk})$, $k = 1,\ldots,N-1$, and $B_N = [t_{m(N-1)}, 1]$. Assume a priori

$$s = \sum_{k=1}^{N} \xi_k \mathbf{1}_{B_k}, \tag{10}$$

where N (the number of bins) is a hyperparameter. Then $s^2 = \sum_{k=1}^{N} \theta_k \mathbf{1}_{B_k}$, where $\theta_k = \xi_k^2$. The prior specification for s is completed by assigning a prior distribution to the coefficients $\theta_{1:N}$ as in Section 2.5, which we repeat here. We introduce auxil-

iary variables $\zeta_{2:N}$, and suppose the sequence $\theta_1, \zeta_2, \theta_2, \ldots, \zeta_k, \theta_k, \ldots, \zeta_N, \theta_N$ forms a Markov chain (in this order of variables). The transition distributions of the chain are defined by

$$\theta_1 \sim \mathrm{IG}(\alpha_1, \beta_1), \quad \zeta_{k+1}|\theta_k \sim \mathrm{IG}(\alpha, \alpha\theta_k^{-1}), \quad \theta_{k+1}|\zeta_{k+1} \sim \mathrm{IG}(\alpha, \alpha\zeta_{k+1}^{-1}), \quad (11)$$

where $\alpha_1, \beta_1, \alpha$ are hyperparameters. We refer to this chain as an inverse Gamma Markov chain, see [4]. The variables $\theta_1, \ldots, \theta_N$ are positively correlated, which imposes *smoothing* across different bins B_k: under the prior, the small or large values of θ's on adjacent bins tend to go together. At the same time, we attain partial conjugacy in the Gibbs sampler that we derive below. This leads to simple and tractable MCMC inference.

An uninformative choice $\alpha_1, \beta_1 \to 0$ performs well in practice. The parameter α plays the role of the smoothing parameter. We equip it with a log-normal prior $\log \alpha \sim N(a, b)$. The hyperparameters $a \in \mathbb{R}, b > 0$ are chosen so as to render the hyperprior on α diffuse. The interpretation of hyperparameter N (or equivalently m) is somewhat different than merely a smoothing parameter. It can also be considered as the resolution at which one wants to infer the volatility function. Under finite amounts of data, this resolution cannot be made arbitrarily fine. On the other hand, inference with the IGMC prior is robust with respect to a wide range of values of N due to the fact that the corresponding Bayesian procedure has an additional *regularisation parameter* α that is estimated from the data. See also the earlier Section 2.5 and Section 4 for related phenomena.

3.4 Gibbs sampler for sampling from the posterior

An expression for the posterior of s can be written down in closed form, but does not scale up computationally. The problem is circumvented by following a data augmentation approach: we treat $x_{0:n}$ as *missing data*, whereas the $y_{1:n}$ are the observed data; cf. [39]. An expression for the joint density of all the quantities involved in the model is easily obtained as

$$\begin{aligned} p(y_{1:n}, x_{0:n}, \theta_{1:N}, \zeta_{2:N}, \alpha, \eta_v) = \\ \left(\prod_{k=1}^{n} p(y_k \mid x_k, \eta_v) \right) p(x_{0:n} \mid \theta_{1:N}) \\ \times p(\theta_1) \prod_{k=1}^{N-1} \left[p(\zeta_{k+1} \mid \theta_k, \alpha) p(\theta_{k+1} \mid \zeta_{k+1}, \alpha) \right] p(\alpha) p(\eta_v). \end{aligned}$$

Except for $p(x_{0:n} \mid \theta_{1:N})$, all the densities have been given in the previous subsection. To obtain an expression for the latter, define (with $\Delta_i \equiv \Delta t_i$)

$$Z_k = \sum_{i=(k-1)m+1}^{km} \frac{(x_i - x_{i-1})^2}{\Delta_i}, \quad k = 1, \ldots, N-1,$$
$$Z_N = \sum_{i=(N-1)m+1}^{n} \frac{(x_i - x_{i-1})^2}{\Delta_i}.$$

Next set $m_k = m$ for $k = 1, \ldots, N-1$, and let $m_N = m + r$. It follows that

$$p(x_{0:N} \mid \theta_{1:N}) \propto e^{(x_0-\mu_0)^2/(2C_0)} \prod_{k=1}^{N} \theta_k^{-m_k/2} \exp\left(-\frac{Z_k}{2\theta_k}\right).$$

Now we can use the Gibbs sampler to sample from the joint conditional distribution of $(x_{0:n}, \theta_{1:N}, \zeta_{2:N}, \eta_v, \alpha)$ given observations $y_{1:n}$. The full conditionals of $\theta_{1:N}$, $\zeta_{2:N}$, η_v required in the Gibbs sampler are easy to derive and are of the inverse Gamma type. Furthermore, the parameter α is updated via a Metropolis-Hastings step; see [16] for all the details. This is quite similar conceptually to Section 4.5 below. On the other hand, for updating $x_{0:N}$, conditional on all other quantities in the model, we use the classical Forward Filtering Backward Simulation (FFBS) algorithm for Gaussian state space models (cf. Section 4.4.3 in [31]). The Gibbs sampler that results from these steps is summarized in Algorithm 1. For details, see [16].

Algorithm 1: Gibbs sampler for volatility learning

Data: Observations $y_{1:n}$
Hyperparameters α_1, β_1, α_v, β_v, a, b, N;
Result: Posterior samples $\theta^i_{1:N} : i = 1, \ldots, M$
Initialization $\theta^0_{1:N}$, $\zeta^0_{1:N}$, η^0_v, α^0;
while $i \leq M$ **do**
  sample $x^i_{0:n}$ via FFBS;
  sample $\theta^i_{1:N}$ from the inverse Gamma full conditionals;
  sample $\zeta^i_{2:N}$ from the inverse Gamma full conditionals;
  sample η^i_v from the inverse Gamma full conditional;
  sample α^i via a Metropolis-Hastings step;
  set $i = i + 1$.
end

3.5 Asymptotics

In this section we provide some results on the limit behaviour of the Gamma Markov Chain prior. They shed additional light on the prior and are expected to be of use when establishing frequentist asymptotic properties of our nonparametric Bayesian volatility estimation procedure.

We start with results on the prior conditional mean and variance of the θ_k. These results will be exploited to find an asymptotic regime for the θ_k when the number of bins tends to infinity. To simplify the exposition, in this section we depart from the setting with α as a random hyperparameter, but instead we assume it to be a deterministic sequence. The proof of the following proposition is in [16].

Proposition 3.1. *If $\alpha > 2$ is a fixed parameter then the IGMC prior satisfies*

$$\mathbb{E}_k[\theta_{k+1} - \theta_k] = \frac{1}{\alpha - 1}\theta_k, \tag{12}$$

$$\mathbb{V}_k(\theta_{k+1} - \theta_k) = \frac{\alpha(2\alpha - 1)}{(\alpha - 1)^2(\alpha - 2)}\theta_k^2, \tag{13}$$

where $\mathbb{E}_k$ and $\mathbb{V}_k$ respectively denote expectation and variance, conditional on θ_k. Hence, the conditional mean squared error $\mathbb{E}_k(\theta_{k+1} - \theta_k)^2$ equals $\frac{2(\alpha+1)\theta_k^2}{(\alpha-1)(\alpha-2)}$.

As seen from Proposition 3.1, the conditional mean squared error in Proposition 3.1 decreases in α. This, in turn, illustrates the regularising property of this parameter.

We are interested in the behaviour of the prior distribution on the θ_k for large values of α. We will suitably scale α with the number of bins N to obtain a limit result by applying Donsker's theorem (cf. Corollary VII.3.11 in [23]). Take $\alpha = \gamma N$, where γ is a positive scaling factor. We are thus interested in the law of θ_k for $N \to \infty$. This entails simultaneously increasing the number of bins, as well as the dependence of the values on the bins. Then, as shown in [16], with θ_1 fixed, the process $t \mapsto s^2(t)$, with $s^2(t) = \sum_{k=1}^{N} \theta_k \mathbf{1}_{B_k}(t)$, converges weakly to the continuous time process $t \mapsto \theta_1 Z_t$, where

$$\log Z_t = \sqrt{\frac{2}{\gamma}} W_t - \frac{1}{\gamma} t. \tag{14}$$

As a result, the continuous time approximation Z of the process θ_k/θ_1, with Z as in (14), is the Doléans exponential $\mathscr{E}(\sqrt{\frac{2}{\gamma}}W)$, which satisfies the stochastic differential equation

$$\mathrm{d}Z_t = \sqrt{\frac{2}{\gamma}} Z_t \, \mathrm{d}W_t. \tag{15}$$

3.6 Numerical results for the Heston model

The Heston model (see [19]) is a popular stochastic volatility model. According to Heston's model, the asset price process, denoted by S, is governed by the SDE

$$\mathrm{d}S_t = \mu S_t \mathrm{d}t + \sqrt{Z_t} S_t \mathrm{d}W_t,$$

where the process Z follows the CIR or square root process (see [6]),

$$\mathrm{d}Z_t = \kappa(\theta - Z_t)\mathrm{d}t + \sigma\sqrt{Z_t}\mathrm{d}B_t. \tag{16}$$

The driving Wiener processes W and B are correlated with correlation ρ. A common approach in quantitative finance is to log transform S to $X_t = \log S_t$. Then X obeys a diffusion equation with volatility $s(t) = \sqrt{Z_t}$,

$$\mathrm{d}X_t = (\mu - \frac{1}{2}Z_t)\mathrm{d}t + \sqrt{Z_t}\mathrm{d}W_t,$$

which can verified by Itô's formula.

