MATH70108 Statistical Methods for Finance Coursework Project

Harrison Lam
Charles Favell
Thomas Pink
Jonathan Guérin

January 9, 2025

Abstract. In this project, we analyse, apply and review the methodologies used in a paper published by Lyons, Nejad and Arribas on pricing exotic derivatives [1]. By applying theoretical results on the approximation of derivative payoffs using signature payoffs, the implied market signature is estimated using a restricted class of derivatives and their market prices. This is then applied to a wider class of derivatives to obtain price estimates.

1 Introduction

The objective of this paper is to estimate the implied market signature - an idea comparable to implied volatility - using current market prices, and ultimately to price path-dependent derivatives. This is done first by estimating the signature payoff for each derivative in the restricted class $\mathscr{F}$, which is a path-dependent function that approximates the payoff of the target derivative. We use these signature payoffs to estimate the implied market signature, which represents the expected characteristics of the price path that best matches these signature payoffs to the market prices, and is built upon the theory of tensor algebras. This implied market signature is then used to price other path-dependent derivatives in a wider class $\mathscr{G}$.

In Section 2, we develop the theory that underpins the subsequent sections. We explore the aforementioned estimation process in Section 3, and in Section 4 we apply it to synthetic derivatives data, both vanilla and exotic, before comparing their performance in predicting the prices of a wider class of derivatives. Finally, in Section 5, we provide an overview of the methodology and design choices made in this paper.

[1] Terry Lyons, Sina Nejad & Imanol Perez Arribas (2019) Numerical Method for Model-free Pricing of Exotic Derivatives in Discrete Time Using Rough Path Signatures, Applied Mathematical Finance, 26:6, 583-597, DOI: 10.1080/1350486X.2020.1726784

2 Theory

The concept of (price) path signatures is the cornerstone of this project, and to define it, we must first define the extended tensor algebra over $\mathbb{R}^d$.

Definition 1. The extended tensor algebra over $\mathbb{R}^d$ is defined as

$$T((\mathbb{R}^d)) = \{(a_0, a_1, a_2, ...) : a_n \in (\mathbb{R}^d)^{\otimes n} \ \ \forall n \geq 0\}$$

where $(\mathbb{R}^d)^{\otimes n}$ is the tensor product of n copies of $\mathbb{R}^d$.

Each element a_n could be viewed as the coefficients of an n^{th} degree homogeneous polynomial in non-commutative variables $X_1, ..., X_d$ [8], and hence a structure with d^n entries in $\mathbb{R}$. We can then define the signature of a continuous path, an element that belongs to this extended tensor algebra.

Definition 2. The signature of a continuous path $X : [0, T] \to \mathbb{R}^d$ with bounded variation is defined as

$$Sig(X) = (1, Z^1, Z^2, ...) \in T((\mathbb{R}^d))$$

where Z^n is the iterated integral

$$Z^n = \int_{0<t_1<...<t_n<T} dX_{t_1} \otimes ... \otimes dX_{t_n}$$

A signature contains high-dimensional information of a process, with Z^1 being the first order increment, $X_T - X_0$, of the path and higher terms representing the more complex features of the path, analogous to higher moments of random variables. However, the signature of a price process does not characterise the process itself uniquely. This can be seen from a simple example.

Remark 3. *By Theorem 2.9 (Chen's Theorem) and Proposition 2.14 in [8], the signature of a concatenated path $X * Y$ of X and Y is*

$$Sig(X * Y) = Sig(X) \otimes Sig(Y)$$

and the signature of the reverse of X, $\overleftarrow{X}$, is

$$Sig(\overleftarrow{X}) = Sig(X)^{-1}$$

where the inverse is defined on $T((\mathbb{R}^d))$ with respect to the tensor product. Therefore,

$$\begin{aligned} Sig(X * \overleftarrow{X}) &= \mathbf{1} \\ &= (1, 0, 0, ...) \in T((\mathbb{R}^d)) \end{aligned}$$

and $X * \overleftarrow{X}$ has the same signature as a constant path.

Rather than defining Sig on the price path, we define it on the lead-lag process instead.

