Sig-SDEs model for quantitative finance.

Imanol Perez Arribas  
imanol.perez@maths.ox.ac.uk  
University of Oxford  
Alan Turing Institute  
United Kingdom

Cristopher Salvi  
cristopher.salvi@maths.ox.ac.uk  
University of Oxford  
Alan Turing Institute  
United Kingdom

Lukasz Szpruch  
l.szpruch@ed.ac.uk  
University of Edinburgh  
Alan Turing Institute  
United Kingdom

ABSTRACT
Mathematical models, calibrated to data, have become ubiquitous to make key decision processes in modern quantitative finance. In this work, we propose a novel framework for data-driven model selection by integrating a classical quantitative setup with a generative modelling approach. Leveraging the properties of the signature, a well-known path-transform from stochastic analysis that recently emerged as leading machine learning technology for learning time-series data, we develop the Sig-SDE model. Sig-SDE provides a new perspective on neural SDEs and can be calibrated to exotic financial products that depend, in a non-linear way, on the whole trajectory of asset prices. Furthermore, we our approach enables to consistently calibrate under the pricing measure Q and real-world measure P. Finally, we demonstrate the ability of Sig-SDE to simulate future possible market scenarios needed for computing risk profiles or hedging strategies. Importantly, this new model is underpinned by rigorous mathematical analysis, that under appropriate conditions provides theoretical guarantees for convergence of the presented algorithms.

CCS CONCEPTS
• Mathematics of computing → Probability and statistics; • Applied computing; • Computing methodologies → Machine learning;

KEYWORDS
market simulation, pricing, signatures, rough path theory

1 INTRODUCTION
The question of finding a parsimonious model that well represents empirical data has been of paramount importance in quantitative finance. The modelling choice is dictated by the desire to fit and explain the available data, but is also subject to computational considerations. Inevitably, all models can only provide an approximation to reality, and the risk of using inadequate ones is hard to detect. A classical approach consists in fixing a class of parametric models, with a number of parameters that is significantly smaller than the number of available data points. Next, in the process called calibration, the goal is to solve a data-dependent optimization problem yielding an optimal choice of model parameters. The main challenge, of course, is to decide what class of models one should choose from. The theory of statistical learning [28] tell us that to simple models cannot fit the data, and to complex one are not expected to generalise to unseen observations. In modern machine learning approaches, one usually starts by defining a highly overparametrised model from some universality class, exhibiting a number of parameters often exceeding the number of data points, and let (stochastic) gradient algorithms find the best configuration of parameters yielding a calibrated model. In this work, we find a middle ground between the two approaches. We develop a new framework for systematic model selection that exhibits universal approximation properties, and we provide a explicit solution to the optimization used in its calibration, that completely removes the need to deploy expensive gradient descent algorithms. Importantly the class of models that we consider builds upon classical risk models that are well underpinned by research on quantitative finance.

The mathematical object at the core of this work is the expected signature of a path, whose properties are well-understood in the field of stochastic analysis. It allows to identify a linear structure underpinning the high non-linearity of the sequential data we work with. This linear structure leads to a massive speed-up of calibration, pricing, and generation of future scenarios. Our approach provides a new systematic model selection mechanism, that can also be deployed to calibrate classical non-Markovian models in a computationally efficient way. Signatures have been deployed to solve various tasks in mathematical finance, such as options pricing and hedging [22, 23], high frequency optimal execution [4, 14] and others [12, 24]. They have also been applied in several areas of machine learning [6, 16, 19, 21, 29–34].

1.1 Sig-SDE Model
Let \(X : [0, T] \rightarrow \mathbb{R}^d\) denote the price process of an arbitrary financial asset under the pricing measure \(Q\). To ensure the no-arbitrage assumption is not violated, \(X\) typically is given by the solution of the following Stochastic Differential Equation (SDE)

\[
dX_t = \Sigma_t dW_t, \quad X_0 = x, \tag{1}
\]

where \(W\) is a one-dimensional Brownian motion and \(\Sigma_t\) is an adapted process (the volatility process). Model (1) accommodates many standard risk models used e.g: the classical Black–Scholes...
model assumes that volatility is proportional to the spot price, i.e. $\Sigma_t := \sigma_t X_t$ with $\sigma \in \mathbb{R}$ constant; the local volatility model assumes that $\Sigma_t := \sigma(t, X_t) X_t$, where $\sigma(\cdot, \cdot)$ (called local volatility surface) depends on both time and spot. Hence, it is a generalisation of the Black–Scholes model; various stochastic volatility models assume that $\Sigma_t := \sigma_t$ with $\sigma_t^2$ following some diffusion process; the SABR model chooses $\Sigma_t := \sigma_t X_t^\beta$, with $\beta \in [0, 1]$ and where $\sigma_t$ follows a diffusion process.

A natural question would be whether one can find a model for the volatility process $\Sigma_t$ that is large enough to include all the classical models such as the ones mentioned above and that would allow for systematic a data driven model selection. We will require such a model to satisfy the following requirements:

1. **Universality.** The model should be able to approximate arbitrarily well the dynamics of classical models.
2. **Efficient calibration.** Given market prices for a family of options, it should be possible to efficiently calibrate the model so that it correctly prices the family of options.
3. **Fast pricing.** Ideally, it should be possible to quickly price (potentially exotic) options under the model without using Monte Carlo techniques.
4. **Efficient simulation.** Sampling trajectories from the model should be computationally cheap and efficient.