Assume high-frequency observations on the log-price process X with additive noise $V_i \sim N(0, \eta_v)$ are available. Note that the volatility s is a random function. No further knowledge of the data generation mechanism is assumed. The Heston model setting is challenging for any nonparametric volatility estimation method and as such serves as a good test example. To generate synthetic data, we used the parameter values $\mu = 0.05$, $\kappa = 7$, $\theta = 0.04$, $\sigma = 0.6$, $\rho = -0.6$. These are similar to the ones obtained via fitting the Heston model to real data (see Table 5.2 in [32]). Furthermore, the noise variance was $\eta_v = 10^{-6}$. This guarantees a large enough signal-to-noise ratio in the model (9), which is a prerequisite for successful estimation of volatility. Finally, the parameter choice $\mu = 0.04$ corresponds to a realistic 4% log-return rate.

Inference results with $N = 80$ bins are given in Figure 6. These are surprisingly accurate, despite the general difficulty of the problem and the amount of assumptions that went into our estimation procedure. Indeed, the Heston model does not formally satisfy assumptions under which our Bayesian method was derived; this concerns specifically the non-deterministic character of the volatility function s.

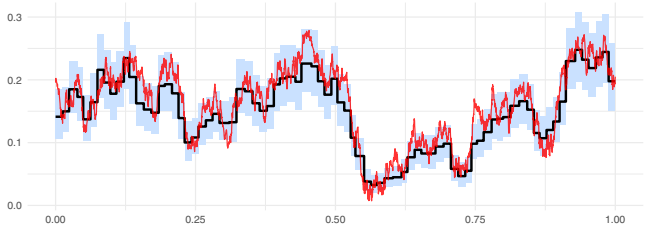

Fig. 6 Posterior mean and pointwise 95% credible band for the Heston model. The true volatility function is plotted in red, the black step function gives the posterior mean, and the credible band is shaded in blue.

We also performed experiments for a variation of the above Heston model, where we replaced the squared volatility process with the limit process $\theta_1 Z$, where Z is as in (14), or, which is the same, as in (15), with $\sqrt{\frac{2}{\gamma}} = 0.6$. These experiments

illustrate the influence of the number of bins N and the starting values, see Figures 7 and 8. Here the results with a smaller number of bins $N = 40$ combined with a low initial value are less satisfactory than with $N = 80$ and higher starting value.

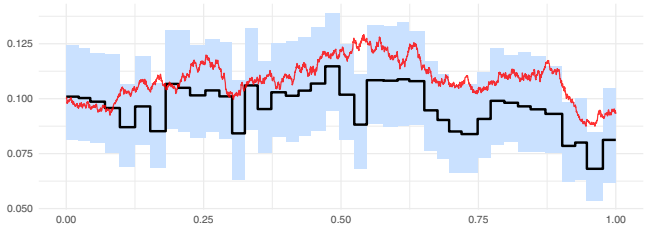

Fig. 7 Posterior results for the Heston model where the CIR volatility is replaced by the root of the continuous time limit of the prior, $N = 40$, starting value of the volatility is 0.1. The colors have the same meaning as in Figure 6.

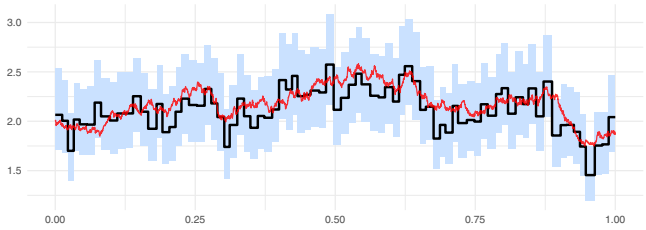

Fig. 8 Posterior results for the Heston model where the CIR volatility is replaced by the root of the continuous time limit of the prior, $N = 80$, starting value of the volatility is 2.0.

f

3.7 Exchange rate data

In this section, we infer volatility of the high frequency foreign exchange rate data made available by Pepperstone Limited, the London based forex broker.[1] As we shall show shortly, the inferred volatility looks realistic, however there is substantial uncertainty surrounding the inferential results (as demonstrated by the wide marginal credible band).

[1] Note that the data are not available from the Pepperstone website any longer, but can be obtained directly from the present authors.

In greater detail, we use the EUR/USD tick data (bid prices) for 2 March 2015. Note that the independent additive measurement error model (6) becomes harder to justify for highly densely spaced data in time. Therefore we opted to subsample the data, retaining every 10th observation. Subsampling is a common strategy in this context; see, e.g., Section 5 in [28]. In our case subsampling results in a total of $n = 13025$ observations over one day, about 9 per minute. As in the case of Heston model, we infer the volatility from the log of the observed time series. We did not apply any other data preprocessing steps (such as detection and removal of jumps from the observed data), because we wanted to focus on our Bayesian procedure itself. The data are shown in Figure 9, top panel. Note that time is rescaled to the interval $[0, 1]$, so that $t = 0$ corresponds to the midnight and $t = 0.5$ to noon.

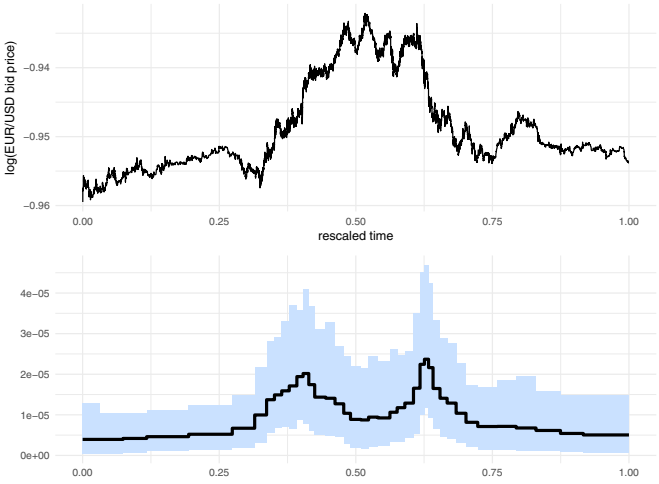

Fig. 9 Top: Natural logarithm of the EUR/USD exchange rate data for 2 March 2015 analysed in Section 3.7. Bottom: Posterior mean (black curve) and pointwise 95% credible band (blue band) for the volatility function. The time axis is rescaled to $[0, 1]$.

The inference results are given in the bottom panel of Figure 9. We used $N = 96$ bins, which corresponds to updating volatility every 15 minutes. There is considerable uncertainty in the inferred volatility. Reassuringly, the volatility is lower during the night hours. It also displays two peaks corresponding to the early morning and late afternoon hours of the day. In Figure 10 we give inference results obtained via further subsampling of the data, retaining 50% of the already subsampled observations. The posterior mean is quite similar to that in Figure 9, whereas the wider credible band reflects greater inferential uncertainty due to a smaller sample size. The stability of the inferred volatility across different observation frequency regimes partially validates the model and the approach that we use.

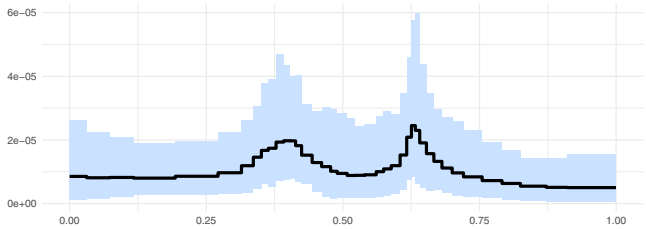

Fig. 10 Posterior mean (black curve) and pointwise 95% credible band (blue band) for the volatility of the further subsampled EUR/USD exchange rate data analysed in Section 3.7. The time axis is rescaled to $[0,1]$.

3.8 Conclusions

Volatility estimation under the presence of microstructure noise is much more challenging than the same task under noiseless observations. As such, accurate inference on volatility under microstructure noise requires large amounts of data. These, however, are readily available in financial applications. In turn, this necessitates finding a volatility estimation method that scales well with data.

The prior that we employ together with and a deliberate, but asymptotically harmless misspecification of the drift (by taking $b \equiv 0$) enable us to combine our earlier work in [17], partially reproduced here in Section 2.5, with the FFBS algorithm for Gaussian linear state space models. This results in an intuitive and fast algorithm to obtain samples from the posterior.

An interesting future research topic in this context is to explicitly account for the possible presence of jumps in financial time series.

4 Poisson intensity estimation

This section is based on [14]. Consider a Poisson point process X on an interval $[0,T]$ of the real line $\mathbb{R}$, which we equip with the Borel σ-field $\mathscr{B}([0,T])$. Such Poisson point processes are also often called non-homogeneous Poisson processes. The process X is a random integer-valued measure on $[0,T]$ (we assume the underlying (complete) probability space $(\Omega,\mathscr{F},\mathbb{Q})$ in the background), such that

1. for any disjoint subsets $B_1,B_2,\ldots,B_m \in \mathscr{B}([0,T])$, the random variables $X(B_1)$, $X(B_2),\ldots,X(B_m)$ are independent, and
2. for any $B \in \mathscr{B}([0,T])$, the random variable $X(B)$ is Poisson distributed with parameter $\Lambda(B)$. Here Λ is a finite measure on $([0,T],\mathscr{B}([0,T]))$, called the intensity measure of the process X. Moreover, it is assumed that Λ admits a density λ with respect to the Lebesgue measure on $\mathscr{B}([0,T])$.

Intuitively, the process X can be thought of as random scattering of points in $[0,T]$, where the way the scattering occurs is determined by properties (i)–(ii).

