1.3 RKHS Regularization of Singular Local Stochastic Volatility McKean–Vlasov Models

Christan Bayer, Oleg Butkovsky, and John Schoenmakers

Introduction

In financial markets, including stock and energy markets, options on future prices of stocks, interest rates, and different kinds of energy are very popular and intensively traded. In this note, we focus on call or put options in stock markets. Suppose S_t is the price of a stock at time $t \geq 0$. The dynamics of S is generally considered as a random process that satisfies so-called *no-arbitrage conditions*. Then, for example, a standard call option with strike price K and maturity time T is the option to receive $S_T - K$ if $S_T \geq K$, and 0, or nothing, if $S_T < K$. By standard financial theory, the value of the option at time $t = 0$, i.e., today, equals

$$C(T, K) = \mathbb{E}\left[\max\left(S_T - K, 0\right)\right], \quad T \geq 0, \quad K \geq 0, \tag{1}$$

under the simplifying assumption that interest rates are zero.

The call prices $C(T, K)$ are quoted for various T and K on the stock market. We here assume the idealized situation where $C(T, K)$ is available for all $0 \leq T \leq T_{\max}$ and all $K \geq 0$. Now, the crucial problem, also called *the calibration problem*, is to find a suitable model for S such that (1) holds, at least approximately but with high accuracy, for all $0 \leq T \leq T_{\max}$ and all $0 \leq K \leq K_{\max}$. One of the most striking results of the late nineties in finance was that (1) can be achieved exactly with a so-called *local volatility model*,

$$dS_t = S_t \sigma_{\text{Dupire}}(t, S_t)\, dW_t \tag{2}$$

with W being a standard Brownian motion, provided that the function $(T, K) \to C(T, K)$ is sufficiently smooth. In particular, σ_{Dupire} in (2) is given by *Dupire's formula,*

$$\sigma_{\text{Dupire}}(T, K) = \frac{1}{K}\sqrt{\frac{2\partial_T C(T, K)}{\partial_{KK} C(T, K)}}.$$

Unfortunately, while Dupire's model (2) is a model that solves (1), i.e., perfectly fits to market prices of call options, it is well understood that it exhibits unrealistic random price behavior. On the other hand, *stochastic volatility models*

$$dS_t = S_t \sqrt{v_t}\, dW_t,$$

for a suitably chosen stochastic variance process v_t, may lead to realistic (in particular, time-homogeneous) dynamics, but are typically difficult or impossible to calibrate to the observed market prices (1).

As a natural way out, one may combine a local volatility model with a stochastic volatility model to a *local stochastic volatility model* that combines the advantages of both local and stochastic volatility models. Indeed, if the stock price is given by

$$dS_t = S_t \sigma(t, S_t)\sqrt{v_t}\, dW_t, \tag{3}$$

then it exactly fits the observed market option prices via (1), provided that

$$\sigma_{\mathsf{Dupire}}(t, x)^2 = \sigma(t, x)^2 \mathbb{E}[v_t \mid S_t = x]. \tag{4}$$

This is a consequence of the celebrated Markovian projection theorem by Gyöngy [4, Theorem 4.6]. By solving for σ in (4) and inserting the result in (3), we get the stochastic dynamics

$$dS_t = S_t \sigma_{\mathsf{Dupire}}(t, S_t)\sqrt{\frac{v_t}{\mathbb{E}[v_t \mid S_t]}}\, dW_t. \tag{5}$$

As recognized by Guyon and Henry-Labordère [3], in principle, the stock price dynamics (5) is the solution of a *McKean-Vlasov Stochastic Differential Equation* (MVSDE) due to the presence of the conditional expectation $\mathbb{E}[v_t \mid S_t]$, which can be regarded as a complicated functional of the joint distribution of (S_t, v_t). MVSDEs are usually solved by Monte Carlo simulation of a particle system, and so was done in [3]. In a nutshell, in [3], one simulates, for $i = 1, \ldots, N$, the "particles,"

$$dS_t^i = S_t^i \sigma_{\mathsf{Dupire}}(t, S_t^i)\sqrt{\frac{v_t^i}{\frac{\sum_{j=1}^N v_t^j \phi_\delta(S_t^j - S_t^i)}{\sum_{j=1}^N \phi_\delta(S_t^j - S_t^i)}}}\, dW_t^i, \quad i = 1, \ldots, N, \tag{6}$$

where ϕ_δ is some "delta-shaped" density function with band width $\delta > 0$. For example, $\phi_\delta(x) = \delta^{-1}\phi(x/\delta)$, where ϕ is the standard normal density. The conditional expectation in the MVSDE (5) is thus approximated by the blue expression in the particle system (6), a kind of weighted Nadaraya-Watson estimator, for some small (but not too small) δ.

