Financial Applications of the Path Signature

Denny Serdarevic (ds22ck@fsu.edu)

Tyler Gorczycki (trg21a@fsu.edu)

Supervisor: Dr. Qi Feng (qfeng2@fsu.edu)

November 17th, 2024

Abstract

We explore the foundations of calibrating a linear combination of signature terms, expressed as polynomials using the signature method, to accurately model stochastic processes in financial applications, aiming to improve the calibration of equity pricing and implied (future) volatility models. Originating from rough path theory, the signature method offers a time-invariant transformation that captures the characteristics of multidimensional time series data. This method relies on iterated integrals, which form a natural linear basis for functions on data streams, allowing for the extraction of meaningful information in a non-parametric manner, without relying on traditional statistical methods. In our simulations, we use a Wiener process as the primary process in computation, providing a foundation for modeling stochastic behavior. Specifically, we assume the asset's price follows a geometric brownian motion and compute the signature of this process in discrete time-steps, regressing the trajectories of our asset against our computed linear combination of signature terms. We then extend this assumption to calibration towards both price and volatility simulated using the Heston model and SABR model, respectively. Furthermore, For our simulated options contracts, we use the Heston model alongside monte carlo simulations, and the Black-Scholes model, to calibrate towards price and implied volatility. In the novel setting, we further choose to adapt our options model to the market-observed "W-shaped" implied volatility curve for equity options. Occasionally, the market witnesses a W-shaped curve during especially volatile environments, such as earnings announcements and global events. Ultimately, this work has improved our ability to ascertain and model financial equities and derivatives, leading to more reliable pricing / volatility models.

1 Introduction

Within computational finance in the present time, there lies an abundance of statistical methodology aiming to both calibrate and predict an asset's dynamics. We consider here an unorthodox process emanating from Dr. Terry Lyon's Rough Path Theory to calibrate to an asset, which in our context concerns capturing a stochastic process and enhancing it. The theory of rough paths provides a powerful pathwise theory of SDEs originating from general classes of stochastic processes, and more precisely, rough paths. Our main focus relies on our parameters lying in geometric brownian motion, our drift and diffusion term, with our drift term defined as expected return, and the volatility term defined as standard deviation. We note that our driving process may equivalently be defined as any alternative stochastic process, if it contains well-observed quantities and can be comprised as an $\mathbb{R}^d$ valued path. We extend from basic calculus principles and build towards a tractable model, describing the tensor algebra operations and monte carlo method necessary for the model to be both well-structured and computationally feasible, respectively. Our asset S is approximated by a process $S_n(\ell)$ given as

$$S_n(\ell)_t := \ell(\hat{\mathbb{X}}_t)$$

where ℓ is a linear map of the signature of our primary process $\hat{X}$ up to degree $n \in \mathbb{N}$. We stress that a core benefit of employing the computations detailed below lies in the linearity of the model, allowing for accurate and fast results.

1.1 Literature Summary

We here comprise the works that have built upon the stochastic analysis necessary for our model. A majority of this paper's contents is a review and implementation of the work by Christa Cuchiero et al. (2022), which is an implementation of the original work performed by Lyons, Mcleod (2024), alongside foundational material provided in Chevyrev, Kormilitzin (2016). We will also note that via the shuffle product $\sqcup\!\sqcup$ stated generally from R. Ree (1958), every polynomial of the signature may be realized as a linear function. Our learning of tensor operations was guided by Hateley, and our introduction to the Heston model and it's implications was guided by the work of Heston (1993), Gatheral (2015), and R. Dunn et al. (2015). Our extension of our model to the W-shaped volatility curve was inspired and motivated by the Guassian-mixture model defined by Glasserman, Pirjol (2023), alongside Alexiou et al. (2021). We stress that we have avoided measure-theoretic principles of the signature, in assumption that we remain under the risk-neutral measure for our option model under Black-Scholes.

1.2 Preliminary Information of Paths

A path X in $\mathbb{R}^d$ is defined as a continuous mapping from a given closed interval $[a,b]$ to $\mathbb{R}^d$, stated as $[a,b] \to \mathbb{R}^d$. We consider the parametrization of this path and further denote as such: X_t: $\{t \in [a,b] \to \mathbb{R}^d\}$. This parametrization of the path generalizes in d-dimensions to:

$$X : [a,b] \to \mathbb{R}^d, X_t = \{X_t^1, X_t^2, X_t^3, ..., X_t^d\}$$

For this preliminary, we will choose to assume that our paths are piecewise differentiable. This property will become necessary later. An example of a smooth path in $\mathbb{R}^2$ is given by:

$$X_t = \{X_t^1, X_t^2\} = \{t, \frac{t^2}{4} - 2t + 5\},\ t \in [0, 10]$$

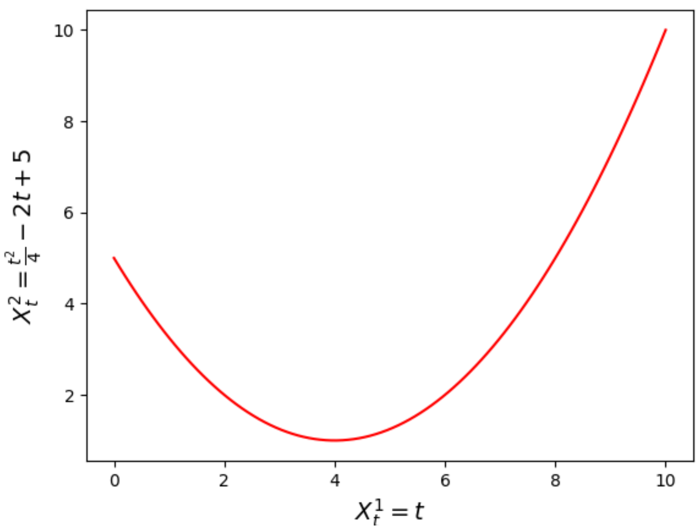

Figure 1

Geometric properties of this path can be yielded via line integration, note that within $\mathbb{R}^2$ we defined the time component itself as a path. Examples include

$$\int_0^{10} X_t^1 dX_t^1, \quad \int_0^{10} X_t^2 dX_t^2$$

One can also capture interactions of X_t^1, X_t^2, understanding X_t^2 to be a function of X_t^1, set $Y_t = f(X_t) = X_t^2$:

$$\int_a^b Y_t dX_t^1 = \int_a^b Y_t \frac{dX_t^1}{dt} dt, \quad Y, X : [a,b] \to \mathbb{R}$$

Stratonovich integration can be thought of as an extension of line integration, which we define later.

An example of a piecewise linear non-smooth path may be constructed as:

$$X_t = \{X_t^1, X_t^2\} = \{t, f(t)\},\ t \in [0,1]$$

Where we can consider $f(t)$ as being a function f of a stock price over time t:

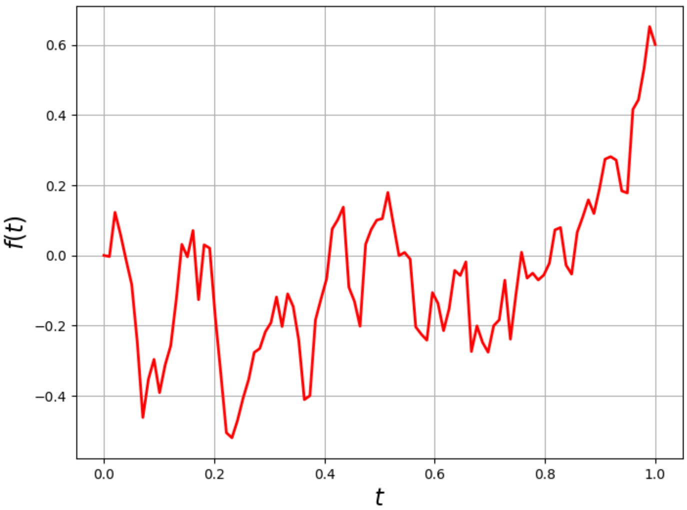

Figure 2

The above represents a path that is derived from time-sequential data, and is not immediately defined by a closed form equation. A real-valued stochastic process is defined as a sequence of Random Variables, indexed by time. A prominent example of a stochastic process is standard Brownian Motion, where in simulation, values are selected from the standard normal distribution at random: $\Delta B_t \sim \mathcal{N}(0, \Delta t)$, and the value-selection is given by a Random Number Generator.