Definition 4. Given a discrete price path $X : \mathbb{T} \to \mathbb{R}$ in the space Ω of all such paths, observed at $(t_i)_{i=0}^n$ with $0 = t_0 < t_1 < ... < t_n = T$, we define the corresponding lead-lag path to be

$$\begin{aligned} \hat{X}_{2k/2n} &= ((t_k, X_{t_k}), X_{t_k}) \\ \hat{X}_{(2k+1)/2n} &= ((t_k, X_{t_k}), X_{t_{k+1}}) \end{aligned}$$

and linear interpolation in between. This process takes values in $\mathbb{R}^2 \oplus \mathbb{R}$.

Lemma 5. *The signature of a lead-lag path $Sig(\hat{X})$ uniquely determines the lead-lag path $\hat{X}$, and hence the price path.*

As an illustration, we generate Black-Scholes paths starting at 1 with two different volatilities and plot their price and corresponding lead-lag paths in Figure 1 and 2.

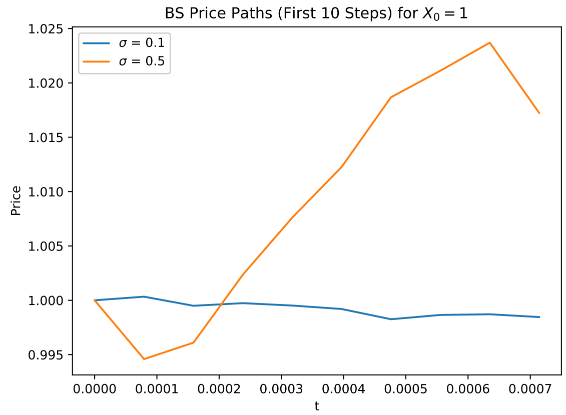

Figure 1: Black-Scholes Price Paths

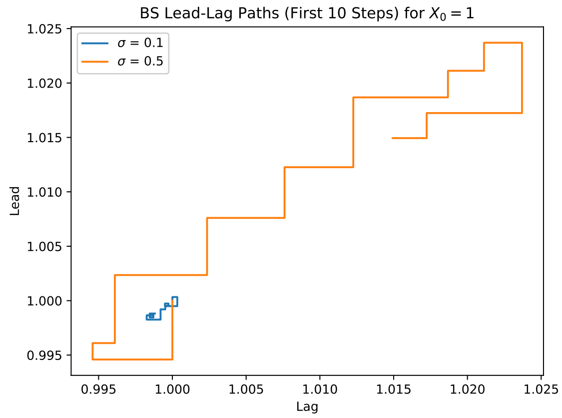

Figure 2: Black-Scholes Lead-Lag Paths

We then define signature payoffs, which are derivatives defined with respect to the underlying price path signature.

Definition 6. The signature payoff S^l is a function that maps a lead-lag path $\hat{X}$ to a real number, and is linear in the elements of $Sig(\hat{X})$:

$$S^l(\hat{X}) \ = < l, Sig(\hat{X}) > \in \mathbb{R}$$

where l is a linear form on $T((\mathbb{R}^d))$.

Then, for any path-dependent derivative $F(\hat{X})$, both its payoff and risk-neutral price could be well-approximated by the payoff and risk-neutral price of a signature payoff S^l. This is proven by Lyons, Nejad and Arribas [6]. Each element of l could be viewed analogously as Arrow-Debreu securities [2], paying a unit of numeraire for a particular path feature rather than a future state.

Since

$$< l_F, Sig(\hat{X}) > \approx \ F(\hat{X}), \tag{1}$$

for each $F \in \mathscr{F}$, we can approximate the linear form l_F corresponding to its signature payoff S^{l_F}. We can then estimate the implied expected signature, $\mathbb{E}^Q[Sig(\hat{X})]$ with our derivative-signature payoff pairs $\{(l_F, F)\}_{F \in \mathscr{F}}$ using the following relation:

$$< l_F, \mathbb{E}^Q[Sig(\hat{X})] > \approx \ P(F), \tag{2}$$

where $P(F)$ is the time-0 price of F and Q is the risk-neutral measure.

Similarly, we approximate the signature payoffs l_G for the out-of-sample derivatives $G \in \mathscr{G}$, obtaining these price estimates

$$\widehat{P(G)} \ = < l_G, \mathbb{E}^Q[Sig(\hat{X})] > .$$

While the signature uniquely determines the lead-lag path, it is an infinite-dimensional object. Therefore, in practice, we use the truncated signature defined below.