An example of a model that satisfies point 1. above is a **neural network model** where the volatility process $\Sigma_t$ is approximated by a neural network $NN^\theta(t, (W_t)_{t \in [0, T]})$ with parameters $\theta$. Such a model would be able to approximate a rich class of classical models. However, the calibration and pricing of such models would involve performing multiple Monte Carlo simulations on each epoch, which might be expensive if done naively. See however, [7, 10]. The aim of this paper is to propose a model for asset price dynamics that, we believe, satisfies all four points above. Our technique models the volatility process $\Sigma_t$ as

$$\Sigma_t = \langle \sigma^N, \tilde{W}_{0,t} \rangle$$

where $\sigma^N$ is the model parameters and $\tilde{W}_{0,t}$ is the signature (c.f definition 2.6) of the stochastic process $W_t := (t, W_t)$. The motivation for choosing the signature as the main building block of this paper is anchored in a very powerful result for universal approximation of functions based on the celebrated Stone–Weierstrass Theorem that we present next in an informal manner (for more technical details see [8, Proposition 3]).

**Theorem 1.1.** Consider a compact set $K$ of continuous $\mathbb{R}^d$-valued paths. Denote by $S$ the function that maps a path $X$ from $K$ to its signature $\mathcal{X}$. Let $f : K \to \mathbb{R}$ be any continuous function. Then, for any $\epsilon > 0$ and any path $X \in K$, there exists a linear function $f^{\epsilon}$ acting on the signature such that

$$||f(X) - \langle f^{\epsilon}, \mathcal{X} \rangle||_\infty < \epsilon$$

In other words, any continuous function on a compact set of paths can be uniformly well approximated by a linear combination of terms of the signature. This universal approximation property is similar to the one provided by Neural Networks (NN). However, as we will discuss below, NN models depend on a very large collection of parameters that need to be optimized via expensive back-propagation-based techniques, whilst the optimization needed in our Sig-SDE model consists of a simple linear regression on the terms of the signature. In this way, the signature can be thought of as a feature map for paths that provides a linear basis for the space of continuous functions on paths. In the setting of SDEs, sample paths are Brownian and solutions are images of these sample trajectories by a continuous functions that one wishes to approximate from a set of observations. Our Sig-SDE model will rely upon the universality of the signature to approximate such functions acting on Brownian trajectories. Importantly, the signature of a realisation of a semimartingale provides a unique representation of the sample trajectory [2, 13]. Similarly, the expected signature – i.e. the collection of the expectations of the iterated integrals – provides a unique representation of the law of the semimartingale [5].

Note that model calibration is an example of generative modelling [11, 18]. Indeed, recall that if one knew prices of traded liquid derivatives, then one can approximate the pricing measure from market data [3, 22]. We denote this by $Q^{\text{cal}}$. We know that when equation (1) admits a strong solution then there exists a measurable map $G : \mathbb{R} \times C([0, T]) \to C([0, T])$ such that

$$X = G(x, (W_t)_{t \in [0, T]}$$

as shown in [15, Corollary 3.23]. If $G_t$ denotes the projection of $G$ given by $X_t := G_t(x, (W_t)_{t \in [0, T]})$, then one can view (1) as a generative model that maps $\mu_0$ supported on $\mathbb{R}^d$ into $(G_t)_{t \in [0, T]} = Q^{\text{cal}}_t$. Note that by construction $G$ is a casual transport map i.e a transport map that is adapted to the filtration $\mathcal{F}_t$ [1]. In practice, one is interested in finding such a transport map from a family of parametrised functions $G^\theta$. One then looks for a $\theta$ such that $G^\theta \mu_0$ is a good approximation of $Q^{\text{cal}}$ with respect to a metric specified by the user. In this paper the family of transport maps $G^\theta$ is given by linear functions on signatures (or linear functionals below).

## 2. NOTATION AND PRELIMINARIES

We begin by introducing some notation and preliminary results that are used in this paper.

### 2.1 Multi-indices

**Definition 2.1.** Let $d \in \mathbb{N}$. For any $n \geq 0$, we call an $n$-dimensional $d$-multi-index any $n$-tuple of non-negative integers of the form $K = (k_1, \ldots, k_n)$ such that $k_i \in \{1, \ldots, d\}$ for all $i \in \{1, \ldots, n\}$. We denote its length by $|K| = n$. The empty multi-index is denoted by $\emptyset$. We denote by $I_d$ the set of all $d$-multi-indices, and by $I_d^\epsilon \subset I_d$ the set of all $d$-multi-indices of length at most $n \in \mathbb{N}$.

**Definition 2.2 (Concatenation of multi-indices).** Let $I = (i_1, \ldots, i_p)$ and $J = (j_1, \ldots, j_q)$ be any two multi-indices in $I_d$. Their concatenation product $\otimes$ as the multi-index $I \otimes J = (i_1, \ldots, i_m, j_1, \ldots, j_n) \in I_d$.

**Example 2.3.**

1. $(1, 3) \otimes (2, 2) = (1, 3, 2, 2)$.
2. $(2, 1, 3) \otimes (1) = (2, 1, 3, 1)$.
3. $(2, 2) \otimes \emptyset = (2, 2)$.

### 2.2 Linear functionals

**Definition 2.4 (Linear functional).** For a given $d \geq 1$, a linear functional is a (possibly infinite) sequence of real numbers indexed by
multi-indices in $I_d$ of the following form

$$F = \{ F^K \in \mathbb{R} : K \in I_d \}.$$  

(5)

We note that a multi-index $K \in I_d$ is always a linear functional. Both concatenation $\otimes$ and $\circ$ can be extended by linearity to operations on linear functionals. We will now define two basic operations on linear functionals that will be used throughout the paper.