1.3 Construction of the Signature

Given a stochastic process such as the function shown above, one may be able to extract information on the respective underlying behavior. In order to correctly group and define this information extraction, the usage of tensor algebra is necessary. When dealing with paths in higher-dimensional space $\mathbb{R}^d$, iterated integration of the path's components from classical calculus fails to encode the full interaction between different dimensions. We will be under a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ for the full length of the paper.

Example 1.1: We define a simple tensor product for clarification, for two vectors V,W, from [3]:

$$V = [1,2,3], \quad W = [4,5,6]$$

$$V \otimes W = \begin{pmatrix} v_1w_1 & v_1w_2 & v_1w_3 \\ v_2w_1 & v_2w_2 & v_2w_3 \\ v_3w_1 & v_3w_2 & v_3w_3 \end{pmatrix} = \begin{pmatrix} 4 & 5 & 6 \\ 8 & 10 & 12 \\ 12 & 15 & 18 \end{pmatrix}$$

We receive a higher-order structure from two lower-order structures. This can generalize from the tensor product of two scalars(a vector) to a tensor product of two lower ordered tensors. We also take note that the tensor product is not commutative, as above we see $V \otimes W \neq W \otimes V$. One can think about the tensor product as taking information from the W vector and meshing it with each data point in the V vector.

Let V be a vector space over a field K. For a non-negative integer k, we define the k-th tensor power of V to be the tensor product of V with itself k times:

$$T^K V = V^{\otimes k} = V \otimes V \otimes ... \otimes V$$

Defining a k-th tensor power is necessary for the context of the signature, as it allows for logging the interactions between the components of a path. Further down this paper, we concern ourselves with the **level** of the signature, loosely defined (rigorously defined later) as the amount of iterated integrals we choose to compute whether it be $\int X, \quad \int\int X, \quad \int\int\int X...$
To the k-th tensor power of $\mathbb{R}^d$, our defined level-1 signature terms are vectors housed in $\mathbb{R}^d$. Level 2 signature terms are tensors, housed in $\mathbb{R}^d \otimes \mathbb{R}^d$ and so on.

With the vector space $\mathbb{R}^d$, and for each $n \in \mathbb{N}_0$, we re-construct the n-fold tensor product of $\mathbb{R}^d$ as

$$(\mathbb{R}^d)^{\otimes n} := \{\underbrace{\mathbb{R}^d \otimes ... \otimes \mathbb{R}^d}_{n}\}, \quad \mathbb{R}^{d\otimes 0} := \mathbb{R}$$

Example:

$$(\mathbb{R}^2)^{\otimes 3} = \{\mathbb{R}^2 \otimes \mathbb{R}^2 \otimes \mathbb{R}^2\}$$

This creates an $(n * d)$-dimensional space out of n vectors from d-dimensional space. In principle, allowing the construction of the n-fold tensor product, in contrast to the $(n-1)$-fold tensor product, we are allowed to encode a richer interaction of the vector spaces. The increase in dimensionality allows for the representation of more complex structures.

Definition 1.1 For $d \in \mathbb{N}$, we construct the extended tensor algebra on $\mathbb{R}^d$ defined from C. Cuchiero et al, [1]

$$T((\mathbb{R}^d)) := \{\mathbf{a} := (a_0, ..., a_n) : a_n \in (\mathbb{R}^d)^{\otimes n}\} \tag{1.1}$$

The employment of an extended tensor algebra allows us to operate with multi-linear maps, which forms a core aspect of the signature method. For a given multi-linear map

$$g : V_1 \times ... \times V_n \to W$$

and a linear map

$$G : V_1 \otimes ... \otimes V_n \to W$$

we have a bijection. The higher-order structure of the tensor product implicitly matches the behavior of multi-linear mapping of several vectors. Directly put, multi-linear maps can be simplified into linear maps by operating in the tensor product space. A simple example in this context could be to take

$$g : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$$

defined as

$$g(x, y) = 3x_1y_1 + x_2y_2$$
$$x = (x_1, x_2) \in \mathbb{R}^2, \quad y = (y_1, y_2) \in \mathbb{R}^2$$

Where $g(x, y)$ is linear in x when y, and the converse defined similarly. Similarly we have

$$G : \mathbb{R}^2 \otimes \mathbb{R}^2 \to \mathbb{R}$$

$$G(x \otimes y) = 3x_1y_1 + x_2y_2$$
$$x \otimes y = \begin{bmatrix} x_1y_1 & x_1y_2 \\ x_2y_1 & x_2y_2 \end{bmatrix}, \quad (x \otimes y) \in \mathbb{R}^4$$

$x \otimes y$ provides a basis for operations in the tensor product space.

Definition 1.2 To this end we form the truncated tensor algebra of order $N \in \mathbb{N}$ defined from [1]:

$$T^{(N)}(\mathbb{R}^d) := \{a \in T((\mathbb{R}^d)) : a_n = 0, \forall n > \mathbb{N}\} \tag{1.2}$$

Where our elements a are tensors themselves. We take note that $T^{(N)}(\mathbb{R}^d)$ has dimension

$$\sum_{i=0}^{N} d^i = \frac{(d^{N+1} - 1)}{(d-1)}$$

Example 1.2: For further clarity, in regards to paths, d represents the dimension of the space in which the path is taking values. If we have a 3-dimensional path $\hat{X} = [X_t, W_t, t]$, then $d = 3$ and $\hat{X}$ is lying in

$\mathbb{R}^3$. Furthermore, n represents to what level we choose to take the signature, so if we are in the midst of a 3-dimensional path, it would be ideal to set $n \geq 3$ in order to capture higher level interactions of the path components. If we fail to do so, and choose $n = 2$, then we would have the following extended tensor algebra from equation (1.2):

$$T^{(2)}(\mathbb{R}^3) := \{a \in T((\mathbb{R}^3)) : a_n = 0, \forall n > 2\} \quad \text{(meaning } a_3 = 0\text{)}$$

Operations: For a given $\mathbf{a}$,$\mathbf{b} \in T((\mathbb{R}^d))$ and $\lambda \in \mathbb{R}$,

$$\mathbf{a} + \mathbf{b} := (a_0 + b_0, ..., a_n + b_n, ...)$$

$$\lambda \mathbf{a} := (\lambda a_0, ..., \lambda a_n, ...)$$

$$\mathbf{a} \otimes \mathbf{b} := (c_0, ..., c_n, ...)$$

Where $c_n := \sum_{k=0}^{n} a_k \otimes b_{n-k} := \mathbf{a} \otimes \mathbf{b}$. We will need to bring organization in relation to this structure, in order to correctly group each element $\mathbf{a}$. We choose to make a multi-index $I := (i_1, i_2, ..., i_n)$, in which we set $|I| := n$. We also choose $I' := (i_1, ..., i_{n-1})$ and $I'' := (i_1, ..., i_{n-2})$ when applicable.