Although the approach in [3] was well received in the financial community, there are two main issues:

(I) The MVSDE (5) is singular in the sense that it does not satisfy the standard regularity conditions in [2] that guarantee a unique solution. In particular, the general existence of a solution to (5) is still an open question.

(II) The choice of the band width in (6) is very delicate and, generally, the simulation of (6) is rather costly. A more efficient numerical approach is called for.

Both issues were addressed in our recent study [1], and below we will sketch the main lines.

Regularizing the MVSDE

In the MVSDE (5), the conditional expectation

$$m(x, \mu) := \mathbb{E}_{(S,v)\sim\mu}[v \mid S = x], \qquad x > 0, \tag{7}$$

is not Lipschitz continuous with respect to μ in the Wasserstein metric, and therefore the existence of a solution to (5) is not guaranteed by the usual conditions in [2]. As a remedy, we regularize (7) in the framework of a *reproducing kernel Hilbert space* (RKHS).

Let $\mathcal{H}$ be an $RKHS$ of functions on $\mathbb{R}_+$ (state space of the stock price model) that is generated by a positive definite symmetric kernel $k(\cdot,\cdot)$ on $\mathbb{R}_+ \times \mathbb{R}_+$. Assuming that $m(\cdot,\mu) \in \mathcal{H}$ for fixed μ, we may formally write

$$\begin{aligned} c^\mu(\cdot) &:= \int_{\mathbb{R}_+\times\mathbb{R}_+} k(\cdot,x) y \mu(dx,dy) \\ &= \int_{\mathbb{R}_+} k(\cdot,x)\mu(dx,\mathbb{R}_+) \int_{\mathbb{R}_+} y\mu(dy|x) \\ &= \int_{\mathbb{R}_+} m(x,\mu) k(\cdot,x)\mu(dx,\mathbb{R}_+) =: \mathcal{C}^\mu m(\cdot,\mu). \end{aligned}$$

That is, formally, $m(\cdot,\mu) = \left(\mathcal{C}^\mu\right)^{-1} c^\mu$. Unfortunately, the symmetric operator $\mathcal{C}^\mu$ is generally not invertible. As $\mathcal{C}^\mu$ is positive definite, it is, however, possible to *regularize* the inversion by replacing $\mathcal{C}^\mu$ by the invertible operator $\mathcal{C}^\mu + \lambda I_{\mathcal{H}}$ for (small) $\lambda > 0$. Furthermore, it turns out that

$$m^\lambda(\cdot,\mu) := \left(\mathcal{C}^\mu + \lambda I_{\mathcal{H}}\right)^{-1} c^\mu \tag{8}$$

is the solution to the minimization problem

$$m^\lambda(\cdot,\mu) = \arg\min_{f\in\mathcal{H}} \left(\mathbb{E}_{(S,v)\sim\mu}\left[(v - f(S))^2\right] + \lambda \|f\|^2_{\mathcal{H}}\right),$$

and therefore, since conditional expectation is an L_2 projection in fact, it is natural to expect that if λ is small enough and $\mathcal{H}$ is large enough, then $m^\lambda(\cdot,\mu)$ will be close to the true conditional expectation $m(\cdot,\mu)$.

Significantly, we were able to prove in [1] that the MVSDE, obtained by replacing in (5) the conditional expectation with its regularized version (8),

$$dS_t = S_t \sigma_{\text{Dupire}}(t,S_t)\sqrt{\frac{v_t}{m^\lambda(S_t,\mu_t)}}dW_t, \qquad \text{where} \qquad (S_t,v_t)\sim\mu_t, \tag{9}$$

is well posed in the sense of [2], and thus has a unique strong solution.