A popular assumption in the statistical literature (see, e.g., [25] and [26]) is that one observes independent copies $X_1,\ldots,X_n$ of the process X. Based on these observations, we aim to construct an estimator of the intensity function λ.

4.1 Likelihood

By Proposition 6.11 in [25] or Theorem 1.3 in [26], the law of X under the parameter value λ admits a density $p(\cdot;\lambda)$ with respect to the measure induced by a standard Poisson point process of rate 1. This density is given by

$$p(\xi;\lambda) = \exp\left(\int_{[0,T]} \log\lambda(x)\mathrm{d}\xi(x) - \int_{[0,T]} (\lambda(x)-1)\mathrm{d}x\right),$$

where $\xi = \sum_{i=1}^m \delta_{x_i}$ is a realisation of X (here δ_{x_i} denotes the Dirac measure at x_i). We assume independent observations $X_1,\ldots,X_n$ with the same distribution as X and write $X_j = \sum_{i=1}^{m_j} \delta_{X_{ij}}$ for $j=1,\ldots,n$. If we define $X^{(n)} = (X_1,X_2,\ldots,X_n)$, then it follows that the likelihood $L(X^{(n)};\lambda)$ can be written as

$$L(X^{(n)};\lambda) = \prod_{j=1}^n \exp\left(\int_{[0,T]} \log\lambda(x)\mathrm{d}X_j(x) - \int_{[0,T]} (\lambda(x)-1)\,\mathrm{d}x\right). \tag{17}$$

4.2 Prior

Fix a positive integer N. Let $0 = b_0 < b_1 < \cdots < b_{N-1} < b_N = T$ be a grid of points on the interval $\mathscr{X} = [b_0,b_N]$, for instance a uniform grid. Using this grid, define the bins $B_k = [b_{k-1},b_k)$, $k=1,\ldots,N-1$, with the last bin $B_N = [b_{N-1},b_N]$. In order to define a prior on λ, introduce a collection of piecewise constant functions λ on bins B_k,

$$\lambda(x) = \sum_{k=1}^N \psi_k \mathbf{1}_{B_k}(x), \quad x\in[0,T]. \tag{18}$$

We assume $\psi_k \overset{\text{i.i.d.}}{\sim} \mathrm{G}(\alpha,\beta)$ and define the prior Π_N on λ to be the law of the random function (18). We will refer to the prior Π_N as the independent gamma prior. If the grid on $\mathscr{X} = [0,T]$ is taken to be uniform, the prior has three hyperparameters: α, β, and N. Depending on the amount of available data and the degree of prior knowledge on λ, the hyperparameters α,β can for instance be chosen to render the $\mathrm{G}(\alpha,\beta)$ prior diffuse (non-informative). This corresponds to the case when one has little information on the magnitude and shape of λ. The hyperparameter N can be viewed as a smoothing parameter and tuned using an empirical Bayes procedure.

Using a prior with piecewise constant realisations as in (18) is not unnatural. In fact, in practical applications it is often the case that Poisson point processes are only partially observed through aggregate counts of points over successive intervals (cf. [18]). The intensity function cannot be learned beyond the resolution of these intervals, which lends support to using priors with piecewise constant realisations to model the intensity function.

4.3 Posterior

Since the function λ in our case is parametrised by the coefficients $\psi_1, \ldots, \psi_N$, the posterior distribution of the intensity function λ given the data $X^{(n)}$ can be equivalently described through the posterior distribution of $\psi_1, \ldots, \psi_N$ given $X^{(n)}$. Transition from the prior to the posterior can be thought of as updating our prior opinion on λ upon seeing the data $X^{(n)}$.

Lemma 4.1. *Define*

$$H_k = \sum_{j=1}^{n} \sum_{i=1}^{m_j} \mathbf{1}_{\{X_{ij} \in B_k\}} \tag{19}$$

and $\Delta_k = b_k - b_{k-1}$ *for* $k = 1, \ldots, N$. *Then* $\psi_1, \ldots, \psi_N$ *are a posteriori independent and*

$$\psi_k \mid X^{(n)} \sim \mathrm{G}(\alpha + H_k, \beta + n\Delta_k), \quad k = 1, \ldots, N. \tag{20}$$

Thus, the posterior for λ is known in closed form. The posterior mean is given by $\hat{\lambda}(x) = \sum_{k=1}^{N} \hat{\psi}_k \mathbf{1}_{B_k}(x)$, $x \in [0, T]$, with

$$\hat{\psi}_k = \frac{H_k + \alpha}{n\Delta_k + \beta}, \quad k = 1, \ldots, N. \tag{21}$$

Marginal $1 - \gamma$ credible bands for λ can be obtained by producing $1 - \gamma$ credible intervals for ψ_k, using the lower and upper $\gamma/2$-quantiles of the gamma distribution. This gives Bayesian uncertainty quantification for estimation of λ.

4.4 Asymptotics

In this subsection we perform the asymptotic frequentist analysis of the Bayesian procedure we introduced above. This concerns the study of the asymptotic properties of the posterior measure as the sample size $n \to \infty$.

For simplicity in the proofs, we assume that the grid $\{b_k\}$ defining the bins B_k is uniform: $b_k = Tk/N$, $k = 0, \ldots, N$, so that the bins are of equal width $\Delta_k = \Delta = T/N$. Our first result shows that the posterior mean $\hat{\lambda}$ is a consistent estimator of λ_0, and establishes its convergence rate. The expectation $\mathbb{E}[\cdot]$ here and in the sequel is always under the law of the observations with the "true" parameter value λ_0.

Theorem 4.2. *Assume the intensity function $\lambda_0\colon [0,T] \to (0,\infty)$ is Hölder continuous: there exist constants $L > 0$ and $0 < h \leq 1$, such that*

$$|\lambda_0(x) - \lambda_0(y)| \leq L|x-y|^h, \quad \forall x, y \in [0,T].$$

If $N \asymp n^{1/(2h+1)}$, then

$$\mathbb{E}[\|\hat{\lambda} - \lambda_0\|_2^2] \lesssim n^{-2h/(2h+1)}.$$

The right-hand side is the optimal rate for estimating an h-Hölder-regular intensity function, see [26].

The next result gives a posterior contraction rate in the L_2-metric. Whereas Theorem 4.2 dealt with the 'centre' of the posterior distribution, the theorem below deals with the entire posterior distribution.

Theorem 4.3. *Let the assumptions of Theorem 4.2 hold, and let $\varepsilon_n \asymp n^{-h/(2h+1)}$. Then, for any sequence $M_n \to \infty$,*

$$\mathbb{E}[\Pi_N(\|\lambda - \lambda_0\|_2 \geq M_n \varepsilon_n \mid X^{(n)})] \to 0$$

as $n \to \infty$.

The first conclusion that follows from Theorem 4.3 is that the proposed Bayesian procedure is consistent: as the sample size $n \to \infty$, the posterior puts most of its mass on (shrinking) L_2-neighbourhoods around the true parameter λ_0. Furthermore, the rate $\varepsilon_n \asymp n^{-h/(2h+1)}$ in Theorem 4.3 is the optimal posterior contraction rate for h-Hölder-regular intensity functions.

4.5 Gamma Markov chain prior

The assumption that the coefficients ψ_k of the intensity function λ are a priori independent and gamma distributed can be relaxed. For this we use ideas that have appeared in various works in the audio signal modelling literature (see, e.g., [4], [7] and [30]). This is similar to the approach taken in Section 2.5. Specifically, we postulate that the $\{\psi_k\}$ form a gamma Markov chain (GMC). This chain is defined as follows: introduce auxiliary variables $\zeta_k, k = 2, \ldots, N$, use the time ordering $\psi_1, \zeta_2, \psi_2, \ldots, \zeta_N, \psi_N$, and set

$$\psi_1 \sim \mathrm{G}(\alpha_1, \beta_1), \quad \zeta_k \mid \psi_{k-1} \sim \mathrm{IG}(\alpha_\zeta, \alpha_\zeta \psi_{k-1}), \quad \psi_k \mid \zeta_k \sim \mathrm{G}\left(\alpha_\psi, \frac{\alpha_\psi}{\zeta_k}\right), \quad (22)$$

where $k \in \{2, \ldots, N\}$. The name of the chain reflects the fact that its transition distributions are (inverse) gamma. The parameters α_1, β_1, α_ζ and α_ψ are the hyperparameters of the GMC prior. The hyperparameters α_1, β_1 allow one to 'release' the chain at the origin. This is important to avoid possible edge effects in nonparametric estimation due to a strong specification of the prior at the time origin $t = 0$. A principal aim in using latent variables ζ_k's in (22) is to attain positive correlation

between ψ_k's. In the intensity function modelling context this induces smoothing across different bins.

The posterior distribution with the GMC prior is not available in closed form. However, "full" conditional distributions are known in closed form which facilitates Gibbs sampling. The full conditional distributions of ψ_k, ζ_k are

$$\zeta_k \mid \psi_k, \psi_{k-1} \sim \mathrm{IG}\left(\alpha_\zeta + \alpha_\psi, \alpha_\zeta \psi_{k-1} + \alpha_\psi \psi_k\right), \quad k = 2, \ldots, N, \tag{23}$$

$$\psi_k \mid \zeta_{k+1}, \zeta_k, X^{(n)} \sim \mathrm{G}\left(\alpha_\psi + \alpha_\zeta + H_k, \frac{\alpha_\psi}{\zeta_k} + \frac{\alpha_\zeta}{\zeta_{k+1}} + n\Delta_k\right), \quad k = 2, \ldots, N-1, \tag{24}$$

$$\psi_1 \mid \zeta_2, X^{(n)} \sim \mathrm{G}\left(\alpha_1 + \alpha_\zeta + H_1, \beta_1 + \frac{\alpha_\zeta}{\zeta_2} + n\Delta_1\right), \tag{25}$$

$$\psi_N \mid \zeta_N, X^{(n)} \sim \mathrm{G}\left(\alpha_\psi + H_N, \frac{\alpha_\psi}{\zeta_N} + n\Delta_N\right). \tag{26}$$

The Gibbs sampler cycles through formulae (23)–(26) to generate samples from the posterior distribution of $\{\psi_k\}$, $\{\zeta_k\}$ given the data $X^{(n)}$. One can initialize the sampler e.g. by providing values for $\psi_1, \ldots, \psi_N$.