This forms $\{I : |I| = n\} := \{1, 2, ..., d\}^n$, upon which we combine the multi-index with the tensor basis:

$$e_I = e_{i_1} \otimes e_{i_2} \otimes ... \otimes e_{i_n} \in (\mathbb{R}^d)^{\otimes n}$$

Where $e_1, ..., e_d$ denotes the canonical basis vectors of $\mathbb{R}^d$. The choice of truncation level N determines the highest length of the multi-index. For example when $N = 2$, we consider the multi-indices up to length 2, such as $(1), (2), (1, 2)$, and $(2, 1)$. We take note that the set $\{e_i : |I| = N\}$ is an orthonormal basis of $(\mathbb{R}^d)^{\otimes n}$. For ease of understanding, multi-indices can be represented as words. Consider an alphabet with 2 letters $A = \{1, 2\}$. One could construct an infinite set B of all words by the set of all letters A

$$B \rightarrow \{1, 2, 11, 12, 21, 22, 111, 112, ...\}$$

Imposing a truncation $T^{(N)}(\mathbb{R}^d)$ of the number of words up to two letters (N = 2) results in:

$$B_{N=2} \rightarrow \{1, 2, 11, 12, 21, 22\}$$

Similarly for a set W,

$$W = \{(i_1, ..., i_n) | n \geq 1, i_1, ..., i_n \in \{1, ..., d\}\}$$

we have a set of *words* on the *alphabet* $A = \{1, ..., d\}$ consisting of d letters, making use of the multi-index. Similarly, we can see how for $\mathbf{a} := (a_0, ..., a_n, ...) : a_n \in (\mathbb{R}^d)^{\otimes n}$ we can group each a_n into a word.

Each element of $\mathbf{a} \in T((\mathbb{R}^d))$ can be written as

$$\mathbf{a} = \sum_{|I| \geq 0} a_I e_I$$

for an $a_I \in \mathbb{R}$. This gives us an element or word multiplied with a basis tensor. Furthermore, for each $\mathbf{a} \in T((\mathbb{R}^d))$ and each $\mathbf{b} \in T((\mathbb{R}^d))$ we set

$$\langle \mathbf{a}, \mathbf{b} \rangle := \sum_{|I| \geq 0} \langle a_I, b_I \rangle$$

Where $b_I = \langle e_I, \mathbf{b} \rangle$

Definition 1.3 Let $(X_t)_{t \in [0,T]}$ be a continuous $\mathbb{R}^d$-valued wiener process. The *Signature* of X is the $T((\mathbb{R}^d))$-valued process $(s, t) \rightarrow \mathbb{X}_{s,t}$ with the following components, from [1]:

$$\langle e_\phi, \mathbb{X}_{s,t} \rangle := 1, \quad \langle e_I, \mathbb{X}_{s,t} \rangle := \int_s^t \langle e_{I'}, \mathbb{X}_{s,r} \rangle \circ dX_r^{i_n} \tag{1.3}$$

$\forall (i_1, ...i_n) \in I$, $(i_1, ..., i_{n-1}) \in I'$, and $0 \leq s \leq t \leq T$, where $\circ$ denotes the Stratonovich integral, with

$$X : [0, T] \times \Omega \rightarrow \mathbb{R}^d, \quad X_{t+u} - X_t \sim \mathcal{N}(0, u)$$

where u is an arbitrary time-step. The inner product $\langle e_I, \mathbb{X}_{s,t}\rangle$ measures how much the tensor $\mathbb{X}_{s,t}$ aligns with the basis tensor e_I, where for a level 1 term

$$\int_s^t 1 \circ dX_s^1 = \lim_{n\to\infty} \sum_{i=0}^{k-1} (X_{t_{i+1}}^1 - X_{t_i}^1) = X_t^1 - X_s^1$$

We employ the use of the Stratonovich integral because, given our geometric considerations, it preserves the covariance properties under change of variables and reparametrization of the path as a random variable. For $\mathbb{X}_t$ and our choice of $\mathbb{R}^d$, our computation housing Stratonovich to level **3** is:

$$\mathbb{X}_{s,t} = \Bigg(1, \int_0^t 1 \circ dX_s^1, \dots, \int_0^t 1 \circ dX_s^d, \int_0^t \left(\int_0^s 1 \circ dX_r^1\right) \circ dX_s^1,$$
$$\int_0^t \left(\int_0^s 1 \circ dX_r^1\right) \circ dX_s^2, \dots, \int_0^t \left(\int_0^s 1 \circ dX_r^d\right) \circ dX_s^d,$$
$$\int_0^t \left(\int_0^s \left(\int_0^r 1 \circ dX_u^1\right) \circ dX_r^1\right) \circ dX_s^1, \dots, \int_0^t \left(\int_0^s \left(\int_0^r 1 \circ dX_u^d\right) \circ dX_r^d\right) \circ dX_s^d\Bigg)$$

Definition 1.4: The projection $\mathbb{X}^N$ on $T^{(N)}(\mathbb{R}^d)$ is given below from [1]:

$$\mathbb{X}_{s,t}^N = \sum_{|I|\le N} \langle e_I, \mathbb{X}_{s,t}\rangle e_I \tag{1.4}$$

and is called the *signature* of X truncated at level N. After performing the decomposition above, the projection allows for the scalar terms computed from $\langle e_I, \mathbb{X}_{s,t}\rangle$ to be drafted on to our tensor bases e_I, thus reconstructing the tensor $\mathbb{X}_{s,t}^N$ as a linear combination of the basis tensors. From this computation we are ensured that $\mathbb{X}_{s,t}^N$ represents $\mathbb{X}_{s,t}$ in the truncated tensor space. Level 2 of the signature $N = 2$ would immediately resolve to the following under the projection:

$$\mathbb{X}_{s,t}^2 = \langle e_{()}, \mathbb{X}_{s,t}\rangle e_{()} + ... + \langle e_{(2,2)}, \mathbb{X}_{s,t}\rangle e_{(2,2)}$$

Example 1.3: Given primary process $\hat{X}_t := (t, X_t)$, we can choose X_t to be an $\mathbb{R}^2$-valued geometric brownian motion

$$X_t = X(0)e^{(\mu - \frac{\sigma^2}{2})t + \sigma W_t}$$

We compute the Signature of X truncated at level 2 for $t \in [0, 5]$, taking note that t itself is the second path alongside X_t:

$$\text{dimension of } T^{(2)}(\mathbb{R}^2) = 7 = \sum_{i=0}^{2} 2^i = \frac{(2^{2+1} - 1)}{(2-1)}$$

$$(\mathbb{R}^2)^{\otimes 1} = \mathbb{R}^2, \quad (\mathbb{R}^2)^{\otimes 2} = \mathbb{R}^2 \otimes \mathbb{R}^2$$

Given $\mathbb{R}^2$ spanned by the basis vectors $\{e_1, e_2\}$, then

$$(\mathbb{R}^2)^{\otimes 1} = \text{span}\{e_i | i \in \{1, 2\}\} = e_1, e_2$$
$$(\mathbb{R}^2)^{\otimes 2} = \text{span}\{e_i \otimes e_j | i, j \in \{1, 2\}\} = e_1 \otimes e_1, e_1 \otimes e_2, e_2 \otimes e_1, e_2 \otimes e_2$$

Then,

$$\mathbb{X}_{(s,t)\in[0,5]} = 1, \int_0^5 1 \circ dX_s^1, \int_0^5 1 \circ dX_s^2, \int_0^5 \left(\int_0^s 1 \circ dX_r^1\right) \circ dX_s^1, \int_0^5 \left(\int_0^s 1 \circ dX_r^1\right) \circ dX_s^2,$$
$$\int_0^5 \left(\int_0^s 1 \circ dX_r^2\right) \circ dX_s^1, \int_0^5 \left(\int_0^s 1 \circ dX_r^2\right) \circ dX_s^2$$

Housed within our $T(\mathbb{R}^d)$-valued process:

$$\langle e_\phi, \mathbb{X}_{s,t}\rangle := 1, \quad \langle e_I, \mathbb{X}_{s,t}\rangle := \int_0^t \langle e_{I'}, \mathbb{X}_{s,r}\rangle \circ dX_r^{i_n}$$

we see:

Level 0 Term:

$$\langle e_\phi, \mathbb{X}_{0,t}\rangle = 1$$

Level 1 Terms:

$$\langle e_{(1)}, \mathbb{X}_{0,t}\rangle = \int_0^5 1 \circ dX_r^1 = t = 5$$

$$\langle e_{(2)}, \mathbb{X}_{0,t}\rangle = \int_0^5 1 \circ dX_r^2$$

Level 2 Terms:

$$\langle e_{(1,1)}, \mathbb{X}_{0,t}\rangle = \int_0^5 \left(\int_0^r 1 \circ dX_u^1\right) \circ dX_r^1$$

$$\langle e_{(1,2)}, \mathbb{X}_{0,t}\rangle = \int_0^5 \left(\int_0^r 1 \circ dX_u^1\right) \circ dX_r^2$$

$$\langle e_{(2,1)}, \mathbb{X}_{0,t}\rangle = \int_0^5 \left(\int_0^r 1 \circ dX_u^2\right) \circ dX_r^1$$

$$\langle e_{(2,2)}, \mathbb{X}_{0,t}\rangle = \int_0^5 \left(\int_0^r 1 \circ dX_u^2\right) \circ dX_r^2$$

We would then use our projection $T^{(N)}(\mathbb{R}^d)$ from (1.2) to receive:

$$\mathbb{X}_{0,t}^2 = \sum_{|I|\leq 2} \langle e_I, \mathbb{X}_{0,t}\rangle e_I = \langle e_{()}, \mathbb{X}_{0,t}\rangle e_{()} + ... + \langle e_{(2,2)}, \mathbb{X}_{0,t}\rangle e_{(2,2)}$$

Definition 1.5: From [2], We define the commutative shuffle algebra $(T(\mathbb{R}^d), +, ⧢)$. For any two multi-indices $I := (i_1, ..., i_n)$ and $J := (j_1, ..., j_m)$ the shuffle product is defined as

$$e_I ⧢ e_j := (e_{I'} ⧢ e_J) \otimes e_{i_n} + (e_I ⧢ e_{J'}) \otimes e_{j_m}$$

where $e_I ⧢ e_\emptyset := e_\emptyset ⧢ e_I = e_I$, and for any words or multi-indices m and n we define $m ⧢ n = n ⧢ m$. This definition extends to **a, b** $\in T(\mathbb{R}^d)$ as:

$$\mathbf{a} ⧢ \mathbf{b} = \sum_{|I|,|J|\geq 0} a_I b_J (e_I ⧢ e_J) \tag{1.5}$$

The shuffle product is a computation which as a by-product eliminates redundancy, and it allows for further algebraic manipulation of the signature. We take note that the shuffle product is not absolutely necessary in computation of the signature, however it can be employed to *relate* different terms, thus promoting the inherent structure of this process. An example of the shuffle for two single integrals $S^{(1)}(X) = \int_0^t dX_s^1, \quad S^{(2)}(X) = \int_0^t dX_s^2$ gives

$$S^{(1)}(X) ⧢ S^{(2)}(X) = S^{(1,2)}(X) + S^{(2,1)}(X)$$

Equivalently,

$$\int_0^t dX_s^1 ⧢ \int_0^t dX_s^2 = \int_0^t \int_0^s dX_r^1 dX_s^2 + \int_0^t \int_0^s dX_r^2 dX_s^1$$

Example 1.4 For pre-defined paths X^1, X^2, $\langle e_1, \mathbb{X}_{s,t}\rangle = \int_0^t dX_s^1$, $\langle e_2, \mathbb{X}_{s,t}\rangle = \int_0^t dX_s^2$ gives us

$$\langle e_1, \mathbb{X}_{s,t}\rangle \sqcup\!\sqcup \langle e_2, \mathbb{X}_{s,t}\rangle = \langle e_{1,2}, \mathbb{X}_{s,t}\rangle + \langle e_{2,1}, \mathbb{X}_{s,t}\rangle$$

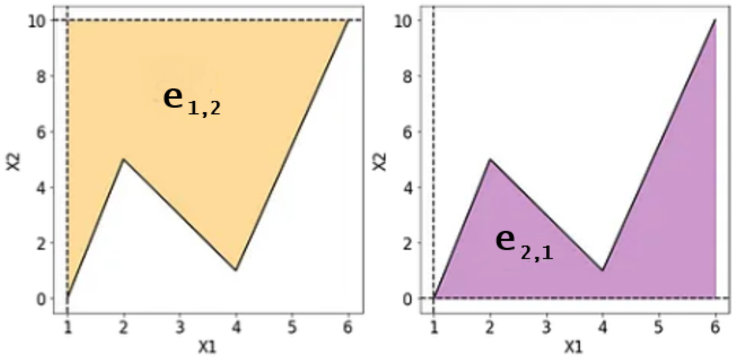

Figure 2

1.4 Signature Model

Since the brownian motion X will be held as our main construct, we call it the *primary process* and define it's extension below. We strive to remind that our stochasic processes is defined on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ where depending on the context, we consider one of the two extensions of X.

(i) $\hat{X}_t := (t, X_t, [X]_t)$ where $[X]$ denotes the d^2-dimensional process given by quadratic variation of X.

(ii) $\hat{X}_t = (t, X_t)$. This extension will always be paired with the assumption that X is a brownian motion with absolutely continuous characteristics.

Remark 1: In several contexts it is also necessary to consider X to be a vector of correlated Brownian motion paths with correlation matrix $\rho \in [-1, 1]$ and $\hat{X}$ given by definition 3.1(ii). Suppose we label two brownian motions B and W respectively, and choose $W_{t_1} = B_{t_1}$. In our context we choose to state the correlation via

$$W_t = \rho B_t + \sqrt{(1-\rho^2)}\tilde{B}_t$$

where W_t and B_t are also housed within the primary process extension $\hat{X}_t := (t, B_t, W_t)$. We define $\tilde{B}_t$ to have the property $Cov(B_t, \tilde{B}_t) = 0$ to ensure that the two are independent and orthogonal. With this construction of $\tilde{B}_t$, we are then capable of correlating W_t, B_t.

Definition 1.6: For $[X]$ as the d^2-dimensional process of quadratic covariation, we define

$$[X]_t = \lim_{||P||\to 0} \sum_{k=1}^{n} \left(X_{t_k} - X_{t_{k-1}}\right)^2 \tag{1.6}$$

where P ranges over partitions of the interval $[0, t]$

Definition 1.7: From [1], In regards to a vector of correlated brownian motions, we involve correlation matrix ρ and $\hat{X}$ given by 1.4(i). Also defining e_0 for the component of $\hat{X}$ corresponding to time, e^k for it's component corresponding to X^k, and ϵ_{ij} for the component of $\hat{X}$ corresponding to $[X^i, X^j]$. We then state the Tilde transformation for our tensor space e_I,

$$\tilde{e}_I^k = e_I \otimes e_k - \frac{1}{2} e_{I'} \otimes \epsilon_{i_{|I|}k} \quad \forall |I| > 0 \tag{1.7}$$

Equivalently,

$$\tilde{e}_I^k = e_I \otimes e_k - \frac{\rho_{i_{|I|},k}}{2} 1_{\{i_{|I|}\neq 0\}} e_{I'} \otimes e_0 \quad \forall |I| > 0$$

In this context, we are concerning ourselves with the adjustment of value of our tensor space e_I in regards to the k-th correlated brownian motion X^k. We adjust the tensor using the correlation ρ specifically between $X^{i_{|I|}}$ and X^k, with the last element of I, only for when it is not the path for time per the indicator function, previously stated in the primary process of both (i) and (ii). The indicator is such that the correction is only applied when the last component of I is the correlated brownian motion brownian motion and not the time path. The factor of $\frac{\rho_{i_{|I|},k}}{2}$ ensures that the subtraction is proportional to two non-time paths. This definition will be crucial later for options simulation. Proof has been given in [1].