Particle system with ridge regression

In comparison to (6), the particle system corresponding to (9) now reads

$$dS_t^i = S_t^i \sigma_{\text{Dupire}}(t,S_t^i)\sqrt{\frac{v_t^i}{m^\lambda(S_t^i,\mu_t^N)}}dW_t^i, \quad i = 1,\dots,N, \tag{10}$$

$$\mu_t^N(dx,dy) = \frac{1}{N}\sum_{j=1}^N \delta_{S_t^j}(dx)\delta_{v_t^j}(dy),$$

where for any Borel measurable C, $\delta_x(C) := 1_C(x)$, i.e. δ_x denotes the Dirac measure at point x. In (10),

$$m^\lambda(\cdot,\mu_t^N) = \arg\min_{f\in\mathcal{H}}\left(\frac{1}{N}\sum_{j=1}^N\left(v_t^j - f(S_t^j)\right)^2 + \lambda\|f\|^2_{\mathcal{H}}\right) \tag{11}$$

is computed via a ridge regression procedure. It follows from the representer theorem for RKHS [5, Theorem 1] that the solution to (11) has a representation

$$m^{\lambda}(\cdot, \mu_t^N) = \sum_{i=1}^{N} \alpha_i k(\cdot, S_t^i), \tag{12}$$

for some $\alpha_i \in \mathbb{R}$, $i = 1, \dots, N$, which can be found by plugging the ansatz 12 into the minimization problem 11.

Numerical example

For any numerical implementation of the approach suggested above, the regularized MV problem (10) together with (11) needs to be discretized in time, for which we use a straight-forward Euler–Maruyama method. In this short note, we do not further consider time discretization, since no phenomena specific to our approximation method seem to occur.

A more important practical concern directly related to (12) is the high cost of solving the minimization problem (11): As the ansatz corresponds to a regression problem with N basis functions and N samples, the expected asymptotic cost is $O(N^2)$, which is prohibitively expensive given that the number of samples may be very large. Hence, instead of using all basis functions $k(\cdot, S_t^i)$, $i = 1, \dots, N$, we pick a representative ensemble $k(\cdot, Z_t^j)$, $j = 1, \dots, L \ll N$. The *representative samples* Z_t^j can be chosen in many different ways, for instance, as the jL/N th order statistic of $(S_t^i)_{i=1}^N$ or even as points on a deterministic grid.

In order to test our method, we chose a synthetic, but nonetheless realistic calibration problem. In order to obtain realistic option prices, we fix a particular stochastic volatility model as a "ground-truth model," in which option prices are computed. In the second step, we choose a different stochastic volatility model as our "backbone model," which needs to be calibrated to the option prices computed in the first step by adding a suitable local volatility factor.

In practice, both our ground-truth model and our backbone model are instances of the popular *Heston model*, i.e.,

$$\begin{aligned} dS_t &= \sqrt{v_t} S_t dW_t, \\ dv_t &= \kappa(\theta - v_t)dt + \xi \sqrt{v_t} dB_t, \end{aligned}$$

driven by standard Brownian motions W, B with correlation ρ. However, different sets of parameters are chosen for the models. The concrete parameter choices are reported in Table 1.

	κ	θ	ξ	ρ	v_0
market	2.19	0.17023	1.04	−0.83	0.0045
backbone	1.0	0.0144	0.5751	−0.9	0.0144

Table 1: *Parameter choices for the Heston models (market) used for generating option prices and the backbone model to be calibrated. Both sets of parameters are realistic in the sense that they are reported in the literature as calibrated Heston parameters.*

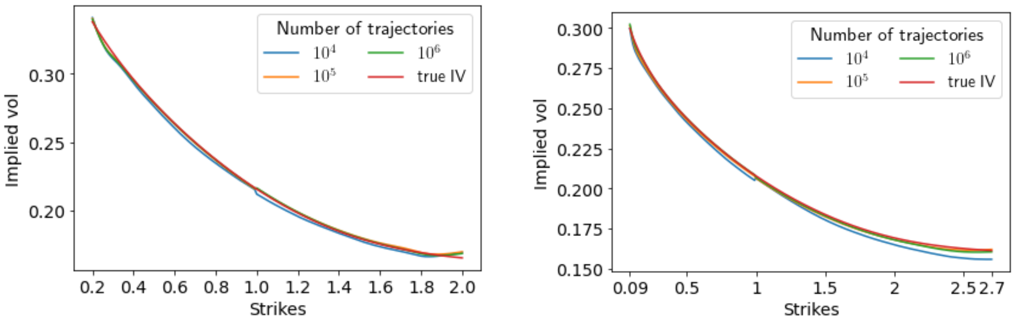

Fig. 1: *Implied volatilities for the calibrated Heston model for maturities $T = 4$ years (left) and $T = 10$ years (right)*