**Definition 2.5.** For any two linear functionals $F, G$ and any real numbers $\alpha, \beta \in \mathbb{R}$ define

$$\alpha F + \beta G = \{ \alpha F^K + \beta G^K \in \mathbb{R} : K \in I_d \}$$  

and

$$\langle F, G \rangle = \sum_{K \in I_d} F^K G^K \in \mathbb{R}$$  

(7)

### 2.3 Signatures

**Rough paths** theory can be briefly described as a non-linear extension of the classical theory of controlled differential equations which is robust enough to allow a deterministic treatment of stochastic differential equations controlled by much rougher signals than semi-martingales [26].

**Definition 2.6 (Signature).** Let $X : [0, T] \to \mathbb{R}^d$ be a continuous semimartingale. The signature of $X$ over a time interval $[s, t] \subseteq [0, T]$ is the linear functional $\mathcal{X}_{s,t} := (\mathcal{X}_{s,t}^a \in \mathbb{R} : K \in I_d)$, such that $\mathcal{X}_{s,t}^a = 1$ and so that for any $n \geq 1$ and $K = \overrightarrow{K} \otimes a \in I_d^n$, with $a \in \{1, \ldots, d\}$ and $\overrightarrow{K} \in I_d^{n-1}$ we have

$$\mathcal{X}_{s,t}^K := \int_s^t \mathcal{X}_{s,u}^{\overrightarrow{K}} \circ dX_u^{(a)}$$  

(8)

where the integral is to be interpreted in the Stratonovich sense.

**Example 2.7.** Let $X : [0, T] \to \mathbb{R}^2$ be a semimartingale.

1. $\mathcal{X}_{s,t}^{(1)} = X_t^{(1)} - X_s^{(1)}$.
2. $\mathcal{X}_{s,t}^{(1,2)} = \int_s^t X_u^{(1)} \circ dX_u^{(2)}$.
3. $\mathcal{X}_{s,t}^{(2,2)} = \frac{1}{2} (X_t^{(2)} - X_s^{(2)})^2$.

A more detailed overview of signatures is included in Appendix A.

### 3 SIGNATURE MODEL

In this section we define the **Signature Model** for asset price dynamics that we propose in this paper. The goal is to approximate the volatility process $\Sigma_t$ (that is a continuous function on the driving Brownian path) by a linear functional on the signature of the Brownian path.

**Definition 3.1 (Signature Model).** Let $W$ be a one-dimensional Brownian motion. Let $N \in \mathbb{N}$ be the *order* of the Signature Model. The Signature Model of parameter $\ell = \{ \ell^K \in I_d^N \}$ is given by $\Sigma_t := (\ell, \mathcal{W}_{0,t})$, where $\mathcal{W}$ denotes the signature of $W$-additive. In other words, the asset price dynamics are given by

$$dX_t = (\ell, \mathcal{W}_{0,t}) dW_t, \quad X_0 = \alpha \in \mathbb{R}.$$  

(9)

We note that the Signature Model has two components: the hyperparameter $N \in \mathbb{N}$, and the model parameter $\ell$. Intuitively, the hyperparameter $N$ plays a similar role to the width of a layer in a neural network. The larger this value is, the richer the range of market dynamics the Signature Models can generate. Once the value of $N$ is fixed, the challenge is to find a suitable model parameter $\ell$. Again, in analogy with neural networks, $\ell$ plays the role of the weights of the network.

The Signature Model possesses the *universality property*, in the sense that given a classical model, there exists a Signature Model that can approximate its dynamics to a given accuracy [20].

We show in the upcoming Sections 5-7 that (a) the Signature Model is efficient to simulate, (b) it is efficient to calibrate, and (c) exotic options can be priced fast under the Signature Model.

**Remark 1.** The Signature Model introduced in Definition 3.1 assumes that the source of noise (i.e. the Brownian motion $W$) is one-dimensional. This was done for simplicity, but the authors would like to emphasise that the model generalises in a straightforward way to multi-dimensional Brownian motion.

### 4 NUMERICAL EXPERIMENTS

We now demonstrate the feasibility of our methodology as outlined in Sections 5-7. Throughout this section, we work with the Signature Model

$$dX_t = (\ell, \mathcal{W}_{0,t}) dW_t, \quad X_0 = 1$$

with $\ell = \{ \ell^K \in I_d^N \}$. We fix $N = 4$. Therefore, the model has $1 + 2 + 2^2 + 2^3 + 2^4 = 31$ parameters that need to be calibrated. We also fix the terminal maturity $T = 1$.

In this section we will show experiments for the calibration of the model, pricing of options under the signature model and simulation. Sections 5-7 will then include the technical details of how calibration, pricing and simulation of signatures model are done.

#### 4.1 Calibration

We assume that the family of options available on the market are a mixture of vanilla and exotic options, given as follows:

- Vanilla call options with strikes $K = 0.5, 0.6, \ldots, 1.1$, and maturities $t = 0.4, 0.45, 0.5, \ldots, 0.9, 0.95$:

  $$\Phi := \max(X_t - K, 0).$$

- Variance options with strikes $K = 0.01, 0.015, \ldots, 0.035, 0.04$ and maturities $t = 0.4, 0.45, 0.5, \ldots, 0.9, 0.95$:

  $$\Phi := \max(\langle X \rangle_t - K, 0).$$

where $\langle X \rangle_t$ is the quadratic variation of $X$.

