HEC PARIS

EXECUTIVE SUMMARY

Calibrating a Singular Local Stochastic Volatility Model using Gaussian Process Regression

Davi Méaille

Keywords: Volatility Surface Calibration, Local Stochastic Volatility Model, McKean-Vlasov Processes, Gaussian Process Regression

We are calibrating the local stochastic volatility model described in Bayer et al. [2024] using Gaussian Process Regressions (GPR) instead of Ridge Regression. The advantages of local stochastic volatility models over other models are the following:

- Local volatility models work well to reproduce current market prices, however they are very volatile when pricing exotic options
- Stochastic volatility models are useful in reproducing realistic dynamics like forward volatility, however they are hard to calibrate and depend on a lot of parameters.
- Jump diffusion models are useful in calibrating actual volatility surfaces, however they lack completeness of the model (as pointed out in Dupire et al. [1994]) since there are no liquid enough instruments to hedge jumps.

Local stochastic volatility models have the potential to be easier to calibrate (under certain calibration methods), while still exhibiting realistic dynamics thanks to their stochastic volatility component.

1 Replicating Bayer et al. [2024] with GPR

In particular, assuming $r = 0$ throughout the paper, the model we will study is as follows:

$$\frac{dX_t}{X_t} = \sqrt{\nu_t}\sigma(X_t, t)dW_t \tag{1}$$

with a stochastic volatility component $\sqrt{\nu_t}$ and a local volatility component $\sigma(X_t, t)$. Using the Dupire's local volatility formula for the local volatility component:

$$\sigma^2_{\text{Dupire}}(x, t) = \sigma^2(x, t)\mathbb{E}\left[\nu_t | X_t = x\right] \tag{2}$$

we end up with:

$$dX_t = \sigma_{\text{Dupire}}(X_t, t)X_t\frac{\sqrt{\nu_t}}{\sqrt{\mathbb{E}[\nu_t|X_t]}}dW_t \tag{3}$$

The particularity of this diffusion model comes from the term $\mathbb{E}[\nu_t|X_t]$, which justifies the use of machine learning techniques. Due to the presence of this common term, shared by all the particles and dependent on the distribution of the whole sample, this diffusion is called a McKean-Vlasov distribution.

As pointed out in the original paper Bayer et al. [2024], coupled distributions of the form:

$$dX_t = H\left(t, X_t, Y_t, \mathbb{E}[A_1(Y_t)|X_t]\right)dt + F\left(t, X_t, Y_t, \mathbb{E}[A_2(Y_t)|X_t]\right)dW_t^X \tag{4}$$

$$dY_t = b\left(t, Y_t\right)dt + \sigma\left(t, Y_t, \right)dW_t^Y \tag{5}$$

do admit a strong solution if: 1) $A_i \in \mathcal{C}^1\left(\mathbb{R}^d, \mathbb{R}\right)$; 2) H, F, b, σ are Lipschitz-continuous, 3) for any fixed $x, y, \in \mathbb{R}^d, z \in \mathbb{R}$, we have

$$\int_0^T \left(|H(t,x,y,z)|^2 + |F(t,x,y,z)|^2 + |b(t,y)|^2 + |\sigma(t,y)|^2\right)dt < \infty$$

4) $\mathbb{E}\left[\left|\widehat{X}_0\right|^2\right] + \mathbb{E}\left[|Y_0|^2\right] < \infty$.

However, the problem is, that for $\mu_t = \mathrm{Law}(X_t, Y_t)$, the following functional is not usually Lipschitz-continuous, even for A_i smooth enough:

$$(x, \mu_t) \to \mathbb{E}[A_i(Y_t)|X_t = x] \tag{6}$$

To circumvent this problem, they approximate the conditional expectation above using Reproducing Kernel Hilbert Spaces (RKHS) and Ridge regression. In particular, they replace the conditional expectation with the result from the following optimization problem:

$$m_A^\lambda(\cdot;\nu) := \arg\min_{f\in\mathcal{H}}\left(\int_{\mathcal{X}\times\mathcal{X}} |A(y) - f(x)|^2 \nu(dx, dy) + \lambda\|f\|_{\mathcal{H}}^2\right) \tag{7}$$

Instead of solving this equation, we resort to using Gaussian Process Regressions (GPR) (see Williams and Rasmussen [2006]). We also separate our data into training data $\{X_t, Y_t\}$ and predicted data for speed purposes. We thus compute the posterior distribution for the predicted points $\{Y_t^*\}$:

$$Y^*|X, Y, X^* \sim \mathcal{N}\left(K(X^*, X)\left[K(X,X) + \sigma_n^2 I_d\right]^{-1} Y, \Sigma\right) \tag{8}$$

with $\Sigma = K(X^*, X^*) - K(X^*, X)\left[K(X,X) + \sigma_n^2 I_d\right]^{-1} K(X, X^*)$, using the Matérn covariance kernel (see Matérn [2013]), defined as:

$$K_{\text{Matérn},\nu}(X, X') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}|X - X'|}{l}\right)^\nu K_\nu\left(\frac{\sqrt{2\nu}r}{l}\right) \tag{9}$$

with K_ν the modified Bessel function. We use the particular case where $\nu = 3/2$.