4.6 Numerical results

Consider

$$\lambda_0(x) = 2e^{-x/5}(5 + 4\cos(x)), \quad x \in [0, 10].$$

A principal challenge in inferring this function consists in the fact that it takes small values in the middle part of its domain and has a slope of changing sign. In Figure 11 the posterior mean and marginal 95% posterior credible bands are shown. From the lower panels one can clearly see the beneficial smoothing effect of the GMC-prior.

We refer to [35] for an application of the proposed methods in estimating heart rate variability.

4.7 Closing remarks

The presented approach is computationally appealing, as it scales with the number of bins and not the sample size. One can think of the chosen number of bins as the resolution at which one estimates the intensity function. If this resolution is too fine, the independent Gamma prior results in overfitting. This can be avoided by using the GMC-prior, which induces a smoothing effect on neighbouring bins.

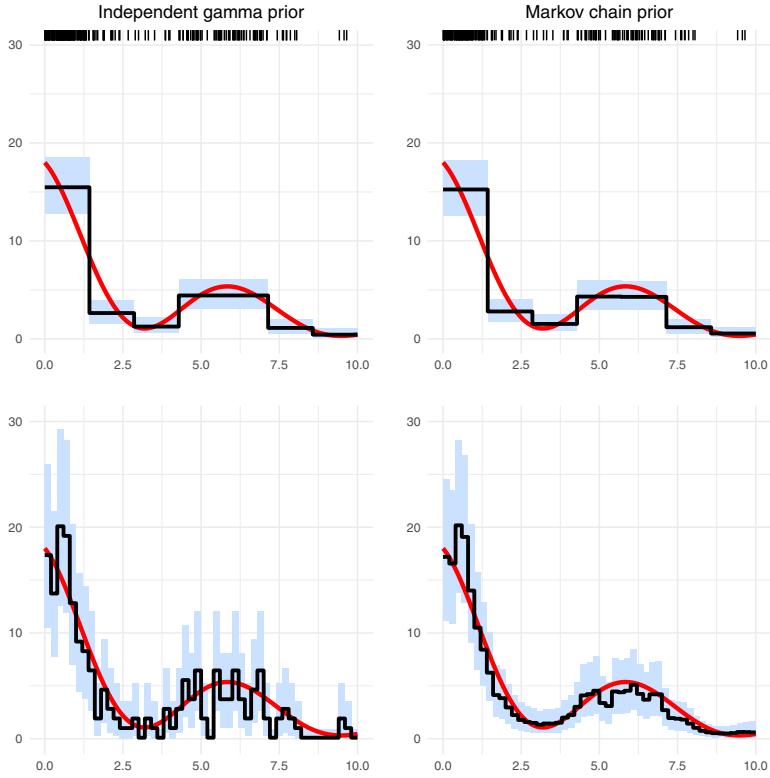

Fig. 11 Estimation results for the oscillating exponential function $\lambda_0(x) = 2e^{-x/5}(5+4\cos(x))$ with $n = 5$ replicates. The top plot corresponds to the method based on the independent gamma prior, with $N = 7$ chosen via the empirical Bayes method (details are in Section 3.3 of [14]). The bottom plot corresponds to the method based on the GMC prior with $N = 50$ bins.

5 Estimation for gamma-driven stochastic differential equations

The goal of the present section, based on [2], is Bayesian nonparametric estimation of the positive local scale function σ that appears in the Lévy-driven stochastic differential equation

$$\mathrm{d}X_t = \sigma(X_{t-})\,\mathrm{d}L_t,\ X_0 = 0, \tag{27}$$

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 ν admitting the Lévy density

$$\nu(x) = \alpha x^{-1}\exp(-\beta x),\ x > 0, \tag{28}$$

where α and β are two positive constants. The process L has independent increments, and $L_t - L_s$ has the Gamma$(\alpha(t-s),\beta)$ distribution for $t > s$, i.e. the distribution with density $p(x) = \frac{\beta^{\alpha(t-s)}}{\Gamma(\alpha(t-s))} x^{\alpha(t-s)-1} \exp(-\beta x)$, $x \geq 0$.

5.1 Preliminaries

Under the assumption that the function σ (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 [3] that Equation (27) admits a weak solution that is unique in law. Note that X is a Markov process. Under the stronger condition that σ is Lipschitz continuous, Equation (27) even has a unique strong solution, see [33, Theorem V.6].

We are ultimately interested in asymptotic properties of our Bayesian procedure, detailed below, for estimating the volatility function σ. To that end, we need a sufficiently rich set of observations and this can be accomplished by the scaled observation process in (29) below, instead of X satisfying (27). So we consider the process X^n given as the solution to

$$\mathrm{d}X_t^n = \frac{1}{n}\sigma(X_{t-}^n)\,\mathrm{d}L_t,\ X_0^n = 0. \tag{29}$$

The scaling factor $\frac{1}{n}$ causes, for large values of n, a 'slow growth' of the process X^n and 'long times' to stay around certain levels. We will thus assume that X^n is observed on a long time interval $[0,T^n]$, where $T^n \to \infty$. In Section 5.4 we will specify T^n and derive asymptotic results for $n \to \infty$.

5.2 Likelihood

For a Bayesian analysis we need to convert a prior distribution into a posterior one, and for this a likelihood ratio is needed. Let us first sketch the set-up.

Let $(\Omega,\mathscr{F},\mathbb{F},\mathbb{P})$ 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 v given by (28). Assume that X is a (weak) solution to (27), and assume that X is observed on an interval $[0,T]$. We denote by $\mathbb{P}_T^\sigma$ its law, a probability measure on the Skorohod space $D[0,T]$ with the Skorohod J_1-topology and the σ-algebra generated by it. 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. Recall (28) and let

$$Y(t,x) := \frac{1}{\sigma(X_{t-})} v\Big(\frac{x}{\sigma(X_{t-})}\Big)/v(x).$$

In [3] the following proposition has been proven.

Proposition 5.1. *Let σ be a positive locally bounded measurable function on $[0,\infty)$ such that* (27) *admits a weak solution that is unique in law. Assume moreover that σ 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 $\mathscr{F}_T^X$ and the corresponding Radon-Nikodym derivative Z_T has the explicit representation:*

$$Z_T = \exp\left(\int_0^T \int_0^\infty \log Y(t,x)\,\mu^X(\mathrm{d}x,\mathrm{d}t) - \int_0^T \int_0^\infty (Y(t,x)-1)\nu(x)\,\mathrm{d}x\,\mathrm{d}t\right), \quad (30)$$

where μ^X is the jump measure of X and both double integrals are a.s. finite.

In the next section we will present, under additional assumptions, an explicit expression for Z_T.

5.3 Prior and posterior

We model σ as a piecewise constant function:

$$\sigma(x) = \sum_{k=1}^K \xi_k \mathbf{1}_{B_k}(x) \quad (31)$$

for bins $B_1 = [0,b_1]$, $B_k = (b_{k-1},b_k], k = 2,\ldots,K-1$, and $B_K = (b_{K-1},b_K)$, with appropriately chosen increasing sequence of bin endpoints $\{b_k\}$ and the bin number K. The $\{\xi_k\}$ are positive numbers (later on positive random variables). Although we use (31) for our model, we emphasize that the 'true' σ does not need to be piecewise constant. As a final remark we note that when σ is given by (31), (29) still has a unique solution, obtained as a concatenation of stopped gamma processes. Under the specification (31), the general likelihood ratio Z_T of (30) gets an explicit form.

Proposition 5.2. *Suppose that X^n is given by* (29) *with σ given by* (31). *Let $\tau_k^n = \inf\{t \geq 0 : X_t^n \geq b_k\}$, $k = 1,\ldots,K$, and write $T = \tau_K^n$. Then*

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

To assign a prior distribution, we take the ξ_k to be independent random variables with an inverse gamma distributions for each of them, that is $\xi_k \sim \mathrm{IG}(\alpha_k,\beta_k)$, and thus have marginal densities $p_k(x) = \beta_k^{\alpha_k}\Gamma(\alpha_k)^{-1}x^{-\alpha_k-1}\exp(-\beta_k/x)$. Therefore, we possibly have to extend the original probability space to carry the ξ_k as well, taking into account that L and the ξ_k have to be independent. We can then use Proposition 5.2, where $\mathbb{P}_T^\sigma$ is to be interpreted as the conditional law of X^n on $[0,T]$ given the ξ_k. It turns out that the posterior distributions of the ξ_k are again of inverse gamma type.