- Down-and-Out barrier call options with maturity $1$, strikes $K = 0.9, 0.92, 0.94, \ldots, 1.01, 1.03$ and barrier levels $L = 0.6, 0.62, 0.64, \ldots, 0.88, 0.9$:

  $$\Phi := \begin{cases} \max(X_t - K, 0) & \text{if} \ \min_{x \in [a,t]} X_x > L \\ 0 & \text{else.} \end{cases}$$

The option prices are generated from a Black-Scholes model with volatility $\sigma = 0.2$.

$$dX_t = \alpha X_t dW_t.$$  

The optimisation (14) was then solved to calibrate the model parameters $\ell = \{ \ell^K \in I_d^N \}$.

Figure 1 shows the absolute error between the real option prices and the option prices of the calibrated model, for the different option types.
4.2 Simulation

Once the Signature Model has been calibrated to the available option prices, we can use Algorithm 1 to simulate realisations of the calibrated Signature Model. Figure 2 shows 1,000 realisations of the Signature Model.

4.3 Pricing

We will now use the calibrated Signature Model to price a new set of options that was not used in the calibration step. This set of options consists of Down-and-In barrier put options with barriers levels $L = 0.7, 0.71, \ldots, 0.81, 0.82$ and strikes $K = 0.9, 0.92, \ldots, 1.01, 1.03$:

$$
\Phi := \begin{cases} 
\max(K - X_t, 0) & \text{if } \min_{s \in [0,t]} X_s < L \\
0 & \text{else} 
\end{cases}
$$

Figure 3 shows the absolute error of the prices under the Signature Model, compared to the real prices. As we see, the calibrated model is able to generate accurate prices for these new exotic options. The error is highest when the barrier is close to the strike price, as expected.

5 SIMULATION

This section will address the question of simulation efficiency of Signature Models. We begin by stating the following two results. The first result rewrites the differential equation (9) solely in terms of the lead-lag signature of the Brownian motion, $\mathcal{W}_{0,t}^{LL}$. Here $\mathcal{W}_{0,t}^{LL}$ denotes the lead-lag transformation of $\mathcal{W}$, see Appendix B. We use the lead-lag transformation because it allows us to rewrite Itô integrals as certain Stratonovich integrals, which in turn can be written as linear functions on signatures. The second result guarantees that the computational cost of computing $\mathcal{W}_{0,t}^{LL}$ is the same as the cost of computing $\{\mathcal{W}_{0,s}^{LL} ; 0 \leq s \leq t\}$. These two results lead to Algorithm 1, which provides an efficient algorithm to sample from a Signature Model.
Theorem 5.2 (Chen's identity, [25, Theorem 2.12]).

Then, for each multi-index \( \ell \in \mathbb{N}^d \) with \( \ell \neq \theta \), the signature \( \mathcal{W}^{LL}_{\ell,0} \) at time 0 is given by

\[
\mathcal{W}^{LL}_{\ell,0} = \mathcal{W}^{LL}(\ell) = \left\langle \mathcal{W}^{LL}(\ell) : K \in \mathbb{N}^{n+1} \right\rangle
\]

where \( \mathcal{W}^{LL}(\ell) \) denotes the lead-lag transformation, introduced in Definition B.1, of the 2-dimensional process \( \mathbb{W} = (t, W_t) \).

PROPOSITION 5.1 ([22, Lemma 3.11]). Let \( X \) follow a Signature Model with parameter \( \ell = \{\ell(K) : K \in \mathbb{N}^n\} \). Then, \( X \) is given by

\[
X_t = \left\langle x(\theta) + \ell \otimes (4), \mathcal{W}^{LL}_{0,t} \right\rangle
\]

where \( \ell \otimes (4) = \{K \otimes (4) : K \in \ell\} \), \( x = X_0 \in \mathbb{R} \), and \( \mathcal{W}^{LL} \) denotes the lead-lag transformation, introduced in Definition B.1, of the 2-dimensional process \( \mathbb{W} = (t, W_t) \).

THEOREM 5.2 (CHEN'S IDENTITY, [25, Theorem 2.12]). Let \( 0 \leq s \leq t \). Then, for each multi-index \( K \in \mathbb{I}^d \) we have

\[
\mathcal{W}^{LL}(K) = \sum_{0 \leq j < s} \mathcal{W}^{LL}(j) \mathcal{W}^{LL}_{0,t} = \left\langle \mathcal{W}^{LL}(j), \mathcal{W}^{LL}_{0,t} \right\rangle
\]

where for any multi-index \( K \in \mathbb{I}^d \) we used the notation \( \mathcal{W}^{LL}(K) = (K, \mathcal{W}^{LL}_{0,t}) \).

These two results lead to Algorithm 1. We note there are a number of publicly available software packages to compute signatures, such as esig \(^1\), iisignature \(^2\) and sigatory \(^3\).

6 PRICING

This section will show that exotic options can be priced fast under a Signature Model. This will be done via a two step procedure. First, it was shown in [22, 23] that prices of exotic options can be approximated with arbitrary precision by a special class of payoffs called signature payoffs, defined below. Hence, we will assume that the exotic option to be priced is a signature payoff, defined as follows.

\(^1\)https://pypi.org/project/esig/
\(^2\)https://github.com/bottler/iisignature, [27]
\(^3\)https://github.com/patrick-kidger/sigatory, [17]

Definition 6.1 (Signature payoffs). A signature payoff of maturity \( T > 0 \) and parameter \( f = \{f(K) : K \in \mathbb{N}^n\} \) is a payoff that pays at time \( T \) an amount given by \( f(\mathcal{W}_{0,T}) \).

Second, the price of a signature payoff is \( (f, \mathbb{E}[\mathcal{W}_{0,T}]) \). To price a signature payoff, all we need is \( \mathbb{E}[\mathcal{W}_{0,T}] \), which doesn’t depend on the signature payoff itself. In particular, it may be reused to price other signature payoffs.

