Nonparametric Bayesian volatility estimation for gamma-driven SDEs

MSTA-V 2021, Kiyv, 1–4 June 2021

Peter Spreij

joint with Denis Belomestny, Shota Gugushvili, Moritz Schauer

Universiteit van Amsterdam

Radboud Universiteit

Outline

Outline

The model and the statistical problem

Goal:
Bayesian nonparametric estimation of $\sigma > 0$ in the Lévy-driven SDE

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

Here L is a gamma process with $L_0 = 0$, Lévy measure ν admitting the Lévy density

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

where $\alpha, \beta > 0$.

Outline

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

The driving gamma process L has parameters $\alpha = 1$, $\beta = 1$.

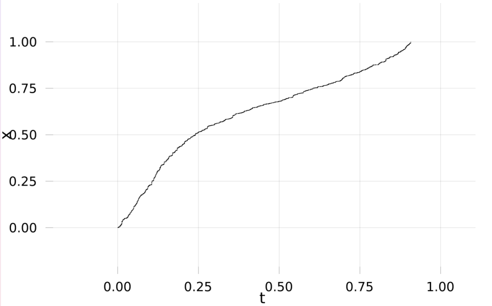

Figure 1: Sample path generated from (1) with an Euler scheme, until the hitting time of the level $b_K = 1$.

Existence of solutions

Well known is that Equation (1) has a unique strong solution if σ is Lipschitz, see Theorem V.6 in Protter (2004). Note that X is a Markov process. We also have

Proposition 1

Assume that $\sigma : [0, \infty) \to [0, \infty)$ is measurable, lower bounded by a constant $\sigma_0 > 0$, assume $K > 0$ such that for all $x \geq 0$ it holds that $\sigma(x) \leq K(1 + x)$. Then, on $[0, \infty)$, Equation (1) admits a weak solution that is unique in law.

Third characteristic

Assume that X is a strictly (weak) positive solution to (1) and denote by $\mathbb{P}_T^\sigma$ its law. Define

$$v^\sigma(t,x) = \frac{1}{\sigma(X_{t-})} v\left(\frac{x}{\sigma(X_{t-})}\right). \tag{3}$$

Lemma 2

Assume that (1) admits a weak solution for a given measurable function σ with $\sigma(X_{s-}) > 0$ a.s. for all $s \geq 0$. Under the measure $\mathbb{P}_T^\sigma$, the third characteristic of the semimartingale X, its compensated jump measure $\nu^\sigma(\mathrm{d}x, \mathrm{d}t)$, is given by $\nu^\sigma(\mathrm{d}x, \mathrm{d}t) = v^\sigma(t,x)\,\mathrm{d}x\mathrm{d}t$.

Likelihood ratio

Let

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

Proposition 3

It holds that $\mathrm{d}\mathbb{P}^\sigma_T \ll \mathrm{d}\mathbb{P}^1_T$. For $Z_T = \frac{\mathrm{d}\mathbb{P}^\sigma_T}{\mathrm{d}\mathbb{P}^1_T}$ one has at any finite (stopping) time $T > 0$ the explicit representation

$$Z_T = \exp\bigg(\int_0^T \int_0^\infty \log Y(t,x)\, \mu^X(\mathrm{d}x, \mathrm{d}t) \\ - \int_0^T \int_0^\infty (Y(t,x) - 1) v(x)\, \mathrm{d}x\, \mathrm{d}t \bigg) .$$

Outline

Earlier results on statistics for Lévy processes

- Jasra, Stephens, Doucet, & Tsagaris (2011),
- Jasra, Kamatani, & Masuda (2019),
- Uehara (2019),
- Gushchin, Pavlyukevich, & Ritsch (2019),
- Eguchi & Uehara (2020) and
- Todorov (2011).

Nonparametric Bayesian literature on inference for the model (1) is *non-existent* so far.

Related works I

- Belomestny, Gugushvili, Schauer, & Spreij (2019): nonparametric Bayesian approach to estimation of the Lévy measure forsubordinators,
- Gugushvili, van der Meulen, & Spreij (2015), Gugushvili, van der Meulen, & Spreij (2018) and Gugushvili, Mariucci, & van der Meulen (2020): compound Poisson processes.
- Nonparametric Bayesian volatility estimation in diffusion models in Batz et al.(2018)Batz (Ruttor) and Nickl & Söhl (2017), most related to the present one is Gugushvili, van der Meulen, Schauer, & Spreij (2019).
- Koskela, Spanò, & Jenkins (2019): frequentist consistency of a Bayesian approach to inference in jump-diffusion models.