Definition 7. The truncated signature of $\hat{X}$ at N is defined to be

$$Sig^N(\hat{X}) = (a_0, a_1, ..., a_N, 0, ...) \in T^N(\mathbb{R}^d)$$

where

$$Sig(\hat{X}) = (a_0, a_1, a_2, ...) \in T((\mathbb{R}^d))$$

and $T^N(\mathbb{R}^d)$ is the truncated tensor algebra at $N \in \mathbb{N}$, defined as

$$T^N(\mathbb{R}^d) = \{(a_0, a_1, a_2, ...) \in T((\mathbb{R}^d)) : a_n = 0 \ \ \forall n \geq N\}$$

We choose N such that the approximations above still hold with $Sig(\hat{X})$ replaced with $Sig^N(\hat{X})$. We use N=4 due to computational limitations, while the original paper uses N=5.

3 Methodology

The restricted family of 100 derivatives $\mathscr{F}$ contains vanilla, up-and-in, up-and-out and variance options (25 each). We consider different subsets of $\mathscr{F}$ as the training set used to estimate the implied expected signature, and compare the results. As OTC option data is difficult to obtain, we use prices generated synthetically with a Heston model (Section 6.2 in [6], [3]) - whose parameters are calibrated to the market prices of vanilla options numerically - considered unknown to the agent to demonstrate the methodology. This allows us to evaluate the effectiveness of this method when generalised to other models not considered in the original paper. The Heston price paths are generated using Euler-Maruyama discretisation (Section 9.1 [5]) with the calibrated parameters. We compute the synthetic prices of each $F \in \mathscr{F}, G \in \mathscr{G}$ by taking the discounted average payoff.

We generate 10000 sample paths under the Black-Scholes model, and compute the truncated signatures $\{Sig^4(\hat{X}^{(i)})\}_{i=1}^{100}$ of the lead-lag paths $\{\hat{X}^{(i)}\}_{i=1}^{10000}$. For each of the derivatives $F \in \mathscr{F}$, their corresponding signature payoffs are estimated by applying linear regression on the truncated signatures, using (1) with $Sig(\hat{X})$ replaced with $Sig^4(\hat{X})$,

$$i.e. < l_F, Sig^4(\hat{X}^{(i)}) > \approx \ F(\hat{X}^{(i)}). \tag{3}$$

In this case, we have the design matrix $(Sig^4(\hat{X}^{(i)}))_{i=1}^{10000} \in \mathcal{M}^{10000,121}$ and target $(F(\hat{X}^{(i)}))_{i=1}^{10000} \in \mathcal{M}^{10000,1}$. We note that due to the fixed time step used in simulation, the increments in the first component of the lead-lag path $\hat{X}$ are equally spaced, and hence the features $Z^n_{0,0} = \frac{1}{n!}$ are linearly dependent by construct. Therefore, we remove these features before computing the Ordinary Least Squares (OLS) estimator.
We then apply linear regression again, this time estimating the implied expected signature using (2), with design matrix $(l_F)_{F\in\mathscr{F}} \in \mathcal{M}^{|\mathscr{F}|,121}$ and target $(P(F))_{F\in\mathscr{F}} \in \mathbb{R}^{|\mathscr{F}|}$. For the out-of-sample derivatives $G \in \mathscr{G}$, we use 50 Asian options with a range of strikes, i.e. options with payoff

$$max(\frac{1}{N}\sum_{i=1}^{N} X_i - K, 0)$$

with prices generated with the Heston model as before. The signature payoffs are estimated using (3) and we obtain price estimates using (2). We then compare $\widehat{P(G)}$ to their (synthetic) market prices $P(G)$.

4 Results

We calibrate the Heston model to *NASDAQ100 (NDX)* European options (provided by CRSP [4]) with least squares estimation, using the closed-form solution to the Heston model [3]. This is done using `quad` and `minimize` in the SciPy package [10]. This estimation results in fitted prices with a mean absolute deviation from mid-price of 14.25, which is smaller than the mean half-spread of 20.82. The negative estimated correlation between the price and variance process is not surprising, and can be attributed to the leverage effect commonly observed in the market. The prices of $F \in \mathscr{F}, G \in \mathscr{G}$ are then generated with the parameters shown in Table 1. This is done with 252 days x 20 observations/day = 5040 observations per paths and 50000 price paths. While generating more price paths would allow us to estimate the theoretical price more accurately, it is restricted by the trade-off between the number of time steps and the number of generated paths.