Lemma 5.3. *Let $x_1,\dots,x_K$ denote realisations of $\xi_1,\dots\xi_K$ and let X^n stand for the path $X^n_t, t\in[0,T]$. With $T=\tau^n_K$, we then have from Proposition 5.2 that the posterior joint density of $(\xi_1,\dots,\xi_K)$ is given by*

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

It follows that the ξ_k are independent under the posterior distribution, and

$$\xi_k \mid X^n \sim \mathrm{IG}(\alpha\Delta\tau^n_k+\alpha_k, n\beta\Delta X^n_{\tau^n_k}+\beta_k), \tag{32}$$

where $\Delta\tau^n_k=\tau^n_k-\tau^n_{k-1}$, $\Delta X^n_{\tau^n_k}=X^n_{\tau^n_k}-X^n_{\tau^n_{k-1}}$.

5.4 Asymptotics

We consider the process X^n satisfying (29), but the standing assumption from here on is that $\sigma:[0,\infty)\to[0,\infty)$ 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, σ is assumed to be bounded from below by a positive constant $\underline{\sigma}$. Here are some further assumptions.

- The number of bins and their width depend on n. So we write $B^n_k=(b^n_{k-1},b^n_k]$, $k=1,\dots,K$, $K=K_n$. 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^n_k as $b^n_k=\frac{b_K k}{K}$, $k=1,\dots,K$. A given $x\in(0,b_K]$ then belongs to bin B^n_k with $k=k_n(x)=\lceil\frac{Kx}{b_K}\rceil$.
- 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^n_k, one has $\Delta b^n_k:=b^n_k-b^n_{k-1}\asymp n^{-\kappa}$ for all k and $x\in B_k$. Note that the bin widths decrease to zero and that the number of bins increases to infinity, with a rate depending on the parameter κ (to be specified below). This behaviour is needed to have better and better approximations of the true function σ with piecewise constant functions.
- We observe the process X^n until it crosses the last bin, i.e. until time $T^n=\tau^n_K$. In [2] it has been shown that the random time T^n is roughly proportional to n.

Although σ is continuous, we assume it to be apriori piecewise constant. That is, parallel to (31) but note that the bins now depend on n, as

$$\xi^n(x)=\sum_{k=0}^{K}\xi_k\mathbf{1}_{B^n_k}(x), \tag{33}$$

where the ξ_k are assigned the inverse gamma prior distributions as in Section 5.3. Consequently, the posterior distributions of the ξ_k are as in (32).

Here is the main result, it says that the contraction rate of the posterior distribution is $n^{-\lambda/(2\lambda+1)}$, λ being the Hölder exponent of σ. We use the notation Π_n

to denote posterior probabilities given the observations of X^n on the time interval $[0,T^n]$.

Theorem 5.4. *Assume the model with volatility* (33) *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 $n^{-\frac{1}{2\lambda+1}}$ (i.e. $\kappa = \frac{1}{2\lambda+1}$), and let (m_n) be any sequence of real numbers diverging to infinity. Then, for $n \to \infty$*

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

Moreover, this contraction rate is minimax optimal.

5.5 Numerical illustration

Here we graphically illustrate our inferential results. In agreement with the notation pertaining to the asymptotic regime detailed in (29) we took X^n as the solution to the Lévy SDE (27) with the volatility function

$$\sigma_0(x) = \frac{3}{2} + \sin(2\pi x) \tag{34}$$

and scaled it with $1/n$, $n = 500$. We have partitioned the unit interval into 10 bins setting $b_k = k\delta x$, $\delta x = 0.1$, $k = 1,\ldots,10$, and used the piecewise constant prior of the form $\sigma(x) = \sum_k \xi_k \mathbf{1}_{B_k}(x)$, where $\xi_1,\ldots,\xi_{10}$ are i.i.d. random variables with the inverse gamma $\mathrm{IG}(\alpha_k,\beta_k)$ distribution with $\alpha_k = \beta_k = 0.1$. The posterior is given in closed form by (32). Figure 12 contrasts the corresponding marginal 90 %-posterior credible bands for σ with the true volatility function σ_0 from (34).

5.6 Closing remarks

There are many extensions conceivable. The inclusion of a drift in Equation (27) is one possibility, another one is to move from a continuous record of observations to a discrete time setting. See [2] for short discussions of such possible extensions and also for an extensive treatment of an application to real data obtained from the North Greenland Ice Core Project. The present section concentrated on the theoretical side.

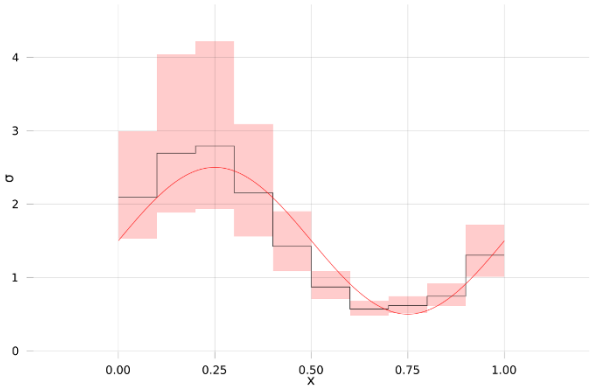

Fig. 12 Red: true volatility σ_0 as a function of x. Shaded: marginal 90 %-posterior credible band for the piecewise constant posterior σ. Black: marginal posterior median.

6 Gamma-type Lévy subordinators

Given discrete time observations over a growing time interval, we consider in this section a nonparametric Bayesian approach to estimation of the Lévy density of a Lévy process belonging to a flexible class of infinite activity subordinators. We establish that our nonparametric Bayesian procedure is consistent: in the low frequency data setting, with equispaced in time observations and intervals between successive observations remaining fixed, the posterior asymptotically, as the sample size $n \to \infty$, concentrates around the Lévy density under which the data have been generated. Finally, we test our method on a classical insurance dataset. This section is based on [1].

6.1 Problem formulation

Consider a univariate Lévy process $X = (X_t : t \geq 0)$ with Lévy triplet (γ, A, ν). By the Lévy-Khintchine formula, see Theorem 8.1 in [36], the characteristic function ϕ_{X_1} of X_1 admits the unique representation of the type

$$\phi_{X_1}(z) = \exp\left(\mathrm{i}\gamma z - \frac{1}{2}Az^2 + \int_{\mathbb{R}} \left(e^{\mathrm{i}zx} - 1 - \mathrm{i}zx\mathbf{1}_{|x|\leq 1}\right)\nu(\mathrm{d}x)\right),$$

and the triplet (γ, A, ν) uniquely determines the law of the process X. We assume that the Lévy measure ν admits the representation

$$\nu(\mathrm{d}x) = \frac{\beta}{x}e^{-\alpha x - \theta(x)}\mathrm{d}x, \quad x > 0, \tag{35}$$

where α and $\theta\colon [0,\infty) \to \mathbb{R}$ are parameters to be estimated, while β is a known or unknown parameter. This entails that θ and α are such that $\nu([1,\infty))$ is finite. We also assume that $A = 0$ and

$$\gamma = \int_{0\le x\le 1} x\nu(\mathrm{d}x) < \infty. \tag{36}$$

It follows that X is a pure jump process with non-decreasing sample paths, or put another way, a subordinator with zero drift, cf. Sections 2.6.1–2.6.2 in [27]. As in our case the Lévy triplet (γ, A, ν) is entirely determined by ν, we will write $\mathbb{P}_\nu$ for the law of X. One may call this class of Lévy processes Gamma-type subordinators, because X is a Gamma process when $\theta \equiv 0$, but we prefer to simply refer to it as θ-subordinators.

Assume that the process X is observed at discrete time instances $0 = t_0 < t_1 < \cdots < t_n = T$, so our observations are $X^{(n)} = (X_{t_i}\colon i \in \{0,\dots,n\})$. Our aim is nonparametric Bayesian estimation for the parameter triple (α,β,θ). This requires specification of the likelihood and the prior in our model, that are next combined via Bayes' formula to form the posterior distribution. This latter encodes all the necessary inferential information within the Bayesian setup.

6.2 Likelihood

We consider the problem with β known and fixed in the present treatment. This choice has been made because of two reasons. First, the parameter β plays a role similar to the diffusion coefficient σ in a stochastic differential equation driven by a Wiener process; different values of σ there result in mutually singular laws of a continuous record of observations, and similar singularity is valid for θ-subordinators. Derivation of nonparametric Bayesian asymptotics for the stochastic differential equations (all of which is a recent work) historically proceeded from the assumption of a known σ to the one where σ is unknown and has to be estimated. In that sense, the fact that we assume β is known at this stage does not appear unexpected or unnatural. There is also a technical reason to assume β to be known. It assists in deriving useful bounds on the Kullback-Leibler and Hellinger distances between marginals of θ-subordinators under different Lévy triplets, which in general, is the key to establishing consistency properties of nonparametric Bayesian procedures. For a treatment of the statistical problem with observations in discrete time when also θ is unknown, we refer to [1, Section 6], and for now, we maintain the assumption of known β.