Definition 1.8: From [1], A *Signature Model* is a stochastic process of the form

$$S_n(\ell)_t := \ell_\emptyset + \sum_{0<|I|\leq n} \ell_I \langle e_I, \hat{\mathbb{X}}_t \rangle \tag{1.8}$$

where $n \in \mathbb{N}$ and $\ell := \{\ell_\emptyset, \ell_I : 0 < |I| \leq n\}$. This model is such that for each $t \in [0,T]$, $S_n(\ell)_t$ is linear in $\hat{\mathbb{X}}_t$. Following $\hat{\mathbb{X}}_t$ to be the iterated integrals of the signature up to level N, we compute $\ell_I \langle e_I, \hat{\mathbb{X}}_t \rangle$ which is a simple scalar product

$$S_n(\ell)_t = \ell_\emptyset + \ell_{i_1} \langle e_{i_1}, \hat{\mathbb{X}}_t \rangle + \ell_{i_1 i_2} \langle e_{i_1 i_2}, \hat{\mathbb{X}}_t \rangle + \cdots + \ell_{i_1 i_2 \ldots i_n} \langle e_{i_1 i_2 \ldots i_n}, \hat{\mathbb{X}}_t \rangle$$

where $\ell_I \langle e_I, \hat{\mathbb{X}}_t \rangle$ lies in tensor space. In our context, the signature model will be applied under discrete time settings.

2 Model Calibration to time-series data

2.1 Calibration to Spot Price

We understand that with our assumptions, the spot price of a stock S in the market roughly follows geometric brownian motion. We strive to calibrate our model detailed above to both simulated prices of S over time. With this assumption for simulated prices, we choose $X_t := (W_t)$ and further set $\hat{X}_t := (t, W_t)$. With our asset S given as geometric brownian motion, we suppose that price data is available on a time grid $t_0, ...t_N$, with $N > 1$. Our standard brownian motion W_t for $t_0, ...t_N$ is computed in simulation as

$$W_{t_{i+1}} = W_{t_i} + \sqrt{t_{i+1} - t_i} \cdot Z$$

$$Z \sim \mathcal{N}(0, t)$$

And our corresponding SDE for S

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

with the closed form solution

$$S_t = S(0) e^{(\mu - \frac{\sigma^2}{2})t + \sigma W_t}$$

The distribution of S_t follows a log-normal distribution, which is concurrent with the brownian motion following a normal distribution. Strictly following from above, one can solve for W_t:

$$W_t = \frac{1}{\sigma} \left(log \frac{S_t}{S_0} - \left(\mu - \frac{1}{2}\sigma^2 \right) t \right)$$

In an ideal environment, one possesses well-stated data for μ and σ. These variables are not directly observable from the market. Therefore we continue below with a formulated approach to recover W_t. Take note that one is allowed to simply possess the underlying W_t of this asset S by having the asset share the underlying randomness in prices, as detailed above. However, to properly **extrapolate** the trajectories of W_t as seen in real market data, the usage of spot quadratic variation is a viable option, as detailed in [1]:

$$\hat{W}_t = \sum_{i=0}^{N-1} \frac{S_{t_{i+1}} - S_{t_i}}{\sqrt{(S_{t_{i+1}} - S_{t_i})^2}}$$

Where the denominator represents the square root of the quadratic variation of our stock price. Finding the quadratic variation of a price is described as in definition 1.6:

$$[S]_t = \sum_{i=0}^{N} (S_{t_{i+1}} - S_{t_i})^2$$

Now with the possession of $X_t \in (\mathbb{R}^2)$, $X_t := (t, \hat{W}_t)$, we choose level 2 of the signature and compute against it

$$\mathbb{X}_t = \left(1, \int_0^t 1 \circ dX_s^1, \int_0^t 1 \circ dX_s^2, \int_0^t \left(\int_0^s 1 \circ dX_r^1\right) \circ dX_s^1,\right.$$

$$\left.\int_0^t \left(\int_0^s 1 \circ dX_r^1\right) \circ dX_s^2, \int_0^t \left(\int_0^s 1 \circ dX_r^2\right) \circ dX_s^1, \int_0^t \left(\int_0^s 1 \circ dX_r^2\right) \circ dX_s^2\right)$$

Taking note that $X^1 := t$, $X^2 := \hat{W}_t$, and clarifying that $T^{(2)}(\mathbb{R}^2) = 7 = \sum_{i=0}^{2} 2^i = \frac{(2^{2+1}-1)}{(2-1)}$, the 7 terms pictured above. At this stage we now employ the aforementioned model $S_n(\ell)_t$

$$S_2(\ell)_t = \ell_\emptyset + \ell_1\langle e_1, \hat{\mathbb{X}}_t\rangle + \ell_2\langle e_2, \hat{\mathbb{X}}_t\rangle + \cdots + \ell_{(2,2)}\langle e_{(2,2)}, \hat{\mathbb{X}}_t\rangle$$

Concerning simulation, $S_n(\ell)_t$ is taken sequentially at each time interval, for example if we measure one tick per day for 252 trading days, we receive the sequence of polynomials $S_2(\ell)_1, ..., S_2(\ell)_{252}$

Definition 2.1: For calibration, we choose to perform Lasso Regression, regressing the trajectories of S on the signature of $\hat{X}$, within the context of $\hat{X}_t = (t, \hat{W}_t)$ (1.5.ii). From [1] we construct the calibration problem at hand

$$\ell^* \in \text{argmin}_\ell L_{\text{price},\alpha}(\ell) \tag{2.1}$$

Definition 2.2: where our loss function $L_{price,\alpha}(\ell)$ is constructed

$$L_{\text{price},\alpha}(\ell) := \sum_{i=1}^{N} \left(S_n(\ell)_{t_i} - S_{t_i}\right)^2 + \alpha(\ell) \tag{2.2}$$

We denote α as a penalization function for regression. With this function, essentially a comparison is made between the approximation $S_n(\ell)_t$ and the true S_t. In generalizing to the least squares Lasso function for clarification,

$$\min_{\beta_0,\beta}\left(\sum_{i=1}^{N}(y_i - \beta_0 - x_i^T\beta)^2\right) \text{ subject to } \quad \sum_{j=1^p} |\beta_j| \leq t$$

We see

- y_i as target variable, $y_i = S_{t_i}$.
- x_i as feature vector for i-th data point, $x_i = \langle e_I, \hat{\mathbb{X}}_{t_i}\rangle$.
- β_0 as intercept term, $\beta_0 = S_0$.
- β as coefficients for the features(ℓ's).
- And we choose $\alpha(\ell) = \lambda||\ell||_i$ as penalization term. A comfortable choice being $\alpha = 10^{-5}$.

Upon retrieving ℓ^* we then calibrate to spot price trajectory S using **Definition 2.3:**

$$\hat{S}_{t_i} = \sum_{0\leq|I|\leq n-1} \ell_I^*\langle e_I, \mathbb{X}_{t_i}\rangle \tag{2.3}$$

Once again, within our respective time grid $t_1, ..., t_N$

Example 2.1: For simulation, we choose the model parameters

S_0	μ	σ
\$1.00	0.01	0.2

with $\alpha(\ell) = 10^{-5}$, and choosing for time increments $N = 1000$ days, we simulate a 6 hour trading day with 360 measurements evenly indexed daily. Bringing total measurements of price to $360,000$. Choosing level 3 of the signature with $\hat{X}_t \in \mathbb{R}^2$, we calibrate:

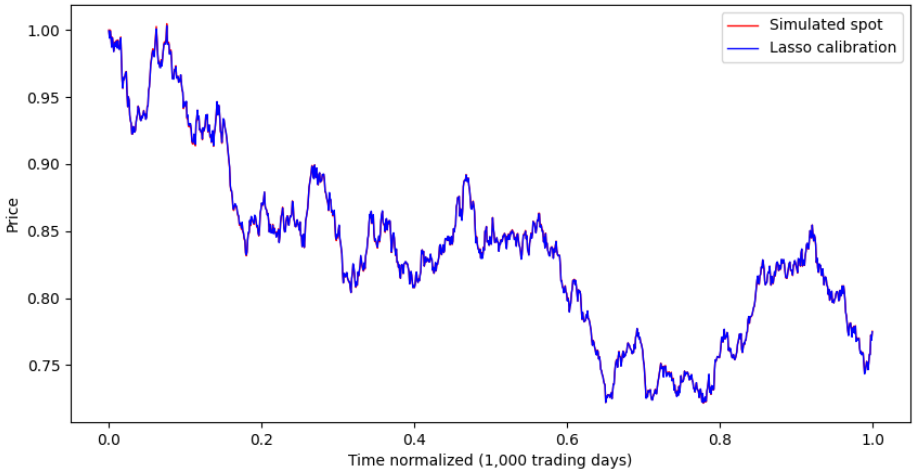

Figure 3

Receiving a Mean Squared Error of $2.76 \cdot 10^{-06}$. Throughout our approximations, we have used the python libraries iisignature and signatory.