We now explicitly derive the expected signature \( \mathbb{E}[\mathcal{W}_{0,T}] \) in terms of the model parameters and the expected signature of the lead-lag Brownian motion \( \mathbb{E}[\mathcal{W}^{LL}_{0,T}] \).

PROPOSITION 6.2. Let \( X \) be a Signature Model of order \( N \in \mathbb{N} \) with parameter \( \ell = \{\ell(K) : K \in \mathbb{I}^d_N\} \). Consider the following linear functionals \( P_1 = (1) \) and \( P_2 = \ell \otimes (4) \). Consider any multi-index \( I = (i_1, \ldots, i_n) \in \mathbb{I}^d_N \) such that \( n \leq N \). Then

\[
\mathbb{E}[\mathcal{W}_{s,t}] = (C_I(f), \mathcal{W}^{LL}_{0,t})
\]

where \( C_I(f) \) is given explicitly in closed-form by

\[
C_I(f) = (\ldots (P_{i_1} \circ P_{i_2} \circ \ldots P_{i_n}) \ldots)
\]

PROOF. By Proposition 5.1 we know that if \( X \) follows a Signature Model with parameter \( \ell = \{\ell(K) : K \in \mathbb{I}^d_N\} \) then

\[
X_t = \left\langle x(\theta) + \ell \otimes (4), \mathcal{W}^{LL}_{0,t} \right\rangle
\]

Let \( I = (i_1, \ldots, i_n) \) be any multi-index in \( \mathbb{I}^d_N \) such that \( n \leq N \). If \( n = 1 \) then \( I = (i_1) \) and we necessarily one of the following two options must hold

- If \( i_1 = 1 \) then \( \mathbb{E}[\mathcal{W}_{s,t}] = (\mathbb{E}[\mathcal{W}_{s,t}], 1) = (P_{i_1}, \mathcal{W}^{LL}_{0,t}) \)
- If \( i_1 = 2 \) then \( \mathbb{E}[\mathcal{W}_{s,t}] = (X_t - X_s, (\ell \otimes (4)), \mathcal{W}^{LL}_{0,t}) = (P_2, \mathcal{W}^{LL}_{0,t}) \).

Hence the statement holds for \( n = 1 \). Let’s assume by induction that the statement holds for any \( 1 \leq n \leq N \). We write \( I = f \otimes (i_n) \) with \( i_n \in \{1, 2\} \) and \( J = (i_1, \ldots, i_{n-1}) \in \mathbb{I}^{n-1}_N \). Clearly \( |(i_n)|, |J| < n \), therefore by induction hypothesis

\[
\mathbb{E}[\mathcal{W}_{s,t}] = \left\langle C_I(f), \mathcal{W}^{LL}_{0,t} \right\rangle = \left\langle P_{i_n}, \mathcal{W}^{LL}_{0,t} \right\rangle
\]

and

\[
\mathbb{E}[\mathcal{W}_{s,t}] = \left\langle C_I(f), \mathcal{W}^{LL}_{0,t} \right\rangle = \left\langle (P_{i_1} \circ \ldots (P_{i_{n-1}}) \ldots P_{i_n}), \mathcal{W}^{LL}_{0,t} \right\rangle
\]

By definition of the signature (see 2.6) we know that

\[
\mathbb{E}[\mathcal{W}_{s,t}] = \int_s^t \mathbb{E}[\mathcal{W}_{s,u}] \, du
\]

which concludes the induction. \( \square \)
7 CALIBRATION

We will now address the task of calibrating a Signature Model. We assume that the market has a family of options \( \{ \Phi_i \}_{i=1}^n \) whose market prices \( \{ p_i \}_{i=1}^n \) are observable. Typically \( \{ \Phi_i \}_{i=1}^n \) will contain vanilla options, together with some exotic options such as various variance or barrier products. Fix \( N \in \mathbb{N} \) be the order of the Signature Model. The challenge here is to find the model parameter \( \ell = \{ \ell(K) : K \in \mathcal{I}^N \} \) that best fits the data, in the sense that the prices of \( \Phi_i \), under the Signature Model with parameter \( \ell \), are approximately given by the observed market prices \( p_i \).

Following Section 6, we assume that the options \( \Phi_i \) are given by signature options. Therefore, we assume that we can write \( \Phi_i \) by

\[
\Phi_i = \langle q_i, \tilde{x}_0, T \rangle, \quad q_i = \{ q_i^K : K \in \mathcal{I}^N \}.
\]

The minimisation problem we aim to solve now is the following:

\[
\min_{\ell = \{ \ell(K) : K \in \mathcal{I}^N \}} \sum_{i=1}^n \left( \langle q_i, \mathbb{E}[\tilde{x}_0, T] \rangle - p_i \right)^2.
\]  

(14)

where \( \mathbb{E}[\tilde{x}_0, T] \) is the expected signature of the Signature Model with parameter \( \ell = \{ \ell(K) : K \in \mathcal{I}^N \} \).

By Proposition 6.2, the price of \( \Phi_i \), which is given by \( \langle q_i, \mathbb{E}[\tilde{x}_0, T] \rangle \), can be written as a polynomial on \( \ell(K) \). Hence, the optimisation (14) is rewritten as a minimisation of a polynomial of variables \( \ell(K) \), for \( K \in \mathcal{I}^N \).

