

# Nonparametric Bayesian volatility learning under microstructure noise

---

Shota Gugushvili<sup>♥</sup>, Frank van der Meulen<sup>♦</sup>, Moritz Schauer<sup>♣</sup> and Peter Spreij<sup>♣</sup>

OICA Conference, 28–29 April 2019

<sup>♥</sup>Wageningen University, <sup>♦</sup>Delft University of Technology, <sup>♣</sup>University of Gothenburg and  
<sup>♣</sup>University of Amsterdam

# Team



Problem formulation

Learning procedure

Numerical experiment

Application to real data

Summary

# Problem formulation

---

# Problem

- **Model:** Stochastic differential equation (SDE)

$$dX_t = b(t, X_t)dt + s(t)dW_t, \quad X_0 = x_0, \quad t \in [0, T].$$

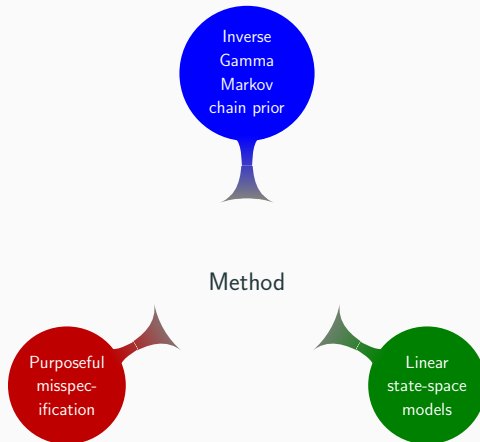
- **Task:** Learn the *volatility function*  $s$  from noisy observations on  $X$  at times  $0 < t_1 < \dots < t_n = T$ :

$$Y_i = X_{t_i} + V_i.$$

The  $V$ 's are unobserved stochastic disturbances.

- Important problem in *finance*; data subject to *microstructure noise* and not on a *uniform* time grid.

# Sources of inspiration



- *Minimalistic* in its assumptions on the volatility function, which in particular can be a stochastic process.
- Intuitive to understand. Ingredients are *well-known techniques*.
- Posterior inference via Gibbs sampler.

## Learning procedure

---



# Main challenge

- *Closed form expression* for the posterior distribution not available; *numerically intractable likelihood*.
- **Remedy:** *misspecification on purpose*.
- Alternative: Data augmentation device and some *intricate form* of an MCMC sampler.

## Misspecification on purpose

- Model misspecification not necessarily bad, if you know what you are doing; cf. GUGUSHVILI ET AL. 'XX and MARTIN ET AL. '18.
- *Misspecify the model purposely* and act as if the drift  $b = 0$ . Ultimate justification: Girsanov's theorem. Also pretend  $s$  is *deterministic*.
- Then

$$X_{t_i} = X_{t_{i-1}} + U_i.$$

Here

$$U_i = \int_{t_{i-1}}^{t_i} s(t) dW_t \sim N(0, w_i),$$

and  $w_i = \int_{t_{i-1}}^{t_i} s^2(t) dt$ . Note that  $\{U_i\}$  are independent.

# Gaussian linear state-space model

- Simplify the notation:  $x_i = X_{t_i}$ ,  $y_i = Y_i$ ,  $u_i = U_i$ ,  $v_i = V_i$ . Then

$$x_i = x_{i-1} + u_i,$$

$$y_i = x_i + v_i.$$

This is a *linear state-space model*.

- Assume that  $\{v_i\} \stackrel{iid}{\sim} N(0, \eta_v)$  are independent of the Wiener process  $W$ , so that  $\{v_i\}$  are independent of  $\{u_i\}$ .
- The model becomes *Gaussian*. This is *very* convenient.
- Notation:

$$\mathcal{X}_n = \{x_i\}, \quad \mathcal{Y}_n = \{y_i\}.$$

# Histogram prior for volatility

- Define  $N$  bins

$$B_k = [t_{m(k-1)}, t_{mk}), \quad k = 1, \dots, N.$$

where  $n = mN$ .

- Assume piecewise constant (squared) volatility,

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

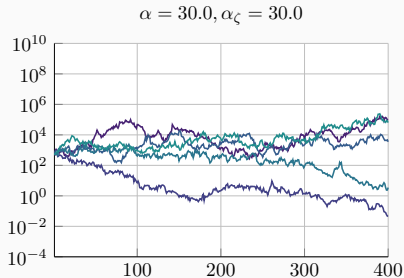
- Assume  $\{\theta_k\}$  form an inverse Gamma Markov chain. . .

- Fix hyperparameters  $\alpha_1, \alpha$  and define a Markov chain on  $\theta_1, \zeta_2, \theta_2, \dots, \zeta_k, \theta_k, \dots, \zeta_N, \theta_N$  with (latent) auxiliary variables  $\zeta_k, k = 2, \dots, N$ , and transition densities specified via

$$\begin{aligned}\theta_1 &\sim \text{IG}(\alpha_1, \alpha_1), \\ \zeta_{k+1} \mid \theta_k &\sim \text{IG}(\alpha, \alpha\theta_k^{-1}), \\ \theta_{k+1} \mid \zeta_{k+1} &\sim \text{IG}(\alpha, \alpha\zeta_{k+1}^{-1}).\end{aligned}$$

- *Smoothing/regularisation* through the prior-induced positive correlation of the volatility function across different bins.

# Realisations of volatility under the prior



**Figure 1:** Realisations of  $\{\theta_k\}$  with  $\alpha = 30$  for  $N = 400$ .

- Drift  $b \equiv 0$ .
- IGMC for  $\{\theta_i\}$ .
- Diffuse prior for  $\alpha$ ; improper for  $\alpha_1$ .
- Noise variance  $\eta_v \sim \text{IG}$ .