Remark 2: One can take note that in the instance of real market data, all underlying functions remain the same, with the elimination of our simulated Geometric Brownian Motion SDE for S. One will instead directly approximate $\hat{W}_t$ from quadratic variation, or other numerical processes.

2.2 Calibration to spot volatility

Definition 2.4: For adaptation to simulating both spot price and spot volatility of our asset S, we employ the coupled SDE Heston model

$$\begin{aligned} dS_t &= \mu S_t dt + S_t \sqrt{V_t} dB_t \\ dV_t &= \kappa(\theta - V_t)dt + \sigma\sqrt{V_t} dW_t \\ dW_t &= \rho B_t + \sqrt{(1-\rho^2)}\tilde{B}_t \end{aligned} \tag{2.4}$$

Where $d[B_t, W_t]_t = \rho dt, \quad \rho \in [-1, 1]$, and primary process denoted as $\hat{X}_t := (t, B_t, W_t)$. We choose to discretize this model with an arbitrary value for S_{t_0}, V_{t_0} along the time grid $t_0, ..., t_N$.

$$S_{t_i} = S_{t_{i-1}} + \mu S_{t_{i-1}} \Delta t + \sqrt{V_{t_i}} S_t (B_{t_i} - B_{t_{i-1}})$$

$$V_{t_i} = V_{t_{i-1}} + \kappa(\theta - V_{t_{i-1}})\Delta t + \sigma\sqrt{V_{t_{i-1}}}(W_{t_i} - W_{t_{i-1}})$$

The calibration process that lays ahead is much the same as 2.1. The quadratic variation of volatility and price is used to compute

$$\hat{W}_t = \sum_{i=0}^{N-1} \frac{V_{t_{i+1}} - V_{t_i}}{\sqrt{(V_{t_{i+1}} - V_{t_i}^2}}$$

$$\hat{B}_t = \sum_{i=0}^{N-1} \frac{S_{t_{i+1}} - S_{t_i}}{\sqrt{(S_{t_{i+1}} - S_{t_i})^2}}$$

Where we once again choose $B_0 = S_0, W_0 = V_0$, and thus approximate our unknown brownian motions B_t, W_t.

Example 2.2: We take note that under the pretense of $\hat{X}_t \in \mathbb{R}^3$ consisting of 3 paths, we are now tasked with finding the corresponding appropriate level of the signature. Denoting $\hat{X}_t$ as $X^1 = t, \hat{X}^2 = \hat{B}_t, \hat{X}^3 = \hat{W}_t$, a sample of a corresponding stratonovich integral at level 3 would be

$$\langle e_{(1,2,3)}, \mathbb{X}_{r,t} \rangle = \int_0^t \left(\int_0^s \left(\int_0^r 1 \circ dX_u^1 \right) \circ dX_r^2 \right) \circ dX_s^3$$

Choosing level 3 allows for a more intricate capture of information. Our corresponding number of terms will be $T^{(3)}(\mathbb{R}^3) = 40 = \sum_{i=0}^{3} 3^i = \frac{(3^{3+1}-1)}{(3-1)}$ Where-as before, intuition allowed for the understanding that the signature of one stochastic path would follow the characteristics of that path, we now strive to convey that this signature process will consolidate both $\hat{W}_t$ and $\hat{B}_t$ all within $\mathbb{X}_{r,t}$.

Upon possession of $\mathbb{X}_t$, we thus regress the signature of $\hat{X}$ against the spot quadratic variation with our corresponding loss function from [1], to find $\ell^* \in \mathbb{R}^{d^*}$ such that

$$\ell^* \in \text{argmin}_\ell L_{\text{vol},\alpha}(\ell)$$

$$L_{\text{vol},\alpha}(\ell) := \sum_{i=1}^{N} \left(S_n(\ell)_{t_i} - \sqrt{\frac{d}{dt}[S]_{t_i}} \right)^2 + \alpha(\ell)$$

Equivalently

$$L_{\text{vol},\alpha}(\ell) := \sum_{i=1}^{N} \left(\ell_\phi + \sum_{0<|I|\leq n} \ell_I \langle e_I, \hat{\mathbb{X}}_t \rangle - \sqrt{\frac{d}{dt}[S]_{t_i}} \right)^2 + \alpha(\ell)$$

$\frac{d}{dt}[S]$ denoting the spot quadratic variation process as in definition 1.6. Where-as before in equation 2.2, we now simply modify S_{t_i} to our $\sqrt{\frac{d}{dt}[S]_{t_i}}$ term. As before, having retrieved ℓ^*, we can then calibrate to spot volatility V for **Definition 2.5:**

$$\hat{V}_{t_i} = \sum_{0\leq|I|\leq n-1} \ell_I^* \langle e_I, \mathbb{X}_{t_i} \rangle \tag{2.5}$$

As noted.

Example 2.3: For simulation, we choose

S_0	V_0	μ	σ	θ	κ	ρ
1	0.08	0.01	0.2	0.15	0.5	-0.5

Once again Setting $\alpha(\ell) = 10^{-5}$ with the signature of level 3. choosing for time increments $N = 1000$ days, we simulate a 6 hour trading day with 360 measurements evenly indexed daily. bringing total measurements of price to $360,000$. We calibrate,

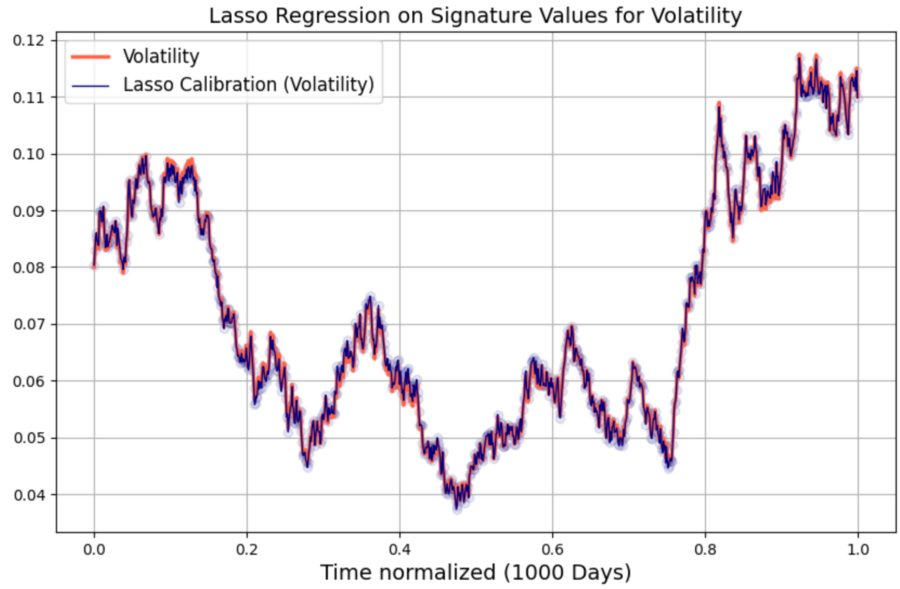

Figure 4

Receiving a Mean Squared Error of $2.33 \cdot 10^{-07}$

Remark 3: Similarly with the calibration solely to price, in an experimental setting with real market data, we are able to recover the underlying brownian motion of volatility as well as price via quadratic variation of real data. An alternative option to approximate underlying volatility V_{t_i} for a given time-window $t_i, ...t_k$ from spot price, one may choose to compute the standard deviation of the logarithmic return of price.

$$r_t = \ln\left(\frac{S_{t_i}}{S_{t_{i-1}}}\right)$$

$$\hat{V} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(r_i - \bar{r})^2}$$

For a pre-determined time window. An alternative to recovery could be performing maximum likelihood estimation.