If the number of parameters \( \ell(K) \) is large compared to the number of available option prices, the optimisation problem might be over-parametrised and there will be multiple solutions to (14). In this case, we are in the robust finance setting where there are multiple equivalent martingale measures that fit the data. If the number of parameters \( \ell(K) \) is small, however, we are in the setting of classical mathematical finance modeling and there will in general be a unique solution to (14).

8 CONCLUSION

In this paper we have proposed a new model for asset price dynamics called the signature model. This model was develop with the objective of satisfying the following properties:

1. Universality.
2. Efficiency of calibration to vanilla and exotic options.
3. Fast pricing of vanilla and exotic options.
4. Efficiency of simulation.

Due to the rich properties of signatures, the signature model satisfies all four properties and is, therefore, capable of generating realistic paths without sacrificing the computational feasibility of calibration, pricing and simulation. Although this paper has focused on the risk-neutral measure \( \mathbb{Q} \), it can also be used to learn the real-world measure \( \mathbb{P} \). One would first calibrate to the risk-neutral measure \( \mathbb{Q} \) and then learn the drift.

ACKNOWLEDGMENTS

This work was supported by The Alan Turing Institute under the EPSRC grant EP/N510129/1.

REFERENCES

[1] Beatrice Acciaio, Julio Backhoff-Veraguas, and Anastasia Zalashko. 2019. Causal optimal transport and its links to enlargement of filtrations and continuous-time stochastic optimization. Stochastic Processes and their Applications (2019).
[2] Horst R Boedihardjo, Xi Geng, Terry Lyons, and Danya Yang. 2016. The signature of a rough path: uniqueness. Advances in Mathematics 293 (2016), 720–737.
[3] Douglas T Breeden and Robert H Litzenberger. 1978. Prices of state-contingent claims implicit in option prices. Journal of Business (1978), 621–651.
[4] Álvaro Cartea, Imanol Perez Arribas, and Leandro Sánchez-Betancourt. 2020. Optimal Execution of Foreign Securities: A Double-Execution Problem with Signatures and Machine Learning. Available at SSRN (2020).
[5] Ilya Chevyrev, Terry Lyons, et al. 2016. Characteristic functions of measures on geometric rough paths. The Annals of Probability 44, 6 (2016), 4049–4082.
[6] Ilya Chevyrev and Harald Oberhauser. 2018. Signature moments to characterize laws of stochastic processes. arXiv preprint arXiv:1810.10971 (2018).
[7] Christa Cuchiero, Wahid Khosrawi, and Josef Teichmann. 2020. A generative adversarial network approach to calibration of local stochastic volatility models. arXiv preprint arXiv:2005.02505 (2020).
[8] Adeline Ferneman. 2019. Embedding and learning with signatures. arXiv preprint arXiv:1911.13211 (2019).
[9] Guy Fint, Ben Hambly, and Terry Lyons. 2016. Discretely sampled signals and the rough Hoeff process. Stochastic Processes and their Applications 126, 9 (2016), 2593–2614.
[10] M. Siska D. Szpruch L. Gierjatowicz P., Sabate-Vidales. 2020. Robust Pricing and Hedging with neural SDEs. to appear (2020).
[11] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In Advances in neural information processing systems. 2672–2680.
[12] László Gergely Gyurics, Terry Lyons, Mark Kontokwioski, and Jonathan Field. 2013. Extracting information from the signature of a financial data stream. arXiv preprint arXiv:1307.7244 (2013).
[13] Ben Hambly and Terry Lyons. 2010. Uniqueness for the signature of a path of bounded variation and the reduced path group. Annals of Mathematics (2010), 109–167.
[14] Jasdeep Kalsi, Terry Lyons, and Imanol Peretx Arribas. 2020. Optimal execution with rough path signatures. SIAM Journal on Financial Mathematics 11, 2 (2020), 470–493.
[15] Ioannis Karatzas and Steven Shreve. 2012. Brownian motion and stochastic calculus. Vol. Vol. 113. Springer Science & Business Media.
[16] Patrick Kidger, Patric Bonnier, Imanol Perez Arribas, Cristopher Salvi, and Terry Lyons. 2019. Deep Signature Transforms. In Advances in Neural Information Processing Systems. 3099–3109.
[17] Patrick Kidger and Terry Lyons. 2020. Signatory: differentiable computations of the signature and logsignature transforms, on both CPU and GPU. arXiv preprint arXiv:2001.00706 (2020).
[18] Diederik F Kingma and Max Welling. 2013. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013).
[19] Franz J Kiraly and Harald Oberhauser. 2016. Kernels for sequentially ordered data. arXiv preprint arXiv:1601.08169 (2016).
[20] Daniel Levin, Terry Lyons, and Hao Ni. 2013. Learning from the past, predicting the statistics for the future, learning an evolving system. arXiv preprint arXiv:1309.0260 (2013).
[21] Chenyang Li, Xin Zhang, and Lianwen Jin. 2017. Lipsnet: A novel log path signature feature based hand gesture recognition framework. In Proceedings of the IEEE International Conference on Computer Vision Workshops. 651–659.
[22] Terry Lyons, Sina Nejad, and Imanol Perez Arribas. 2019. Nonparametric pricing and hedging of exotic derivatives. arXiv preprint arXiv:1905.00711 (2019).
[23] Terry Lyons, Sina Nejad, and Imanol Perez Arribas. 2020. Numerical Method for Model-free Pricing of Exotic Derivatives in Discrete Time Using Rough Path Signatures. Applied Mathematical Finance (2020), 1–15.
[24] Terry Lyons, Hao Ni, and Harald Oberhauser. 2014. A feature set for streams and an application to high-frequency financial tick data. In Proceedings of the 2014 International Conference on Big Data Science and Computing. 1–8.
[25] Terry J Lyons. 1998. Differential equations driven by rough signals. Revista Matemática Iberoamericana 14, 2 (1998), 215–310.
[26] Terry J Lyons, Michael Caruana, and Thierry Lévy. 2007. Differential equations driven by rough paths. Springer.
[27] Jeremy F Reizenstein and Benjamin Graham. 2020. Algorithm 1004: The iisignature: Efficient calculation of iterated-integral signatures and log signatures. Annals of Mathematics (2020).
[28] Vladimir Vapnik. 2013. The nature of statistical learning theory. Springer science & business media.
[29] Zecheng Xie, Zenghui Sun, Lianwen Jin, Hao Ni, and Terry Lyons. 2017. Learning spatial-semantic context with fully convolutional recurrent network for online handwritten chinese text recognition. IEEE transactions on pattern analysis and machine intelligence 40, 8 (2017), 1903–1917.
A OVERVIEW OF SIGNATURES