Our contribution

- Nonparametric Bayesian volatility estimation for Lévy-driven SDEs,
- based on a *piecewise constant approximation* with a proper prior on the corresponding parameters.
- Concrete results on contraction of the posterior for Hölder continuous or piecewise contant volatility.

Example result with piecewise constant prior

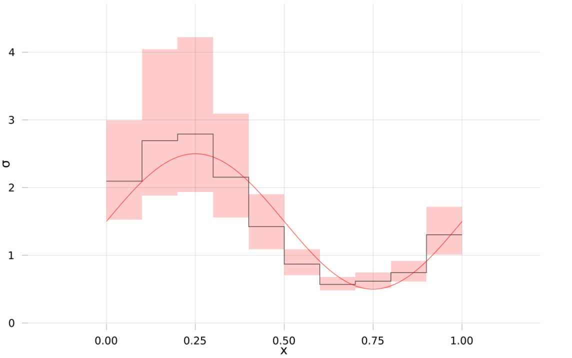

Figure 2: Red: true volatility σ_0 as a function of x. Shaded: marginal 90 %-posterior credible band for the piecewise constant posterior σ. Black: marginal posterior median.

Asymptotic setting

We consider the process X^n given as the solution to

$$\mathrm{d}X_t^n = \frac{1}{n}\sigma(X_{t-}^n)\,\mathrm{d}L_t. \tag{4}$$

The scaling factor $\frac{1}{n}$ causes for large values of n a 'slow growth' of the process X^n and 'long times' to reach certain levels.

X^n is observed on a long time interval $[0, T^n]$, where $T^n \to \infty$.

Later on: T^n grows roughly proportionally with n. Asymptotic results will be derived for $n \to \infty$.

Equivalent description, similarity with small noise

The $Y^n_t := X^n_{nt}$ satisfy

$$Y^n_t = \int_0^t \sigma(Y^n_{s-})\,\mathrm{d}L^n_s = \frac{\alpha}{\beta}\int_0^t \sigma(Y^n_{s-})\,\mathrm{d}s + M^n_t, \tag{5}$$

where $L^n_t = \frac{L_{nt}}{n}$. Here $L^n \rightsquigarrow L^\infty$ given by $L^\infty_t = \frac{\alpha}{\beta}t$ and $M^n \rightsquigarrow 0$.

Similar to a diffusion process Y^ε with small diffusion coefficient,

$$Y^\varepsilon_t = \int_0^t a(Y^\varepsilon_s)\,\mathrm{d}s + \varepsilon W_t,$$

which also has a deterministic limit as $\varepsilon \to 0$.

Prior model

We a priori model σ as a *piecewise constant* function:

$$\sigma(x) = \sum_{k=1}^{K} \xi_k \mathbf{1}_{B_k}(x) \tag{6}$$

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 $\{b_k\}$ and the bin number K.

The $\{\xi_k\}$ are positive numbers (later on positive random variables). Although we use (6) for our model, we emphasize that the 'true' σ *does not need to be piecewise constant*.

Likelihood specified

Corollary 4

Suppose that X^n is given by (4) with σ given by (6). Let $\tau_k^n = \inf\{t \geq 0 : X_t^n \geq b_k\}$, $k = 1, \ldots, K$, be the bin crossing times and $T = \tau_K^n$. Then

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

Piecewise constant volatility - towards the Posterior

Assume the *true* volatility function σ is piecewise constant. So,

$$\sigma(x) = \sum_k \sigma_k \mathbf{1}_{B_k}(x).$$

The σ_k are a priori independent random variables ξ_k with inverse gamma distributions

$$\xi_k \sim \mathrm{IG}(\alpha_k, \beta_k).$$

We use Corollary 4, with $\mathbb{P}^\sigma_T$ as the conditional law of X^n on $[0, T]$ given the ξ_k.