3 Calibration to Equity Options Contract

3.1 Market Options Contracts

We now begin calibration to european call option prices, alongside calibrating to implied volatility values via the signature method.

Definition 3.1: We construct the Black-Scholes equation for a european call option contract

$$C(S_t, T, K) = N(d_+)S_t - N(d_-)Ke^{-r(T-t)} \tag{3.1}$$

$$d_+ = \frac{1}{\sigma\sqrt{T-t}}\left(\ln\left(\frac{S_t}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)\right)$$

$$d_- = d_+ - \sigma\sqrt{T-t}$$

$$\text{subject to} \quad C(S_t, K) = \max\{S_t - K, 0\}$$

In order to construct a volatility surface or volatility smile, we fix a time T_f at which we measure our contracts for our asset S, as the surface is time-variant. Thus we possess a set, either from the market, or from simulation, of N call options $C_N := \{C^*(T_1,K_1),...,C^*(T_N,K_N)\}$ with maturities T_i and strikes K_i, each measured at time T_f such that for all $T_i \in C_N, T_i \leq T_f$.

Definition 3.2: In section 2, we concerned ourselves with stochastic volatility, as defined by our pre-determined time grid $T_1,...,T_N$. In the instance of options pricing, one needs to compute for implied volatility at time T_f. Given $C(T,K)$ as the price of a call option written on the asset S, the implied volatility of $C(T,K)$ is defined as the volatility $\sigma(T,K)$ that solves the equation

$$C^{\mathrm{BS}}(T,K,\sigma(T,K)) = C(T,K) \tag{3.2}$$

where C^{BS} is our Black-Scholes option price taken at time T_f. Then, $(\sigma(T,K))_{T,K}$ is called the implied volatility surface, and $(\sigma(T,K))_K$ is called the implied volatility smile for each fixed maturity T.

We introduce the current calibration process to finding ℓ such that

$$\ell \in \mathrm{argmin}_\ell L_{\mathrm{option}}(\ell)$$

where

$$L_{\mathrm{options}}(\ell) = \sum_{i=1}^{N} g_i^* \left(C^*(T_i,K_i) - C^{\mathrm{model}}(T_i,K_i,\ell)\right)^2 \tag{3.3}$$

Seeking to minimize the difference between our forthcoming C^{model} and $C^*(T,K)$. For each respective strike K_i and maturity T_i, and g_i^* representing the vega weights, defined as

$$g_i^* = \max\left(\frac{1}{\gamma(T_t,K_t)}, 100\right)$$

$$\gamma(T_t,K_t) = S(0)\sqrt{T-t}N'(d_+) \quad N'(d_+) = \frac{e^{\frac{-d_+^2}{2}}}{\sqrt{2\pi}}$$

g_i^* is constructed such that it is proportional to the inverse of vega. Under this, options with smaller vega values will receive higher weighting, which will assist us in the instance of severe volatility, smoothing out the process. In order to achieve an acceptable error figure between the call pay-off of the Heston model and the signature payoff, a signature order of quite high level is required, due to the set of multiple strikes and maturities. To combat this burdensome computation, opting for a Monte Carlo approach is viable. The signature is taken individually on each of our samples $(X_t(\omega_i))_{t\in[0,T]}$ of underlying asset paths, compared with each strike price, **Definition 3.4** from [1]:

$$C^{\mathrm{model}}(T,K,\ell) \approx \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} (S_n(\ell)_T(\omega_i) - K)^+ \tag{3.4}$$

Once each $(X_t(\omega_i))_{t\in[0,T]}$ is simulated, we compute $\langle e_I, \hat{\mathbb{X}}_T(\omega_i)\rangle$ for all $i = 1,...,N_{MC}$ and for each multi-index I such that it's magnitude is less than n, $|I| \leq n$. We then choose to take linear combinations to compute $\langle \tilde{e}_I, \hat{\mathbb{X}}(\omega_i)\rangle$ for all $i = 1,...,N_{MC}$ and for all multi-indices $|I| \leq n-1$, described by definition 1.7 in Section 1. Since we are now dealing with multiple correlated brownian motions X^k under Monte Carlo, we have to account for these correlations. For clarification, from definition 1.7,

$$\tilde{e}_I^k = e_I \otimes e_k - \frac{\rho_{i_{|I|},k}}{2} 1_{\{i_{|I|}\neq 0\}} e_{I'} \otimes e_0 \quad \forall |I| > 0$$

The first term $e_I \otimes e^k$ extends the signature to include the interaction with the brownian motion X^k. The second term introduces a correction based on ρ, necessary to maintain accuracy. Involving e_0 accounts for proper handling outside of the non-stochastic time path. To summarize, $\langle \tilde{e}_I, \hat{\mathbb{X}}(\omega_i)\rangle$ captures the contribution of the modified tensor $\tilde{e}_I^k$, which includes corrections for the correlation, to the simulated path $\hat{\mathbb{X}}_T(\omega_i)$. Without this correction, one would experience an inaccurate monte carlo simulation, since the paths generated for S_t would not fully capture the impact of the correlations.

When computed, we receive $(S_n(\ell)_T(\omega_i))_{i=1}^{N_{MC}}$ Upon which we employ our loss function in (3.2). After receiving our ℓ^*'s from (3.2), we solve numerically for $C(T,K) = C^{\mathrm{model}}(T,K,\ell^*)$ to then survey the absolute error in basis points between the implied volatilities from the market and the model.

This simulation can be structured as:

- Receive an array of options contracts with prices: $\{C^*(T_1, K_1), ..., C^*(T_N, K_N)\}$
- Either possess, or derive the market implied volatility for each via:

$$C(S_t, T_i, K_i) = C^{\mathrm{BS}}(T_i, K_i, \sigma(T_i, K_i))$$

- For the consolidated strikes and maturities, perform afore-mentioned monte carlo simulation (3.4) on each contract
- Perform calibration argument in (3.3) to compare signature pay-off from monte carlo to market black-scholes pay-off, receive the coefficients ℓ^*.
- Solve equation (3.2) numerically given the ℓ^*'s, but rather modify to perform

$$C^*(T, K) = C^{\mathrm{model}}(T, K, \ell^*)$$

  to measure the absolute error in basis points between the market and the model.

Remark 4: We are effectively relying on the law of large numbers to achieve accuracy with Monte Carlo, our absolute difference in value from C^{model} to $C^{BS}(T, K)$ theoretically becomes smaller if we increase number of paths in N_{MC} under sig-payoff P for each respective call:

$$\frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} P_i \xrightarrow{P} E[P]$$

$$P_i = (S_n(\ell)_T(\omega_i) - K)^+$$

Where $E[P]$ represents the Black-Scholes payoff of our market options.