In this section we state some of the main properties of signatures that are used in this paper.

Definition A.1 (Shuffle of multi-indices). For any two multi-indices $I, J \in I_d$ and 1-dimensional multi-indices $a, b \in I_1 = \{1, \ldots, d\}$ we define the shuffle product $\omega$ recursively as follows:

$$\omega a = 1 \omega = I$$ (15)

and

$$(I \otimes a)\omega (j \otimes b) = ((I \otimes a)\omega j) \otimes b \oplus (\omega (j \otimes b)) \oplus a$$ (16)

Example A.2. We have the following examples for $I_d$:

1. $(1, 2)\omega (3) = (1, 2, 3) + (1, 3, 2) + (3, 1, 2)$.
2. $(1, 2)\omega (3, 4) = (2, 1, 2, 4, 3) + (2, 1, 4, 2, 3) + (2, 1, 4, 2)$.
3. $(2)\omega (2) = (2, 1)$.
4. $(4)\omega (2, 1) = (2, 1)$.

Proposition A.3 (Shuffle identity). Let $X : [0, T] \rightarrow \mathbb{R}^d$ be a continuous semimartingale. For any two multi-indices $I, J \in I_d$ the following identity on the Signature of $X$ holds

$$\langle I \omega, X_{s,t} \rangle = \langle I, X_{s,t} \rangle \cdot \langle J, X_{s,t} \rangle = \mathcal{X}^{(I)}_{s,t} \cdot \mathcal{X}^{(J)}_{s,t}$$ (17)

Proof. Theorem 2.15 in [26].

Proposition A.4 (Uniqueness of the Signature). Let $X : [0, T] \rightarrow \mathbb{R}^d$, $Y : [0, T] \rightarrow \mathbb{R}^d$ be two continuous semimartingales. Then

$$\forall t \in [0, T], X_t = Y_t \iff \forall K \in I_d, \mathcal{X}^{(K)}_{s,t} = \mathcal{Y}^{(K)}_{s,t}$$ (18)

Proof. See main result in [13].

Proposition A.5 (Factorial decay). Given a semimartingale $X : [0, T] \rightarrow \mathbb{R}^d$, for any time interval $[s, t] \subset [0, T]$ and any multi-index $K \in I_d$ such that $|K| = n$ we have

$$\mathcal{X}^{(K)}_{s,t} = O\left(\frac{1}{n!}\right)$$ (19)

Proof. Proposition 2.2 in [26].

Definition A.6. For a given time interval $[0, T]$ we call a continuous, surjective, increasing function $\psi : [0, T] \rightarrow [0, T]$ a time-reparametrization.

Proposition A.7 (Invariance to time-reparametrizations). Let $X : [0, T] \rightarrow \mathbb{R}^d$ be a semimartingale and $\psi : [0, T] \rightarrow [0, T]$ be a time-reparametrization. Then the Signature of $X$ has the following invariance property

$$\mathcal{X}_{s,t} = \mathcal{X}_{\psi(s),\psi(t)} \quad \forall s, t \in [0, T]$$ (20)

Definition A.8 (Half-Shuffle). Let $F$ and $G$ be any two linear functionals. We define their half-shuffle product $\triangleright$ on $\mathcal{X}_{s,t}$ as the following (Stratonovich) iterated integral on the real line

$$F \triangleright G, X_{s,t} = \int_s^t \langle F, X_{s,u} \rangle \circ d \langle G, X_{s,u} \rangle$$ (21)

Let $\mathbb{B}$ be a 2-dimensional Brownian motion, defined for example on $[0, 1]$. Consider two linear functionals $F = \{F(K) : K \in I_d\}$ and $G = \{G(K) : K \in I_d\}$ defined as

$$F(K) = \begin{cases} 1 & \text{if } K = (1, 2) \\ 0 & \text{otherwise} \end{cases}$$ (22)

and

$$G(K) = \begin{cases} 1 & \text{if } K = (2, 1) \\ 0 & \text{otherwise} \end{cases}$$ (23)

Then the following quantity

$$\mathcal{I}_{s,t} = \frac{1}{2} \{F \triangleright G - G \triangleright F, \mathbb{B}_{s,t}\}$$ (24)

is the Levy area of the Brownian motion $\mathbb{B}$ on $[s, t] \subset [0, 1]$.

A.0.1 Expected signature. We will now define the expected signature of a semimartingale.