Parameter	**Value**
Initial Variance, v_0	0.0358
Long-Run Variance, θ	0.0741
Mean-Reversion Speed, κ	2.0465
Vol of Vol, ξ	0.8617
Correlation, ρ	-0.5475

Table 1: Calibrated Parameters of the Heston Model

We generate 10000 Black-Scholes paths with constant volatility ranging from 5% to 40%, and compute the signature using the esig package [9]. This is a computationally intensive task, and for 252 days x 50 observations/day = 12600 observations per path, this requires ≈ 17 minutes.

4.1 Estimating Signature Payoffs

One of the biggest challenges in estimating the linear forms is the multicollinearity in the design matrix. The second and third components of the lead-lag paths contain very similar information, which causes linear dependency between features. In particular, by removing columns with > 0.99 correlation with any other column, only 66 out of 121 columns remain. While this reduced the multicollinearity, this is not sufficient for the inversion $(X^T X)^{-1}$ in the OLS computation to be numerically stable. In addition, we use the Moore-Penrose inverse (pseudo-inverse) rather than the simple inverse [11], and results in a much more stable solution in terms of the magnitude of coefficients.

On the other hand, the derivative payoffs are zero for a lot of paths, e.g. paths whose terminal value is smaller than the strike for a European call. This worsens the fit of the signature payoffs, shown in Figure 3, and many fitted signature payoffs have a negative price. Therefore, we repeat this process with only the paths with strictly positive payoff, resulting in a much better fit, shown in Figure 4.

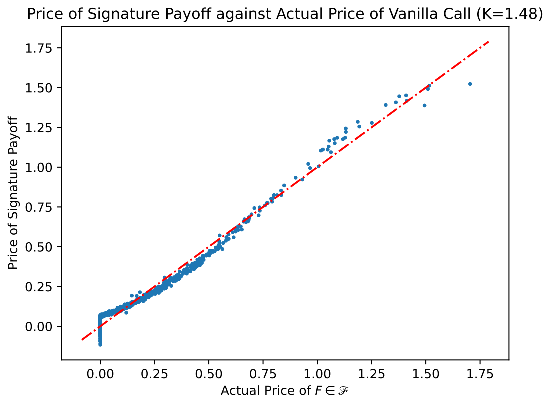

Figure 3: Price of Signature Payoff against Actual Price

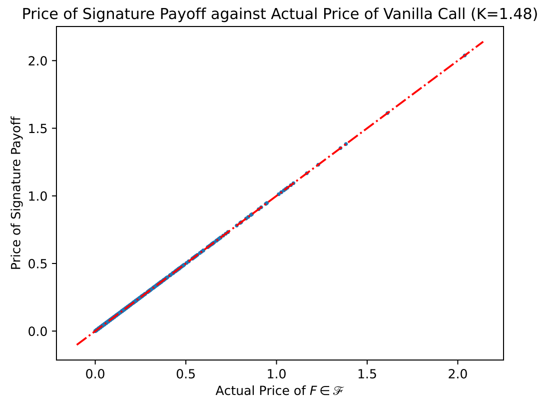

Figure 4: Price of Signature Payoff against Actual Price (Filtered)

We also note that the fit for the variance options is noticeably worse (Figure 5). The coefficients of the linear forms l_F for variance options F are much greater in magnitude, with range $[-305.974, 293.473]$ compared to $[-0.099, 0.912]$ for other $F \in \mathscr{F}$. Although regularisation (both L1 and L2) shrinks the coefficients to a range consistent with other derivatives in $\mathscr{F}$, it also worsens the fit considerably, and results in negative prices for signature payoffs. Computing the least squares estimator with log-prices circumvents this issue, but it also worsens the regression fit. Therefore, we procced with unregularised OLS results, regarding this difference in magnitude as the difference in characteristics of the derivatives considered.