Posterior

Let $x_1, \ldots, x_K$ denote realisations of $\xi_1, \ldots \xi_K$. The posterior joint density of $(\xi_1, \ldots, \xi_K)$ is proportional to

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

It follows that the ξ_k are a posteriori independent, and

$$\xi_k \mid X^n \sim \mathrm{IG}(\alpha \Delta\tau^n_k + \alpha_k, n\beta \Delta X_{\tau^n_k} + \beta_k), \qquad (7)$$

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

Overshoot, a crucial lemma

The overshoot POT_k^n is defined as $\mathrm{POT}_k^n := X_{\tau_k^n}^n - b_k \geq 0$ for $k = 0, \ldots, K$. Note that $\mathrm{POT}_0^n = 0$. The inaccurracy of the overshoot at τ_k^n of the level b_k is controlled by

Lemma 5

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

$$\mathbb{P}(\mathrm{POT}_k^n < \delta) \geq \frac{b_K}{b_K + \delta}(1 - \exp(-n\delta\beta/\sigma_K^*)). \tag{8}$$

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

Rate for the τ_k^n and $\Delta\tau_k^n$

Proposition 6

Let $\Delta b_k = b_k - b_{k-1}$ and $\Delta\bar{\tau}_k^n = \frac{\Delta b_k \beta}{\alpha \sigma_k} n$.

For $c_n \to \infty$ such that $c_n n^{-\frac{1}{2}} \to 0$ it holds that

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

Hence, with high probability, $\Delta\tau_k^n \sim n$.

Posterior convergence rate

Lemma 7

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

Theorem 8

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

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

Hence, the posterior contraction rate for estimating σ_k is $n^{-1/2}$.

Assumptions and notation

- Hölder continuity of σ: $|\sigma(x) - \sigma(y)| \leq H|x-y|^\lambda$ and is bounded from below by $\underline{\sigma} > 0$.
- The number of bins and their width depend on n, $B_k^n = (b_{k-1}^n, b_k^n]$, $k = 1, \ldots, K$, $K = K_n$. The endpoint of the last, bin b_K is fixed. The other bin boundaries are $b_k^n = \frac{b_K}{K_n} k$, $k = 1, \ldots, K_n$. $\Delta b_k^n = b_k^n - b_{k-1}^n$.
- Note that for $x \in B_k^n$ it holds that $|\sigma(x) - \sigma_k^n| \leq H(\Delta b_k^n)^\lambda$ for $\sigma_k^n \in \{\sigma(b_{k-1}^n), \sigma(b_k^n)\}$.
- The number of bins $K = K_n \asymp n^\kappa$ for $0 < \kappa < 1$.

Model and observations

The process X^n is observed until it crosses the last bin.

Although σ is *continuous*, we *model* it in our Bayesian approach as *piecewise constant*, that is, as

$$\xi^n(x) = \sum_{k=0}^{K} \xi_k \mathbf{1}_{B_k^n}(x), \tag{9}$$

where the *independent* ξ_k are assigned the inverse gamma prior distributions as before.

Controlling the overshoot

We need a variation on Lemma 5. Let $\mathrm{POT}_k^n = X^n_{\tau^n_k} - b^n_k$.

Lemma 9

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

$$\mathbb{P}(\mathrm{POT}_k^n < \delta) \geq \frac{b_K}{b_K + \delta}(1 - \exp(-n\delta\beta/\sigma^*)). \tag{10}$$

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

The behaviour of the $\Delta\tau_k^n$

Recall the condition on the bin width, $\Delta b_k^n \asymp n^{-\kappa}$, and require

$$\kappa \geq \frac{1}{2\lambda + 1}. \tag{11}$$

Condition on the overshoot level: $\delta_n \asymp n^{-\delta}$, and require

$$\frac{1+\kappa}{2} \leq \delta < 1. \tag{12}$$

Proposition 10

Let $x \in (0, b_K)$ and $x \in B_k^n$, for $k = k_n(x)$. Let $\Delta\bar{\tau}_k^n(x) = \frac{\Delta b_k^n \beta}{\alpha\sigma(x)} n$. Under conditions (11) and (12) and for $c_n \to \infty$ such that $c_n n^{-\frac{1}{2}(1-\kappa)} \to 0$ it holds that

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

Contraction rate for Hölder continuous σ

Lemma 11

Assume $\Delta b_k^n \asymp n^{-\kappa}$, condition (11) and $c_n n^{-\frac{1}{2}(1-\kappa)} \to 0$. Then

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

with probability tending to one.