2 Confidence Interval Construction

Our second contribution consists in leveraging on the GPR results, to use the covariance and build confidence intervals around our predictions for the conditional expectation term. In what follows, we detail our computations.

Denoting $\mathcal{F}_t = \mathbb{E}[\nu_t | X_t]$, we rewrite the local stochastic volatility diffusion in its discrete time version, under its Euler scheme:

$$X_{t+1} = X_t + \sigma_{\text{Dupire}}(X_t, t) X_t \frac{\sqrt{\nu_t}}{\sqrt{\mathcal{F}_t}} B_t \tag{10}$$

where B_t is a random walk term.

Because $\mathcal{F}_t$ is the result of our GPR, we have that $\mathcal{F}_t \sim \mathcal{N}(\mu_t, \sigma_t)$, where μ is the mean prediction and Σ the uncertainty around μ. Let's denote $\mathcal{G}_t = \dfrac{1}{\sqrt{\mathcal{F}_t}}$, differentiable over $\mathbb{R}^{+*}$. Using the delta method on $\mathcal{G}_t$, since $\dfrac{\partial g(x)}{\partial x} = -\dfrac{1}{2} x^{-3/2}$ for $g(x) = \dfrac{1}{\sqrt{x}}$, we get that:

$$\mathcal{G}_t(\mathcal{F}_t) \sim \mathcal{N}\left(\mathcal{G}(\mu_t), \frac{\sigma^2}{4\mu^3}\right) \tag{11}$$

This allows us to compute some point-wise bounds, that is: the uncertainty delivered by our estimation at each step t and for each particle X_t^i. We get the following bounds for X_t^i:

$$\begin{cases} X_{t+1}^{up} = X_t + X_t \left(\dfrac{1}{\sqrt{\mu}} + 1.96 * \sqrt{\dfrac{\sigma^2}{4\mu^3}} \right) \\ X_{t+1}^{low} = X_t + X_t \left(\dfrac{1}{\sqrt{\mu}} - 1.96 * \sqrt{\dfrac{\sigma^2}{4\mu^3}} \right) \end{cases} \tag{12}$$

For more details on the algorithm implementation, please refer to the main paper of the thesis.

3 Results

We display below our results to several cases, first using a Black Scholes pricer, then a Heston pricer (thus following Bayer et al. [2024]).

We first start by comparing our results for the Black-Scholes and the Heston simulations. We can observe in Figure 1 that our volatility surface calibrations are more accurate for longer maturities in the Heston case. The limitations in the Black-Scholes case come from numerical limits due to the computations of the $N \times N$ Gram matrix in the GPR, limiting the number of particles we could run.

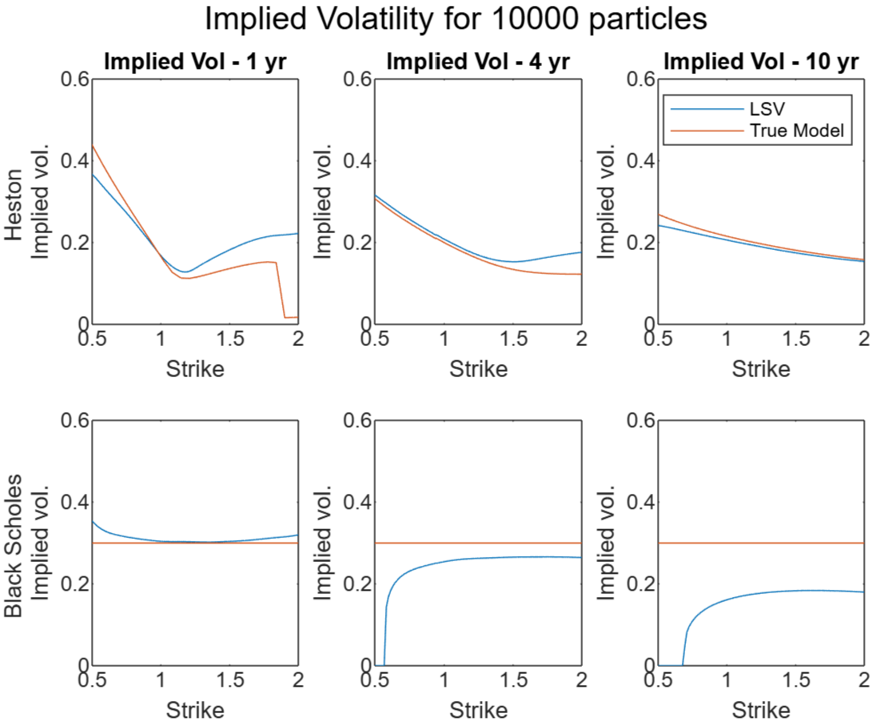

Figure 1: Comparison of performance to replicate the prices from Heston vs Black-Scholes