All processes and their laws in this section are restricted to the time interval $[0,T]$ for a fixed $T > 0$. Note that for any two Lévy measures ν and ν_0 given by (35) with parameters β,α,θ and β,α_0,θ_0, respectively, provided $\theta(0) = \theta_0(0) = 0$ and both functions θ and θ_0 are Lipschitz continuous in some neighbourhood of zero, we have ν and ν_0 are equivalent. Moreover,

$$d^2_{\mathscr{H}}(\nu,\nu_0) = \frac{1}{2}\int_{(0,\infty)} (\sqrt{\mathrm{d}\nu} - \sqrt{\mathrm{d}\nu_0})^2 < \infty, \tag{37}$$

where $d_{\mathscr{H}}(\cdot,\cdot)$ is the Hellinger distance between two (infinite) measures. By assumption (36) and property (37), together with Theorem 33.1 in [36], it follows that the laws $\mathbb{P}_\nu$ and $\mathbb{P}_{\nu_0}$ of $X = (X_t : t \in [0,T])$ are equivalent. Furthermore, Theorem 33.2 in [36] implies that a.s.

$$U_T = \log\left(\frac{\mathrm{d}\mathbb{P}_\nu}{\mathrm{d}\mathbb{P}_{\nu_0}}(X)\right) = \sum_{(s,\Delta X_s)\in(0,T]\times\{\Delta X_s>0\}} \phi(\Delta X_s) - T\int_{(0,\infty)} (e^{\phi(x)}-1)\nu_0(\mathrm{d}x),$$

where $\Delta X_s = X_s - X_{s-}$, and

$$\phi(x) = \log\left(\frac{\mathrm{d}\nu}{\mathrm{d}\nu_0}(x)\right) = -(\alpha x + \theta(x) - \alpha_0 x - \theta_0(x)), \quad x > 0.$$

We can also write the log-likelihood ratio U_T as

$$U_T = \int_{(0,T]}\int_{(0,\infty)} \phi(x)\mu(\mathrm{d}s,\mathrm{d}x) - T\int_{(0,\infty)} (\nu-\nu_0)(\mathrm{d}x),$$

where the jump measure μ is defined by

$$\mu((0,t]\times B) = \#\{s:\ (s,\Delta X_s) \in (0,t]\times B\}$$

for any Borel subset B of $(0,\infty)$. We can view $\mathbb{P}_{\nu_0}$ as the dominating measure for $\mathbb{P}_\nu$. From an inferential point of view the specific choice of the dominating measure is immaterial. A convenient choice of ν_0 for the theoretical development in Section 6.4 is to actually take ν_0 to be the 'true' Lévy measure ν_0 with parameters α_0 and θ_0.

6.3 Prior

To define the prior, we consider a subclass of processes defined in (35), where the parameter θ in the Lévy measure ν has the following form. Fix a sequence $0 < b_1 < \cdots < b_N < \infty$, set for convenience $b_0 = 0$ and $b_{N+1} = \infty$, and define bins B_k by

$$B_k = [b_k, b_{k+1}), \quad k = 0,\ldots,N.$$

Given bins B_k, assume the function θ is, unlike those in previous sections, piecewise linear, i.e.,

$$\theta(x) = \sum_{k=1}^{N} (\rho_k + \theta_k x)\mathbf{1}_{B_k}, \tag{38}$$

where $\rho_k \in \mathbb{R}$, $k = 1,\dots,N$, $\theta_k \in \mathbb{R}$, $k = 1,\dots,N$, and $\theta_N > -\alpha$. Together with α, the parameter θ_k determines the slope of the function $\theta(x) + \alpha x$ on the bin B_k, while ρ_k gives the intercept. The process X with the law $\mathbb{P}_\nu$ can be viewed as a Gamma process with rate parameter α and shape parameter β, subjected to local deviations in the behaviour of jumps of sizes falling in bins B_k compared to that of a Gamma process. The parameters θ_k, ρ_k quantify the extent of these local deviations on the bin B_k.

We model α, θ_k, ρ_k as independent under the prior distribution. Note such a prior on α, θ_k, ρ_k implicitly defines a prior on the Lévy measure ν as well. The specific form of the prior is not crucial for many arguments that follow, but is convenient computationally. In fact, theoretical results in Section 6.4 can be derived for other series priors as well. However, the local linear structure in (38) (which also means that the prior could be rewritten as a series prior with basis functions having compact support) is important to derive some simple update formulae below. For a realisation ν from the implicit prior on ν as above in the present section, let us work out the integral

$$\nu(B_k) = \int_{b_k}^{b_{k+1}} \frac{\beta}{x} e^{-(\alpha+\theta_k)x - \rho_k} \mathrm{d}x,$$

which enters the expression for the likelihood. To that end, remember the definition of the exponential integral, $E_1(z) = \int_z^\infty t^{-1} e^{-t} \mathrm{d}t$, see, e.g., §15.09 in [24] for its basic properties. Then a change of the integration variable gives

$$\nu(B_k) = \beta e^{-\rho_k} \{E_1((\theta_k + \alpha)b_k) - E_1((\theta_k + \alpha)b_{k+1})\}, \quad k = 1,\dots,N. \tag{39}$$

Observe that $\nu(B_N) = \beta e^{-\rho_N} E_1((\theta_k + \alpha)b_N)$. Similar to the case of ν,

$$\nu_0(B_k) = \beta \{E_1(\alpha_0 b_k) - E_1(\alpha_0 b_{k+1})\}, \quad k = 1,\dots,N.$$

Also here remark that $\nu_0(B_N) = \beta E_1(\alpha_0 b_N)$. For future reference, note that for any α, α',

$$\lim_{x \to 0} \{E_1(\alpha x) - E_1(\alpha' x)\} = \log\left(\frac{\alpha'}{\alpha}\right), \tag{40}$$

which follows from the formula for Frullani's integral, see §12.16 in [24].

6.4 Asymptotics

Let the process X be observed at equidistant times t_i, $i = 1,\dots,n$. Without loss of generality we assume that the observations are $X_1,\dots,X_n$. We pass to the increments $Z_i = X_i - X_{i-1}$, $i = 1,\dots,n$, and denote our observations by $\mathscr{Z}_n = \{Z_1,\dots,Z_n\}$. We assume that under the true Lévy density v_0, $Z_1 \sim \mathbb{Q}_{v_0}$. In general, $\mathbb{Q}_v$ will stand for the law of the increment Z_1 under the Lévy density v. Furthermore, we introduce the law $(X_t : t \in [0,1]) \sim \mathbb{R}_{v_0}$ under the true Lévy density v_0. The law of this path under the Lévy density v will be denoted by $\mathbb{R}_v$. For asymptotic results, we will let the

number of bins N depend on the sample size n, and write N_n instead. We postulate a condition on the underlying model and on the chosen prior distribution.

Condition 1 Let the function θ_0 have a compact support on the interval $[\underline{b}, \overline{b}]$ where the boundary points $0 < \underline{b} < \overline{b} < \infty$ are known, $\|\theta_0\|_\infty < \bar{\theta}$, and suppose θ_0 is λ-Hölder continuous, that is, $|\theta_0(x) - \theta_0(y)| \le L|x-y|^\lambda$ ($\lambda \in (0,1]$, $L > 0$). Suppose also that $\alpha_0 \in [\underline{\alpha}, \overline{\alpha}]$ with known boundary points $0 < \underline{\alpha} < \overline{\alpha} < \infty$. Finally, assume that the parameter β_0 is known (and, without loss of generality, equal to 1.

Condition 2 The prior Π_n is defined on a special class of Lévy densities, V_n. These are the densities that on the bins $B_k = (b_{k-1}, b_k]$, $k = 1, \ldots, N$, $b_0 = 0$, $b_1 = \underline{b}$, $b_N = \overline{b}$, have the form $v(x) = \frac{\beta_0}{x} \exp(-\alpha x - \theta_k(x))$, with $\theta_k(x) = \rho_k + \theta_k x$, with the special choice $\rho_0 = \theta_0 = 0$ and $\beta_0 = 1$. So, with the above notation,

$$V_n = \left\{ v \colon v_{|B_k}(x) = \frac{\beta_0}{x} \exp(-\alpha x - \theta_k(x)), k = 1, \ldots, N \right\}.$$

Under the prior distribution, the coefficients θ_i, $i = 1, \ldots N-1$, are independent and uniformly distributed on the known interval $[-\overline{\theta}, \overline{\theta}]$, $\overline{\theta} > 0$. Likewise, the coefficients ρ_i, $i = 1, \ldots, N-1$, are independent uniform on the interval $[-\overline{\theta}, \overline{\theta}]$, whereas α is uniform on $[\underline{\alpha}, \overline{\alpha}]$, $\overline{\alpha} > 0$. We assume a priori independence of all ρ_i, θ_i and α. Implicitly, this defines a prior Π_n on the class of Lévy densities V_n, which are realisations from the prior.

Theorem 6.1 below implies that our Bayesian procedure is consistent in probability; this in turn implies the existence of consistent Bayesian point estimates, see, e.g., [12], pp. 506–507. We use the notation $\Pi_n(\mathrm{d}v \mid \mathscr{Z}_n)$ for the posterior measure. Also, $\mathbb{Q}^n_{v_0}$ denotes the law of the sample $\mathscr{Z}_n$ under the true Lévy density v_0.

Theorem 6.1. *Assume that Conditions 1 and 2 hold and that $N_n \to \infty$ and $N_n/n \to 0$ as $n \to \infty$. Let $d_{\mathscr{H}}$ be the Hellinger metric. Then, for any fixed $\varepsilon, \varepsilon' > 0$,*

$$\mathbb{Q}^n_{v_0} \left(\Pi_n(v \colon d_{\mathscr{H}}(\mathbb{Q}_{v_0}, \mathbb{Q}_v) > \varepsilon \mid \mathscr{Z}_n) > \varepsilon' \right) \to 0$$

as $n \to \infty$.

6.5 Numerical results

Over the last two decades there has been an increasing interest in applying Bayesian methods to insurance problems, see, e.g., [21] and references therein. [22] apply a Dirichlet process mixture prior to model the density of insurance claim sizes, and provide motivation for using a nonparametric Bayesian approach in the actuarial science. In this section we will apply our Bayesian approach to the Danish data on large fire losses. This dataset is a standard test example in extreme value theory, and from that point of view it has been a subject of several deep studies, such as [29]

and [34]. Our goals here are more modest, and aim at demonstrating the facts that firstly, θ-subordinators can be potentially used to capture some aggregate features of the Danish data on large fire losses, and secondly, statistical inference for real data modelled through such processes can be successfully performed using the Bayesian methodology developed here.