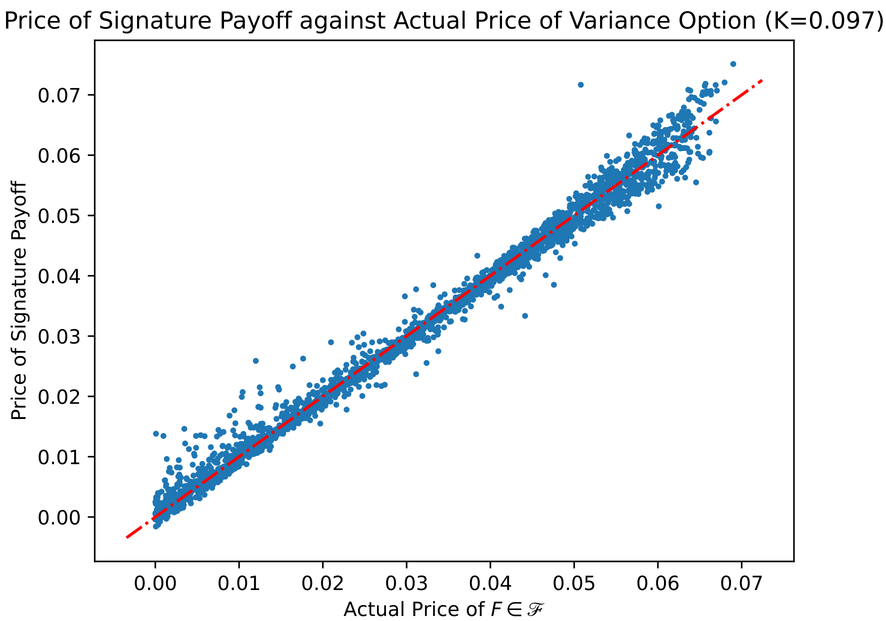

Figure 5: Signature Payoff against Actual Price of Variance Option

4.2 Estimating Implied Expected Signature

This issues encountered in estimating the implied expected signature is very similar to the ones in estimating the signature payoffs, i.e. multicollinearity and negative predictions. In addition, for $|\mathscr{F}|$ small, we have a smaller number of samples than features, since the design matrix $(l_F)_{F\in\mathscr{F}} \in \mathcal{M}^{|\mathscr{F}|,65}$. This causes the system of equations to be underdetermined. We observe that three components explain $> 99.99\%$ of the variance in the signature payoffs even after normalising (Figure 6).

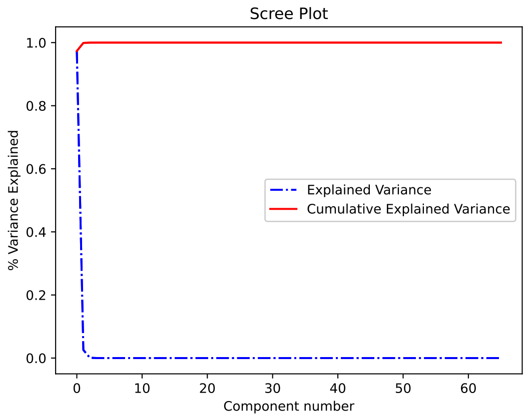

Figure 6: Explained Variance of Components

The numerical instability causes the implied expected signature to be large in magnitude, even reaching values on the order of 10^{10}, while the coefficients of the original Black-Scholes path signatures have a range of $[-0.826, 3.461]$. We note that the only subset of $\mathscr{F}$ that gives a reasonable implied expected signature and predicted price range is the subset of variance options. Since $n < p$, it is a perfect fit in-sample (Figure 7), but the out-of-sample price estimates we obtain are far from the actual prices (Figure 8).