We can then observe in Figure 2 that we get very similar results to what Bayer et al. [2024] obtain. This is mainly due to the proximity between Ridge regression on Hilbert Spaces and GPR when it comes to the estimation of the mean. The observable differences come in large part because we replicated Bayer et al. [2024] faithfully, leaving a hyperparameter for the gaussian kernel fixed at $\sigma_f = 0.1$, whereas we are constantly optimizing when running the GPR regression.

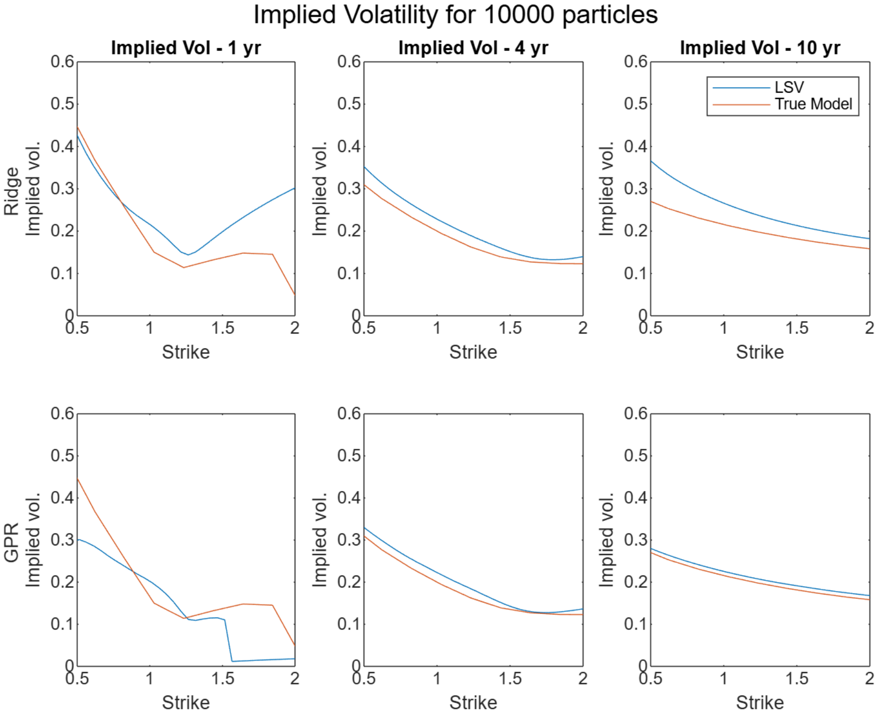

Figure 2: Comparison between Ridge and GPR methods for Smile calibration

Additionally, we observe that we manage to get some point wise upper and lower confidence bounds for the volatility surface, as in Figure 3. This is an important improvement on top of what Bayer et al. [2024] manage to obtain. It is also interesting to note that there is an apparent systematic upward bias in the estimations of the smile. This is only due to the too small number of particles used to simulate. Indeed, when we use up to 100,000 particles in the Ridge scheme, we obtain very close results. The advantage of Bayer et al. [2024] being that they compute no larger a Gram matrix than $100 \times 100,000$, we managed to reproduce these results. In our GPR case, we didn't implement optimization method such as Nyström approximation or Subset selections and thus our computational capabilities were limited by the large Gram matrix of size $N \times N$, with N the number of particles. We restricted ourselves to only computing up to 10,000 particles.