Data description and visualisation

A succinct description of the Danish data on large fire losses can be found on p. 298 in [9]. The dataset (scaled for privacy reasons) comprises 2167 fire losses (adjusted suitably for inflation to reflect the 1985 values) in Denmark over the 10 year period starting on 6 January 1980 and ending on 30 December 1990, that exceed in size one million DKK, and that were registered by Copenhagen Reinsurance. The rationale for thresholding losses at one million DKK is given in [29], pp. 119–120, and consists in the fact that in practice it is virtually impossible to collect exhaustive data on small losses: insurance is typically provided against significant losses, while small losses are dealt with by insured parties directly. The data can be accessed through the **QRM** package in **R** under the name `danish`. The time plot of the data is given in the left panel of Figure 13.

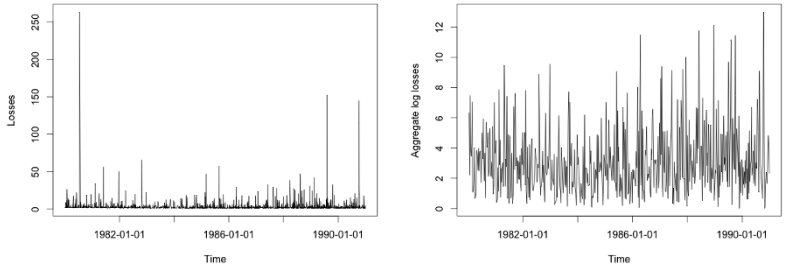

Fig. 13 Danish data on large fire losses. *Left*: original daily data (the unit is one million DKK). *Right*: logarithmically transformed and aggregated data.

According to the exploratory analysis of the transformed data, the data can be modelled as an i.i.d. sequence that follows a Gamma-like distribution, but perhaps is not genuinely Gamma. This suggests a possibility of using a θ-subordinator to model the data.

Modelling fire losses with a θ-subordinator

Because the sample size is much smaller compared to our simulation examples, we chose $N = 1$ corresponding to a single grid point $b_1 = 2$ and four parameters

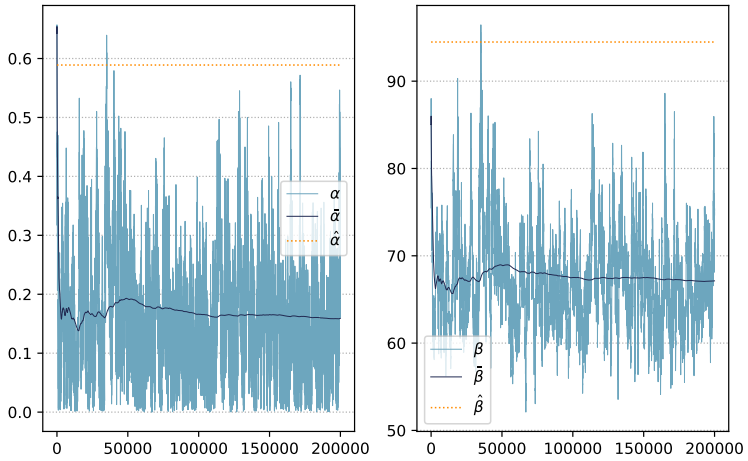

Fig. 14 Trace plots of the parameters α and β for the fire loss data. Left: trace and running average of samples of α. (The latter indicated by $\bar{\alpha}$.) The maximum likelihood estimate $\hat{\alpha}$ of α using a Gamma process model is marked as the dotted yellow line. Right: trace and running average of samples of β. (The latter indicated by $\bar{\beta}$.) The maximum likelihood estimate $\hat{\beta}$ of β using a Gamma process model is marked as the yellow dotted line.

α, β, θ_1, ρ_1. In order to improve mixing of the chain, we use a reparametrization $\alpha_1 = \alpha + \theta_1$, $\beta_1 = \beta \exp(-\rho)$, and work with four parameters α, β, α_1, β_1, so that

$$v(x) = \begin{cases} \frac{\beta}{x} \exp(-\alpha x) & x \le b_1, \\ \frac{\beta_1}{x} \exp(-\alpha_1 x) & x > b_1. \end{cases}$$

A priori we equip these four parameters with Gamma distributions, with mean 0.75 and variance 0.36 for the parameters α, α_1, and mean 90 and variance 2500 for the parameters β, β_1, and in such a way that they become independent random variables. In the data augmentation step we take intermediate points at distance 0.0192, corresponding to $m = 1000$.

For the parameter updates we took independent Gaussian innovations with standard deviations $\sigma_\alpha = \sigma_{\alpha_1} = 0.03$, $\sigma_\beta = 1$ and $\sigma_{\beta_1} = 6$, respectively. In the Gibbs sampler in each step new Gamma bridges are proposed in the data augmentation step, followed by a parameter update Metropolis-Hastings step cycling through updates of β in the first and second and the other parameters jointly in each of the remaining three of in total 5 stages. With these choices, the chains mix sufficiently well. The MCMC algorithm was ran for 200 000 iterations. Figure 14 shows trace plots and running averages of the posterior samples of the parameters α and β, whereas Figure 15 shows similar plots for the parameters α_1 and β_1.

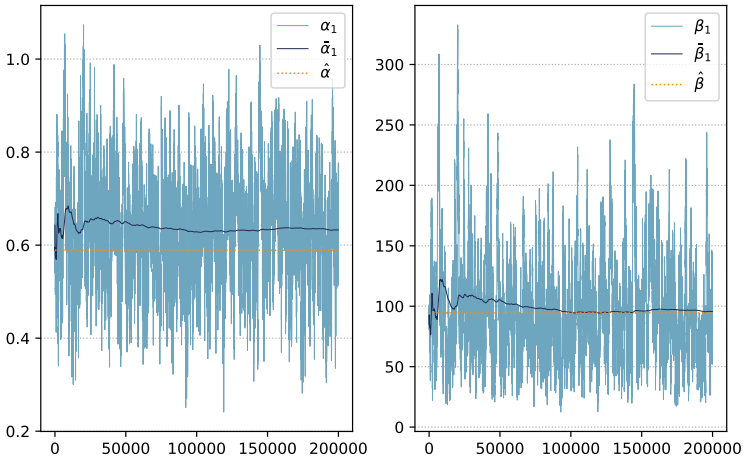

Fig. 15 Trace plots of the parameters used for the bin (b_1, ∞) for the fire loss data. Left: trace and running average of the samples of α_1. The maximum likelihood estimate of α using a Gamma process model is marked as yellow line. Right: trace and running average of the samples of β_1. The maximum likelihood estimate of β using a Gamma process model is marked as yellow line.

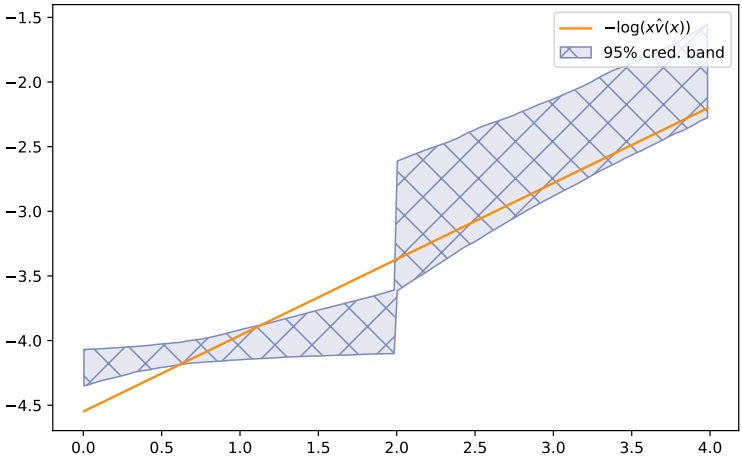

Fig. 16 Marginal Bayesian credible bands for the fire loss data for the function $-\log(xv(x))$ based on all samples. Yellow: maximum likelihood estimate $-\log(x\hat{v}(x))$ assuming a Gamma process.

Figure 16 shows the 95% marginal Bayesian credible band for the function $\theta(x) + \alpha x$ contrasted to the maximum likelihood estimate that assumes the observations come from a Gamma process. This plot suggests that modelling the losses with a Gamma process leads to overestimation of the number of small jumps and possibly of large jumps too; however, more data is necessary to make a definitive statement (unfortunately, as observed in [5], it is difficult for academia to gain access to the insurance data). In connection to this, we note that a difference in the estimates of the rate of decay of the Lévy density (value of α_1 in the model) has serious implications of practical relevance for the assessment of the risk of very large fire losses.

6.6 Conclusions and possible extensions

We have considered in this section a nonparametric Bayesian approach to estimation of the Lévy density of a Lévy process belonging to a flexible class of infinite activity subordinators and established that this procedure is consistent. We have also tested our method on a classical insurance dataset.

There are various directions for extensions of the model studied in this section. One may consider a class of increasing, infinite activity Lévy processes X for which the characteristic function of X_1 is given by

$$\phi(z) = E[e^{\mathrm{i}zX_1}] = \int (e^{\mathrm{i}zx} - 1)\nu(dx),$$

where $\theta(\cdot)$ is nonnegative and the Lévy measure ν is given by

$$\nu(\mathrm{d}x) = \frac{1}{x^{1+a}} e^{-\theta(x)} \mathbf{1}_{(0,\infty)}(x)\mathrm{d}x.$$

This model generalizes the Gamma process with $a = 0$ and $\theta(x) \equiv \lambda$ and also covers the class of one-sided tempered stable processes. Here one could think of estimating the stability index α from a Bayesian point of view. This is difficult due to singularity of the measures induced by two Lévy processes with different stability indices, and so this poses an interesting challenge that deserves to be explored.