Theorem 12

Assume the *model* with piecewise constant volatility (9) whereas the true $x \mapsto \sigma(x)$ is Hölder continuous of order $\lambda \leq 1$ and bounded from below. Let the Δb_k^n be proportional to $n^{-\frac{1}{2\lambda+1}}$, and let $m_n \to \infty$ (arbitrarily slow). Then, for $n \to \infty$

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

The right contraction rate, lower bound

Let $\Sigma(\lambda, L)$ denote the Hölder class of functions f on $[0,1]$ satisfying $|f(x) - f(y)| \leq L|x-y|^\lambda, \quad x, y \in [0,1]$. Denote by $\mathbb{P}_T^{\sigma,n}$ the law of the process $(X_t^n)_{t\in[0,T]}$.

Proposition 13

There are constants $c_0, c_1 > 0$ not depending on n such that

$$\liminf_{n\to\infty} \inf_{\widehat{\sigma}_n} \sup_{\sigma\in\Sigma(\lambda,L)} \mathbb{P}_T^{\sigma,n}\Big(\sup_{x\in[0,1]} |\sigma(x) - \widehat{\sigma}_n(x)| \geq c_0 n^{-\lambda/(1+2\lambda)}\Big) \geq c_1,$$

where the infimum is taken over all estimators $\widehat{\sigma}_n$.

Inclusion of drift, modification of the original SDE

$$X_t = \int_0^t a(X_s)\,\mathrm{d}s + \int_0^t \sigma(X_{s-})\,\mathrm{d}L_s. \tag{13}$$

Consequence: the laws of X on $[0, T]$ under presence and absence of a drift are automatically mutually singular.

Implication: the drift part, $\int_0^t a(X_s)\,\mathrm{d}s$, can be identified with probability one.

For estimating σ, one can apply a modification of our procedure.

Example: If next to σ also a is piecewise constant on bins, with values a_k, in Proposition 6 one takes $\Delta\bar{\tau}_k^n$ as $\frac{n\Delta b_k \beta}{\alpha\sigma_k + \beta a_k}$.

Outline

Ice core data

Ice core data may be modelled as a diffusion.

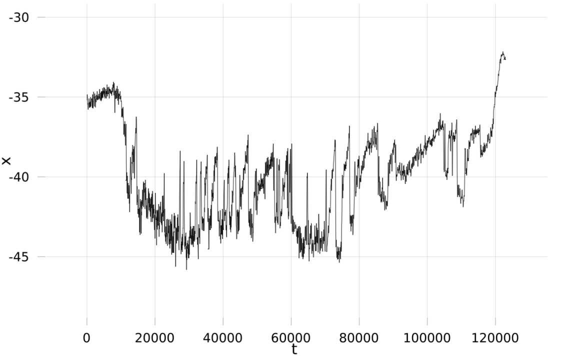

Figure 3: NGRIP oxygen isotope $\delta_{18}\mathrm{O}$ measurements against time (years BP 2000).

Ice core data, quadratic variation

Realised quadratic variation process q of a process Y satisfies $\Delta q_t := q_{t+\Delta t} - q_t = (\Delta y_t)^2 = (y_{t+\Delta t} - y_t)^2$, $q_0 = 0$, where y stands for a realization of Y, $t = 0, \ldots, n\Delta t$.

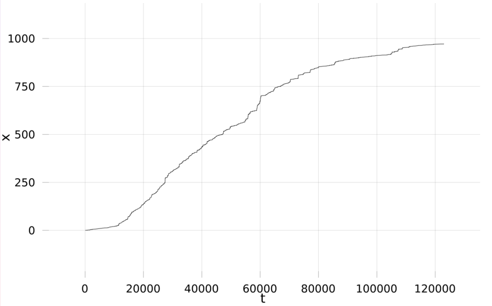

Figure 4: Realized quadratic variation process for NGRIP oxygen isotope $\delta_{18}\mathrm{O}$ measurements against time (years BP 2000).

Volatility estimation

For $\mathrm{d}Y_t = b_t\,\mathrm{d}t + \sigma_t \mathrm{d}W_t$ one has $(\Delta Y_t)^2 \approx \text{Gamma}(\frac{1}{2}, \frac{1}{2\Delta t \sigma_t^2})$.

Compare to the Lévy-driven SDE

$$\mathrm{d}X_t = c\varsigma_t^2\,\mathrm{d}L_t, \tag{14}$$

where L is a gamma process with parameters α, β.