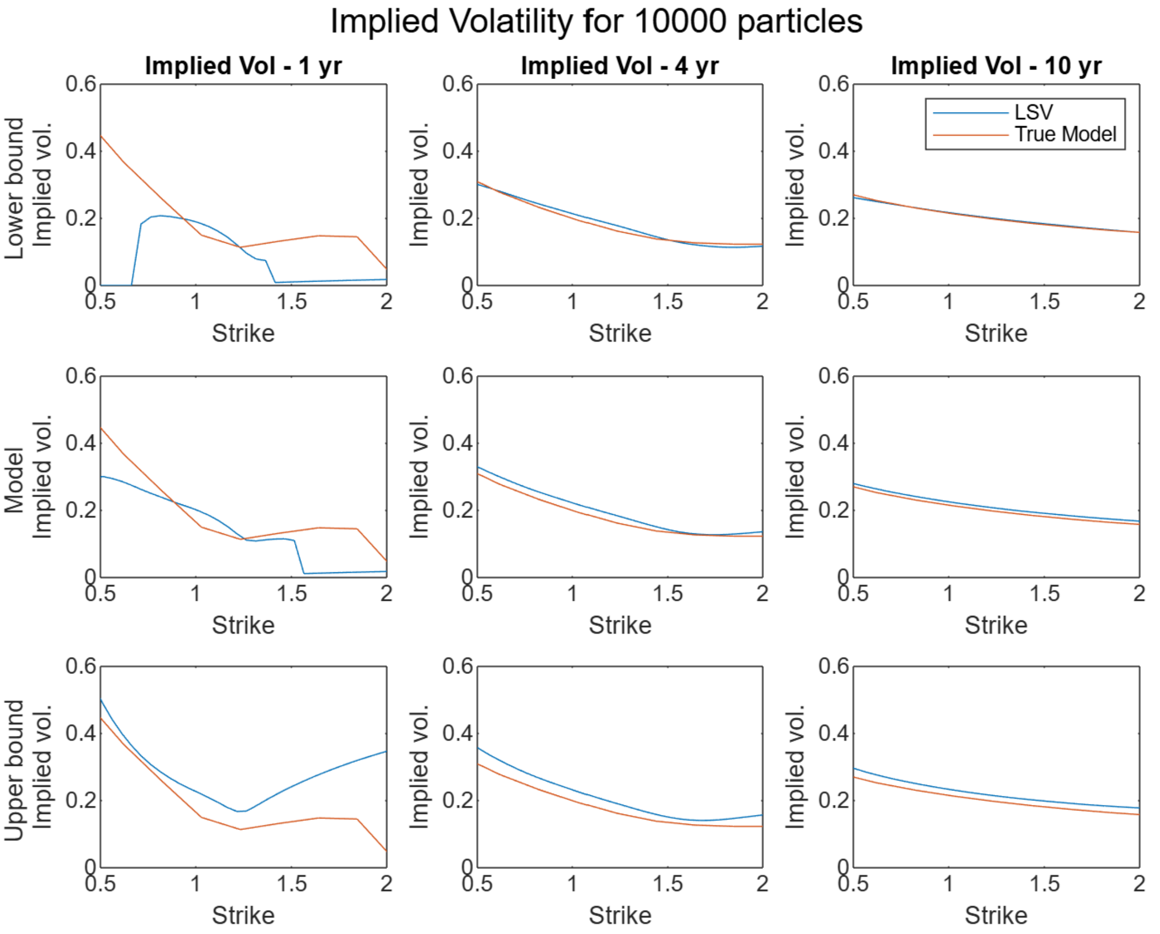

Figure 3: Point wise Lower and Upper bounds with GPR

4 Conclusion

GPR are a very versatile and useful techniques in machine learning, with interesting applications in finance. We have developed its use in the context of a Local Stochastic Volatility model to estimate a conditional expectation term. This conditional expectation terms renders the corresponding McKean-Vlasov diffusion analytically singular, but using kernel methods such as Ridge in Bayer et al. [2024] or GPR in our current paper allows to circumvent this difficulty.

We first showed that we managed to reproduce the original authors' results using GPR instead of Ridge regression. This is standard, as both method are equivalent when considering Reproducing Kernel Hilbert Space (Kanagawa et al. [2018]). We get better results with our algorithms are we reoptimize the hyperparameters at each steps of the diffusion, unlike the original paper.

We then moved on to show that, utilizing the covariance predicted by the GPR, we can build upper and lower bounds for the estimated volatility surface. This can be very useful, in contexts where the uncertainty is larger, for instance with smaller datasets (calibrating on market prices).

5 Supplementary Material

See my github: https://github.com/dmeaille

References

Christian Bayer, Denis Belomestny, Oleg Butkovsky, and John Schoenmakers. A reproducing kernel hilbert space approach to singular local stochastic volatility mckean–vlasov models. *Finance and Stochastics*, 28(4):1147–1178, 2024.

Bruno Dupire et al. Pricing with a smile. *Risk*, 7(1):18–20, 1994.

Motonobu Kanagawa, Philipp Hennig, Dino Sejdinovic, and Bharath K Sriperumbudur. Gaussian processes and kernel methods: A review on connections and equivalences. *arXiv preprint arXiv:1807.02582*, 2018.

Bertil Matérn. *Spatial variation*, volume 36. Springer Science & Business Media, 2013.

Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.