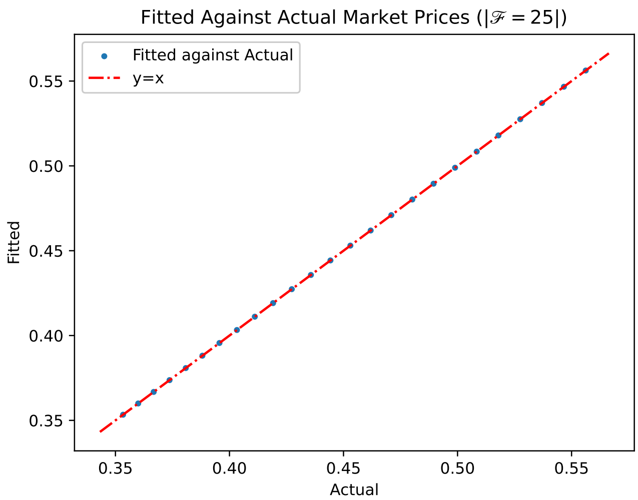

Figure 7: In-Sample Price Prediction

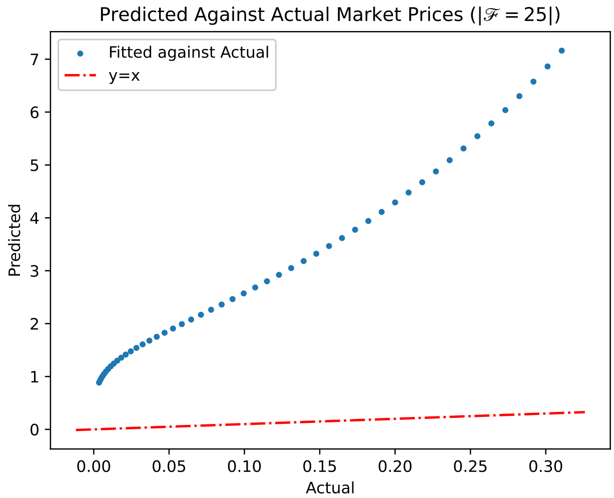

Figure 8: Out-of-Sample Price Prediction

5 Conclusion

In this project, we reviewed the theory and attempted to replicate the methodology used in a paper [7] on pricing exotic derivatives. We considered market prices generated by a calibrated Heston stochastic volatility model, which was not considered in the original paper. While using Euler-Maruyama discretisation allowed us to generate discrete observations of the stochastic volatility paths, the trade-off between the number of time steps and the number of paths could result in the synthetic prices deviating from their theoretical price. This could be improved by using more efficient simulation methods. Although we were able to demonstrate the estimation process of signature payoffs, the matrix inversion was highly unstable when computing the implied expected signature and yielded inaccurate results. The accuracy of the proposed method greatly depends on the invertibility of the linear map from signature payoffs to market prices, as mentioned in the original paper. While increasing the truncation threshold to $N = 5$ would increase the information encoded in the signature payoffs and might yield more accurate results, it also requires exponentially more computational resources. The increase in the number of features could also exacerbate the multicollinearity challenges faced in the estimation process. One of the potential solutions to the lack of invertibility would be to consider types of path other than the lead-lag path, for example, the augmented path used in a preceding paper by Arribas (Section 5.1 [1]). In the Black-Scholes simulation process, rather than a fixed time step, we could also use a varying or random time step. This could reduce the correlation between the features considered and improve the prediction fit.

References

[1] Imanol Perez Arribas. Derivatives pricing using signature payoffs. https://doi.org/10.48550/arXiv.1809.09466, 2018.

[2] Kenneth Arrow. The role of securities in the optimal allocation of risk-bearing. *The Review of Economic Studies*, 31(2):91–96, 1964.

[3] Ricardo Crisóstomo. An analysis of the heston stochastic volatility model: Implementation and calibration using matlab. https://arxiv.org/pdf/1502.02963, 2014.

[4] CRSP. Nasdaq european options data. https://wrds-www.wharton.upenn.edu/.

[5] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, 1992.

[6] Terry Lyons, Sina Nejad, and Imanol Perez Arribas. Nonparametric pricing and hedging of exotic derivatives. https://doi.org/10.48550/arXiv.1905.00711, 2019.

[7] Terry Lyons, Sina Nejad, and Imanol Perez Arribas. Numerical method for model-free pricing of exotic derivatives in discrete time using rough path signatures. https://doi.org/10.1080/1350486X.2020.1726784, 2019.

[8] Terry J. Lyons, Michael J. Caruana, and Thierry Lévy. *Differential Equations Driven by Rough Paths*. Springer, 2004.

[9] David Maxwell and Terry Lyons. Esig documentation. https://esig.readthedocs.io/en/latest/.

[10] SciPy. Scipy documentation. https://docs.scipy.org/doc/scipy/.

[11] statsmodels OLS. Package documentation. https://www.statsmodels.org/dev/generated/statsmodels.regression.linear_model.OLS.html.