We consider the RKHS associated with the Gaussian kernel $k(x, y) := \exp(-|x - y|^2/(2\sigma^2))$ with $\sigma^2 = 5$. The representative samples Z_t^j are chosen as the $\frac{j \cdot 100}{L+1}$ percentiles of the empirical distribution of the samples $(S_t^i)_{i=1}^N$ based on the total number of samples $N \in \{10^4, 10^5, 10^6\}$. The weight λ for the ridge regression is chosen to be $\lambda = 10^{-5}$. In order to further mollify the singularity in (10), we truncate the denominator at $\varepsilon = 10^{-3}$, i.e., we choose

$$dS_t^i = S_t^i \sigma_{\text{Dupire}}(t, S_t^i) \frac{\sqrt{v_t^i}}{\sqrt{m^\lambda(S_t^i, \mu_t^N) \vee \varepsilon}} dW_t^i.$$

We plot implied volatility surfaces at different maturities for the calibrated Heston model by the described procedure in Figure 1. We note that the calibration is essentially exact for $N \geq 10^5$ samples.

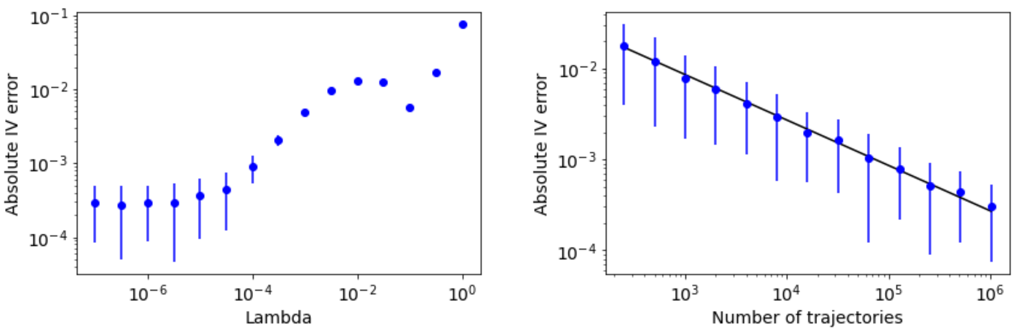

Fig. 2: *Error of the implied volatility smile for the calibrated Heston model as a function of λ (left) and as a function of the number of samples N (right)*

Figure 2 shows the dependence of the error in the calibrated Heston model (i.e., difference between the prices in the ground-truth model and the model (10)) in terms of the weight parameter λ in the ridge regression and the number of samples N. For the former, we note that the choice of λ is not important as long as it is small enough, which might indicate that the problem is actually more robust than suggested by the theoretical analysis. On the other hand, the results converge with rate $1/2$ in terms of N, faster than rate $1/4$ obtained in the error analysis.

Conclusions and outlook

We suggest a novel RKHS-based regularization method for the problem of calibrating local stochastic volatility models and prove that this regularization guarantees well-posedness of the underlying MVSDE. Our numerical results suggest that the proposed approach is quite efficient for the calibration of various local stochastic volatility models and can outperform widely used local kernel methods as in [3]. Nonetheless, it remains unclear whether the regularized MVSDE remains well posed when the regularization parameter λ tends to zero. This limiting case needs a subsequent study. Another important issue for future research is the choice of the RKHS kernel that ideally should be adapted to the problem at hand.

References

[1] C. Bayer, D. Belomestny, O. Butkovsky, J.G.M. Schoenmakers, *RKHS regularization of singular local stochastic volatility McKean–Vlasov models*, WIAS Preprint no. 2921, 2022.

[2] R. Carmona, F. Delarue, *Probabilistic Theory of Mean Field Games with Applications I+II*, Springer, vol. 83 of Probability Theory and Stochastic Modelling, 2016.

[3] J. Guyon, P. Henry-Labordère, *Being particular about calibration*, Risk, **25**:1 (2012), p. 88.

[4] I. Gyöngy, *Mimicking the one-dimensional marginal distributions of processes having an Itô differential*, Probab. Theory Related Fields, **71**:4 (1986), pp. 501–516.

[5] B. Schölkopf, R. Herbrich, A. Smola, *A Generalized Representer Theorem*, In: D. Helmbold, B. Williamson, eds., Computational Learning Theory, COLT 2001, vol 2111 of Lecture Notes in Computer Science, Springer, Berlin Heidelberg, pp. 416–426.