7 Summary and concluding remarks

In this paper, we have reviewed some of our contributions, obtained with our collaborators, to Bayesian nonparametric function estimation. We have considered a large variety of models (volatility estimation in diffusion models, Poisson intensity estimation, estimation for gamma-driven stochastic differential equations, estimation in the presence of microstructure noise, and Lévy density estimation for Gamma-

type subordinators), which all had a common denominator, the function to be estimated, is well approximated by piecewise constant functions (with one exception that deals with piecewise linear functions) giving rise to tractable prior distributions for Bayesian inference.

The goal of inference in all models under consideration was to obtain posterior consistency of the proposed estimators or even to quantify posterior contraction rates. We have seen that rougher continuous functions are more difficult to estimate than smoother ones. The degree of smoothness is in our approach measured by the Hölder exponent; the smoother the function is, the better the contraction rate of the estimation procedure is. The precise results are formulated as Theorems 2.4, 4.3, and 5.4. The conclusion that emerges from these theorems is that despite the entirely different settings and the different kinds of functions (their arguments being time or space variables), the contraction rates, all depending on the Hölder exponent, are the same across all cases. A deeper analysis in terms of equivalence of statistical experiments is likely to explain the similarity of the rates. These rates were obtained under an independence assumption on the prior, i.e., the approximating functions with constant levels on each of their bins are such that these levels are a priori independent. The setting of Section 6 was somewhat different, and we had to content ourselves with posterior consistency only, Theorem 6.1.

It also appeared that better numerical results, visible in the plots, can be obtained by relaxing the independence assumption on the prior distribution. This observation has been accomplished by using a specific Markov chain prior on the levels of the piecewise constant functions, introducing a priori dependence between the levels on adjacent bins of the piecewise constant approximating functions. However, the contraction rates induced by such a prior cannot improve upon the one for the prior that renders the levels independent, as the latter has been seen to be optimal.

Throughout the paper, we have given ample numerical illustrations that support the theoretical results, ranging from simulated experiments to real-world cases.

Acknowledgements The authors gratefully acknowledge the contributions to this paper by Shota Gugushvili. The work of D. Belomestny was done within the framework of the HSE University Basic Research Program.

References

1. Belomestny, D., Gugushvili, S., Schauer, M., Spreij, P.: Nonparametric Bayesian inference for Gamma-type Lévy subordinators. Commun. Math. Sci. **17**(3), 781–816 (2019). DOI 10.4310/CMS.2019.v17.n3.a8
2. Belomestny, D., Gugushvili, S., Schauer, M., Spreij, P.: Nonparametric Bayesian volatility estimation for gamma-driven stochastic differential equations. Bernoulli **28**(4), 2151–2180 (2022). DOI 10.3150/21-BEJ1413. URL https://doi.org/10.3150/21-BEJ1413
3. Belomestny, D., Gugushvili, S., Schauer, M., Spreij, P.: Weak solutions to gamma-driven stochastic differential equations. Indagationes Mathematicae **34**(4), 820–829 (2023). DOI 10.1016/j.indag.2023.03.004. URL https://doi.org/10.1016/j.indag.2023.03.004

4. 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)
5. Chavez-Demoulin, V., Embrechts, P., Hofert, M.: An extreme value approach for modeling operational risk losses depending on covariates. Journal of Risk and Insurance **83**(3), 735–776 (2016)
6. 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)
7. Dikmen, O., Cemgil, A.T.: Gamma Markov random fields for audio source modeling. IEEE Trans. Audio, Speech, Language Process. **18**(3), 589–601 (2010)
8. Donoho, D.L., Johnstone, I.M.: Adapting to unknown smoothness via wavelet shrinkage. J. Amer. Statist. Assoc. **90**(432), 1200–1224 (1995)
9. Embrechts, P., Klüppelberg, C., Mikosch, T.: Modelling extremal events: for insurance and finance, vol. 33. Springer Science & Business Media (2013)
10. Fan, J., Gijbels, I.: Local polynomial modelling and its applications. Routledge (2018)
11. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B.: Bayesian data analysis, third edn. Texts in Statistical Science Series. CRC Press, Boca Raton, FL (2014)
12. Ghosal, S., Ghosh, J.K., van der Vaart, A.W.: Convergence rates of posterior distributions. Ann. Statist. **28**(2), 500–531 (2000). DOI 10.1214/aos/1016218228. URL https://doi.org/10.1214/aos/1016218228
13. 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)
14. Gugushvili, S., van der Meulen, F., Schauer, M., Spreij, P.: Fast and scalable non-parametric bayesian inference for poisson point processes. Researchers.One (2020). URL https://researchers.one/articles/19.06.00001v1
15. Gugushvili, S., van der Meulen, F., Schauer, M., Spreij, P.: Nonparametric Bayesian estimation of a Hölder continuous diffusion coefficient. Braz. J. Probab. Stat. **34**(3), 537–579 (2020)
16. Gugushvili, S., van der Meulen, F., Schauer, M., Spreij, P.: Nonparametric bayesian volatility learning under microstructure noise. Japanese Journal of Statistics and Data Science **online** (2022). URL https://doi.org/10.1007/s42081-022-00185-9
17. Gugushvili, S., van der Meulen, F., Schauer, M., Spreij, P.: Nonparametric Bayesian volatility estimation. In: J. de Gier, C.E. Praeger, T. Tao (eds.) 2017 MATRIX Annals, pp. 279–302. Springer International Publishing, Cham (2019). DOI 10.1007/978-3-030-04161-8_19
18. Henderson, S.G.: Estimation for nonhomogeneous Poisson processes from aggregated data. Oper. Res. Lett. **31**(5), 375–382 (2003). DOI 10.1016/S0167-6377(03)00027-0
19. Heston, S.L.: A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev. Financial Stud. **6**(2), 327–343 (1993)
20. Hoffmann, M.: Minimax estimation of the diffusion coefficient through irregular samplings. Statistics & probability letters **32**(1), 11–24 (1997)
21. Hong, L., Martin, R.: A review of bayesian asymptotics in general insurance applications. European Actuarial Journal **7**, 231–255 (2017)
22. Hong, L., Martin, R.: Dirichlet process mixture models for insurance loss data. Scandinavian Actuarial Journal **2018**(6), 545–554 (2018)
23. Jacod, J., Shiryaev, A.: Limit theorems for stochastic processes, vol. 288. Springer Science & Business Media (2013)
24. Jeffreys, H., Jeffreys, B., Swirles, B.: Methods of mathematical physics. Cambridge university press (1999)
25. Karr, A.F.: Point processes and their statistical inference. Probability: Pure and Applied, 2. New York-Basel: Marcel Dekker, Inc. (1986)
26. Kutoyants, Y.A.: Statistical inference for spatial Poisson processes, *Lecture Notes in Statistics*, vol. 134. Springer-Verlag, New York (1998). DOI 10.1007/978-1-4612-1706-0

27. Kyprianou, A.E.: Fluctuations of Lévy processes with applications. Introductory lectures, 2nd edn. Universitext. Springer, Heidelberg (2014)
28. Mancini, C., Mattiussi, V., Renò, R.: Spot volatility estimation using delta sequences. Finance and Stochastics **19**(2), 261–293 (2015)
29. McNeil, A.J.: Estimating the tails of loss severity distributions using extreme value theory. ASTIN Bulletin: The Journal of the IAA **27**(1), 117–137 (1997)
30. Peeling, P., Cemgil, A.T., Godsill, S.: Bayesian hierarchical models and inference for musical audio processing. In: 2008 3rd International Symposium on Wireless Pervasive Computing, pp. 278–282 (2008). DOI 10.1109/ISWPC.2008.4556214
31. Petris, G., Petrone, S., Campagnoli, P.: Dynamic linear models with R. Use R! Springer, New York (2009)
32. van der Ploeg, A.P.C.: Stochastic volatility and the pricing of financial derivatives. Ph.D. thesis, University of Amsterdam (2006)
33. Protter, P.E.: Stochastic integration and differential equations, *Applications of Mathematics (New York)*, vol. 21, 2nd edn. Springer-Verlag, Berlin (2004). Stochastic Modelling and Applied Probability
34. Resnick, S.I.: Discussion of the danish data on large fire insurance losses. ASTIN Bulletin: The Journal of the IAA **27**(1), 139–151 (1997)
35. Rosas, F.E., Candia-Rivera, D., Luppi, A.I., Guo, Y., Mediano, P.A.: Bayesian at heart: Towards autonomic outflow estimation via generative state-space modelling of heart rate dynamics. arXiv preprint arXiv:2303.04863 (2023)
36. Sato, K.i.: Lévy processes and infinitely divisible distributions, *Cambridge Studies in Advanced Mathematics*, vol. 68. Cambridge University Press, Cambridge (2013). Translated from the 1990 Japanese original, Revised edition of the 1999 English translation
37. Scricciolo, C.: On rates of convergence for bayesian density estimation. Scandinavian Journal of Statistics **34**(3), 626–642 (2007)
38. Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A.: Bayesian measures of model complexity and fit. J. R. Stat. Soc. Ser. B Stat. Methodol. **64**(4), 583–639 (2002)
39. 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