- Sequential updates:

$$\zeta_2, \dots, \zeta_N \mid \mathcal{X}_n, \alpha, \theta_1, \dots, \theta_N \quad (\text{conjugate})$$

$$\theta_1, \dots, \theta_N \mid \mathcal{X}_n, \alpha, \zeta_2, \dots, \zeta_N \quad (\text{conjugate})$$

$$\alpha \mid \theta_1, \zeta_2, \dots, \zeta_N, \theta_N \quad (\text{generic MH})$$

$$\mathcal{X}_n \mid \mathcal{Y}_n, \eta_v, \theta_1, \dots, \theta_N \quad (\text{Kalman-based})$$

$$\eta_v \mid \mathcal{V}_n \quad (\text{conjugate})$$



# Numerical experiment

---

- Widely used stochastic volatility model (HESTON '93).
- Price process  $S$  evolving according to the SDE

$$dS_t = \mu S_t dt + \sqrt{Z_t} S_t dW_t,$$

where the process  $Z$  follows the CIR (COX ET AL. '85) or square root process,

$$dZ_t = \kappa(\theta - Z_t)dt + \sigma\sqrt{Z_t}dB_t.$$

$W$  and  $B$  are correlated Wiener processes with correlation  $\rho$ .

- By Itô's formula, logarithm  $X_t = \log S_t$  obeys the diffusion equation with volatility

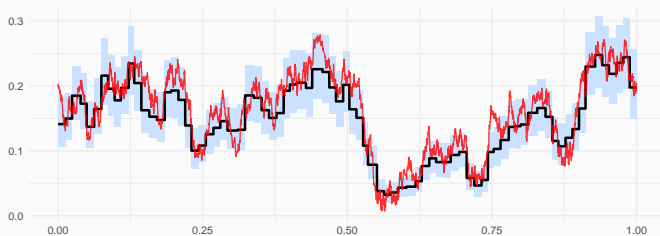
$$s(t) = \sqrt{Z_t}.$$

- In the Heston model the volatility function  $s$  is random, with its law not known in closed form.

- High frequency observations on the log-price process  $X$  with additive noise  $V_i \sim N(0, \eta)$ . No further knowledge of the data generation mechanism.
- Simulated path using “realistic” values (cf. Table 1 in HESTON '93)

$$\mu = 0.05, \quad \kappa = 7, \quad \theta = 0.04, \quad \sigma = 0.6, \quad \rho = -0.6.$$

- Noise variance  $\eta_v = 10^{-6}$  yielding a meaningful signal-to-noise ratio.



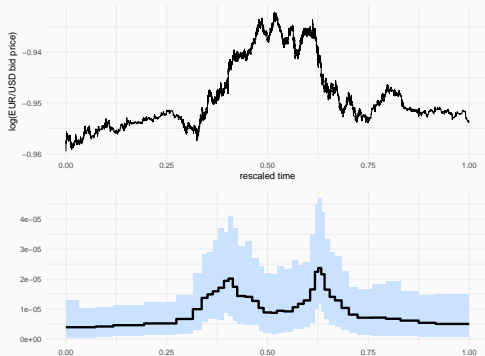
**Figure 2:** Posterior mean and pointwise 95% credible band for  $N = 80$  bins. True volatility function plotted in red; the black step function is the posterior mean.

## **Application to real data**

---

- Volatility learning from the high frequency exchange rate data.
- EUR/USD tick data (bid prices) for 2 March 2015, subsampled by retaining every 10th observation. A total of  $n = 13\,025$  observations (subsampling accounts for the assumption of independent additive measurement noise).
- Log-transform the observed time series and assume the additive measurement error model.

## Exchange rate data ii



**Figure 3:** Top: Logarithm of the EUR/USD exchange rate data for 2 March 2015. Bottom: Posterior mean (black curve) and pointwise 95% credible band (blue band) for the volatility function.









# Summary



---

# Key points

- A conceptually straightforward, clever nonparametric Bayesian method for volatility learning under microstructure noise.
- Available as **julia** package; see SCHAUER AND GUGUSHVILI '20.
- Partial fulfillment of prediction in GODSILL ET AL. '07 that some ideas developed originally in the context of audio and music processing “*will also find use in other areas of science and engineering, such as financial or biomedical data analysis*” .

-  A.T. Cemgil and O. Dikmen. Conjugate gamma Markov random fields for modelling nonstationary sources. *7th International Conference on Independent Component Analysis and Signal Separation (ICA 2007)*, pages 697–705.
-  J. C. Cox, J. E. Ingersoll and S. A. Ross (1995). A theory of the term structure of interest rates. *Econometrica*, 53:385–407.
-  S. L. Heston (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financial Stud.*, 6:327–343.

-  S. J. Godsill, A. T. Cemgil, C. Févotte and P. J. Wolfe (2007). Bayesian computational methods for sparse audio and music processing, *15th European Signal Processing Conference (EURASIP)*, pages 345–349.
-  S. Gugushvili, F. van der Meulen, M. Schauer and P. Spreij (20xx). Nonparametric Bayesian estimation of a Hölder continuous diffusion coefficient. To appear in *Braz. J. Probab. Stat.*
-  S. Gugushvili, F. van der Meulen, M. Schauer and P. Spreij (2018). Nonparametric Bayesian volatility learning under microstructure noise. arXiv:1805.05606 [stat.ME].

-  R. Martin, C. Ouyang and F. Domagni (2018). 'Purposely misspecified' posterior inference on the volatility of a jump diffusion process. *Stat. Probab. Lett.*, 134:106–113.
-  M. Schauer and S. Gugushvili. (2020). *MicrostructureNoise.jl*: v0.11.0. *Zenodo*. DOI: 10.5281/zenodo.1241010.