Example 3.1:

- $S_0 = 4288.24$
- Set of strikes $K = \{3800, 3900, 4000, 4100, 4200, 4300, 4400, 4500, 4600\}$
- Set of maturities (in years) $T = \{0.082, 0.164, 0.246, 0.493, 1, 1.717, 2.215\}$ (approximately $30 - 810$ days respectively)

We decided to import real \$SPX index European option data, dated October 2nd, 2023, furthermore, we chose to calculate implied volatility of each contract using the bisection minimzation function from the heston characteristic function in Python, since we found it difficult to download actual implied volatility data from the market. We chose a signature level of 4, and $\hat{X}_t \in \mathbb{R}^3$, and $N_{MC} = 10^5$. We received:

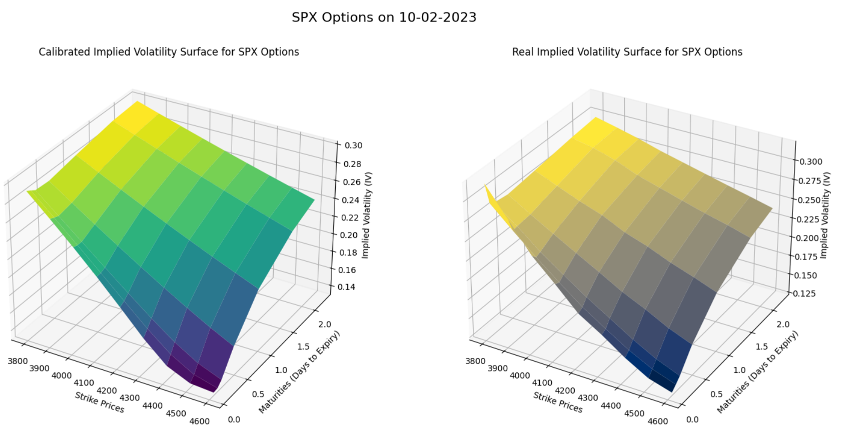

Figure 4

With I.V error:

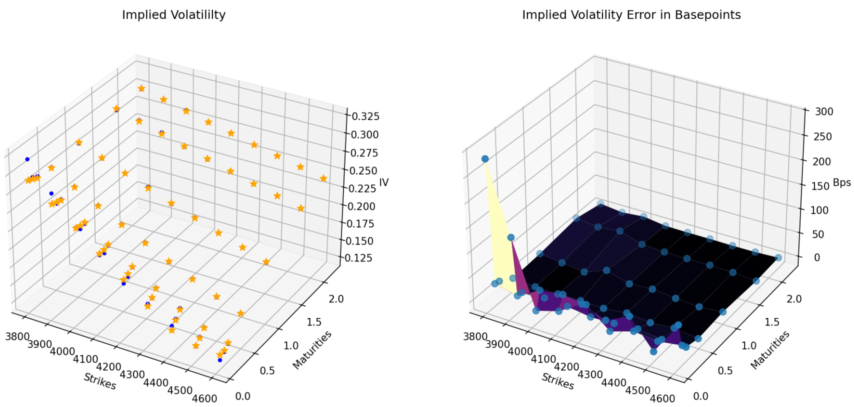

Figure 5

Remark 5: In this section we are concerned with market data, however in the instance of simulated options contracts, one could choose to model an asset S using either the SABR or Heston model, upon which they list a set of strikes and maturities, with the limitation defined above. For example we've chosen the Heston model:

$$dS_t = \mu S_t dt + S_t \sqrt{V_t} dB_t$$

$$dV_t = \kappa(\theta - V_t)dt + \sigma \sqrt{V_t} dW_t$$

Where the prices of the set

$$C_N := \{C^*(T_1, K_1), ..., C^*(T_N, K_N)\}$$

Could be determined using the Heston call function on the log-price of the asset

$$C(S_t, T, K) := S_0 P_1 - K e^{-rT} P_2$$

where P_1 and P_2 are functions involving the Heston characteristic function of the log of price

$$P_1 = \frac{\kappa - \theta}{\sigma^2}$$

$$P_2 = \frac{1}{\sigma^2}\left(\kappa - \theta - \frac{(r-\delta)}{2}\right).$$

For further information regarding this, we direct to the reference *Steven Heston, A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options* (1993).

4 W-Shaped Volatility Smile

4.1 Structure

In the financial markets there exists scenarios in which one witnesses implied volatility values which are higher for at-the-money strikes as compared to a selection of options contracts further in and out of the money. A study by *Alexiou et al.* (2021) cited that often, implied volatility curves represent this behavior directly before earnings announcements, an event often associated with high volatility. We now seek to employ our model detailed in section 3 to calibrate real market data that exhibits this behavior.

4.2 Amazon Example

For modeling this W-Shaped behavior, we will use a particularly volatile example of Amazon on the 26th of April in 2018 just before the earnings report and 1 day before the expiration. Since, we are examining options close to the expiration date (approximately 1 5 days out) and since Amazon does not pay dividends, we will model the behavior of the American options as European options for simplicity.

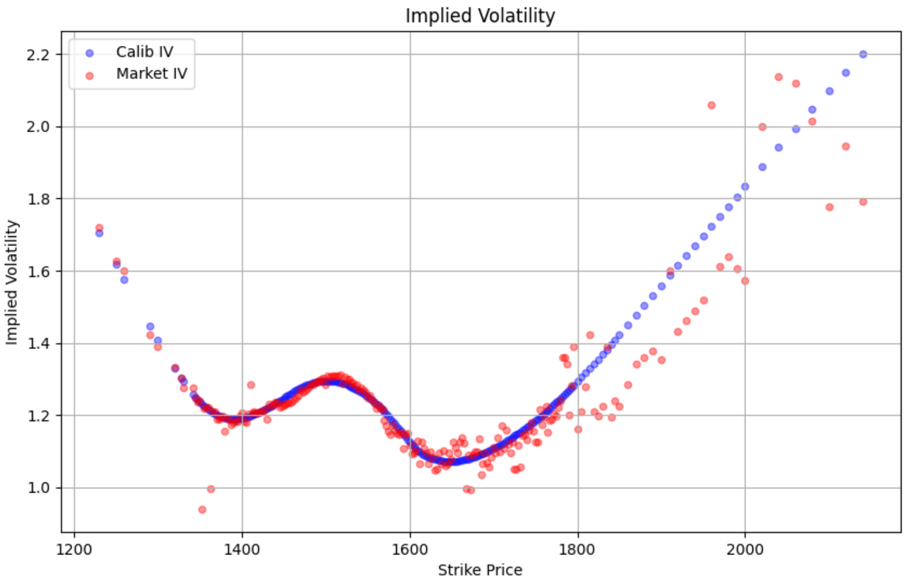

Figure 4

Here we have performed the necessary computation found in section 3 to calibrate towards market-observed implied volatilities. Contracts further in or out of the money are inherently less predictable and thus more volatile and scattered.

5 References

5.1 Direct References

[1] Cuchiero C., Gazzani, G., Svaluto-Ferro, S. (2022). *Signature-based models: theory and calibration.* From arXiv:2207.13136v1 [q-fin.MF]

[2] Ree, R. (1958). *Lie elements and an algebra associated with shuffles.* Annals of Mathematics, pages 210–220

[3] Hateley, J. *Introduction to the Tensor Product, UC Santa Barbara Mathematics*

[4] Lyons, T., Mcleod, D., A, (2024). *Signature Methods in Machine Learning,* Pages 20-23.

5.2 Supplemental References

Alexiou, L., Goyal, A., Kostakis, A., and Rompolis, L. (2021). *Pricing Event Risk: Evidence from Concave Implied Volatility Curves*

Carr, P., and Madan, D. B. (1999). *Option valuation using the fast Fourier transform, Journal of Computational Finance* Volume 2/Number 4.

Glasserman, P., and Pirjol, D. (2021). *W-shaped implied volatility curves and the Gaussian mixture model*

Chevyrev, I., and Kormilitzin, A. (2016). *A Primer on the Signature Method in Machine Learning.*

Dunn R., Hauser P., Seibold T., Gong, H. (2015). *Estimating Option Prices with Heston's Stochastic Volatility Model.*

Heston, S. (1993). *A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options.*

Gatheral, J. (2015). *Session 2: Computationally tractable stochastic volatility models.*