With $\alpha = \frac{1}{2\Delta t}$ and $\beta = \frac{c}{2\Delta t}$, the conditional distributions of ΔX_t and $(\Delta Y_t)^2$ are approximately gamma with the same parameters, for any choice of $c > 0$. We used $c = n\Delta t / q_{n\Delta t}$, which implies $\mathbb{E}L_{n\Delta t} = q_{n\Delta t}$.

This is used to model volatility with gamma driven SDEs.

Results

We set $K = 20$, a good compromise in terms of bias-variance trade-off. For the prior we took weakly informative parameters $\alpha_k = \beta_k \equiv 0.1$. Figure 5 gives the marginal posterior band for σ.

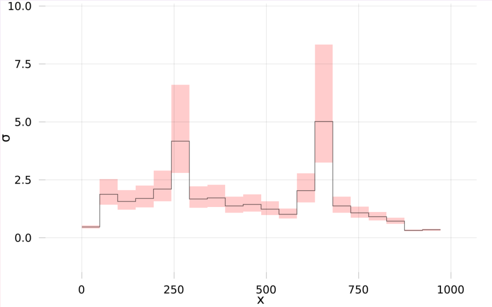

Figure 5: Shaded: marginal 90 %-posterior credible band for the piecewise constant posterior for the scale of the quadratic variation process of the ice core time series. Black: marginal posterior median.

Outline

References I

Batz, P., Ruttor, A., & Opper, M. (2018). Approximate Bayes learning of stochastic differential equations. *Phys. Rev. E*, *98*, 022109.

Belomestny, D., Gugushvili, S., Schauer, M., & Spreij, P. (2019). Nonparametric Bayesian inference for Gamma-type Lévy subordinators. *Commun. Math. Sci.*, *17*(3), 781–816.

Eguchi, S., & Uehara, Y. (2020). Schwartz type model selection for ergodic stochastic differential equation models. *ArXiv*, 1904.12398.

Gugushvili, S., Mariucci, E., & van der Meulen, F. (2020). Decompounding discrete distributions: A non-parametric Bayesian approach. *Scand J. Statist.*, *47*(2), 464–492.

Gugushvili, S., van der Meulen, F., Schauer, M., & Spreij, P. (2019). Nonparametric Bayesian volatility estimation. In J. de Gier, C. E. Praeger, & T. Tao (Eds.) *2017 MATRIX Annals*, (pp. 279–302). Cham: Springer International Publishing.

References II

Gugushvili, S., van der Meulen, F., & Spreij, P. (2015). Nonparametric Bayesian inference for multidimensional compound Poisson processes. *Mod. Stoch. Theory Appl.*, *2*(1), 1–15.

Gugushvili, S., van der Meulen, F., & Spreij, P. (2018). A non-parametric Bayesian approach to decompounding from high frequency data. *Stat. Inference Stoch. Process.*, *21*(1), 53–79.

Gushchin, A., Pavlyukevich, I., & Ritsch, M. (2019). Drift estimation for a Lévy-driven Ornstein-Uhlenbeck process with heavy tails. *ArXiv*, 1911.11202.

Jasra, A., Kamatani, K., & Masuda, H. (2019). Bayesian inference for stable Lévy–driven stochastic differential equations with high-frequency data. *Scandinavian Journal of Statistics*, *46*(2), 545–574.

Jasra, A., Stephens, D. A., Doucet, A., & Tsagaris, T. (2011). Inference for Lévy-driven stochastic volatility models via adaptive sequential Monte Carlo. *Scandinavian Journal of Statistics*, *38*(1), 1–22.

References III

Koskela, J., Spanò, D., & Jenkins, P. A. (2019). Consistency of Bayesian nonparametric inference for discretely observed jump diffusions. *Bernoulli*, *25*(3), 2183–2205.

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

Todorov, V. (2011). Econometric analysis of jump-driven stochastic volatility models. *Journal of Econometrics*, *160*(1), 12–21.

Uehara, Y. (2019). Statistical inference for misspecified ergodic Lévy driven stochastic differential equation models. *Stochastic Processes and their Applications*, *129*(10), 4051–4081.

This presentation: Denis Belomestny, Shota Gugushvili, Moritz Schauer, Peter Spreij (2020). Nonparametric Bayesian volatility estimation for gamma-driven stochastic differential equations. *Bernoulli, to appear (eventually), now under revision*; *arXiv:2011.08321*.

Thank you for your attention!