Definition A.9 (Expected signature). Let $X : [0, T] \rightarrow \mathbb{R}^d$ be a continuous semimartingale, and let $\mathcal{X}_{s,t} = \{\mathcal{X}^{(K)}_{s,t} : K \in I_d\}$ be its signature. The expected signature of $X$ is defined by

$$\mathbb{E}[\mathcal{X}_{s,t}] := \{\mathbb{E}[\mathcal{X}^{(K)}_{s,t}] : K \in I_d\}.$$ (25)

The expected signature – i.e. the expectation of the iterated integrals (8) – behaves analogously to the moments of random variables, in the sense that under certain assumptions it characterises the law of the stochastic process:

Theorem A.10 ([13]). Let $X : [0, T] \rightarrow \mathbb{R}^d$ be a semimartingale. Then, under certain assumptions (see [5]) the expected signature $\mathbb{E}[\mathcal{X}_{0,T}]$ characterises the law of $X$.

B TIME AND LEAD-LAG TRANSFORMATION

The invariance of the signature of a semimartingale to time-reparametrizations allows to handle irregularly sampled sample paths (prices etc.) by completely eliminating the need to retain information about the original time-parametrization. Nonetheless, for the pricing of many options, especially ones resulting from payoffs calculated pathwise (such as integrals for American options), the time represents an important information that we are required to retain. To do so it suffices to augment the state space of the input semimartingale $X$ by adding time $t$ as an extra dimension to get $X^\text{add-time} = (t, X_t)$.

We report another basic transformation that can be applied to semimartingales and that will be useful in the sequel of the paper: the lead-lag transformation. This transformation allows us to write Itô integrals as linear functions on the signature of the lead-lag transformed path.
Definition B.1 (Lead-lag transformation). Let \( Z : [0, T] \to \mathbb{R}^d \) be a semimartingale. For each partition \( D = \{ t_i \}_i \subset [0, T] \) of mesh size \(|D|\), define the piecewise linear path \( Z^D : [0, T] \to \mathbb{R}^d \) given by
\[
Z^D_{2kT/2n} := (Z_{tk}, Z_{tk+1}), \quad (25)
\]
and linear interpolation in between. Figure 4 shows the lead-lag transformation of a Brownian motion. As we see, the lead component leads the lag component, hence the name. The lead component can be seen as the future of the path, and the lag component as the past.

Denote by \( Z^D \) the signature of \( Z^D \). Then, we define the lead-lag transformation of \( Z \), denoted by \( Z^{\text{LL}} \), as the limit of signatures of \( Z^D \):
\[
Z^{\text{LL}} := \lim_{|D| \to 0} Z^D.
\]
The work in [9] showed the convergence of this limit and studied some of its properties.

B.1 Expected signature of the lead-lag Brownian motion

Definition B.2. Let \( I = (i_1, \ldots, i_n) \in I^n_3 \) be a multi-index. We denote by \( \mathcal{P}(I) \) the set of all possible tuples of non-empty multi-indices from \( I^{n-1}_3 \) such that their concatenation is equal to \( I \) and their length doesn’t exceed 2, i.e.
\[
\mathcal{P}(I) = \{(i_1, \ldots, i_k) \in (I^{-1}_3)^k : i_1 \otimes \ldots \otimes i_k = I \text{ and } |i_j| \in \{1,2\} \}
\]

Example B.3.

(1) \( \mathcal{P}((1,2,3)) = \{(1,2,3), (1, (2,3)), ((1,2),3)\} \).

(2) \( \mathcal{P}((1,3,2,2)) = \{(1,3,2,2), (1,3,2,2), (1,3,2,2), (1,3,2,2), (1,3,2,2), (1,3,2,2)\} \).

(3) \( \mathcal{P}((3,2)) = \{(3,2), (3,2)\} \).

Definition B.4 (Exponential of a linear functional). Let \( F = \{F^K \} \in \mathbb{R} : K \in I^n_3 \) be a linear functional. We define the exponential of \( F \) as the following linear functional
\[
\exp(F) = (\omega) + \sum_{k=1}^{\infty} \frac{1}{k!} F^\otimes k
\]
where for any \( k \geq 1, F^\otimes k = F \otimes \ldots \otimes F \) for \( k \) times.

Proposition B.5. Define the function \( \alpha : I_3 \to I_3 \) that maps a multi-index to another multi-index in the following way: \( \forall I \in I_3 \nexists (1)
\]
\[
(2) \quad -\frac{1}{2} (1) \text{ if } I = (2, 3) \]
\[
(3) \quad 0 \cdot (\cdot) \text{ otherwise}
\]

Given a final time \( T \) define the linear functional \( E_T := \exp(T + \frac{T}{2}, (2, 2)). \) Then we have the explicit closed-form expression for the Expected Signature of the lead-lag Brownian motion: given any multi-index \( I \in I_3 \)
\[
E \left[ \hat{\mathcal{W}}^I_{0,T} \right] = \sum_{(J_1, \ldots, J_k) \in \mathcal{P}(I)} \langle \alpha(I_1) \otimes \ldots \otimes \alpha(I_k) \rangle E_T
\]

Proof. Follows from [22, Lemma B.1] and the fact that \( E \left[ \hat{\mathcal{W}}^I_{0,T} \right] = E_T. \) \( \square \)

Example B.6. If \( I = (3, 2, 3), \mathcal{P}(I) = \{(3, 2, 3), ((3, 2), 3), (3, 2, 3)\}. \) Hence,
\[
E \left[ \hat{\mathcal{W}}^I_{0,T} \right] = \langle \alpha(3) \otimes \alpha(2) \otimes \alpha(3), E_T \rangle + \langle \alpha((3, 2)) \otimes \alpha(3), E_T \rangle + \langle \alpha(3) \otimes \alpha(2, 3), E_T \rangle
\]
\[
= \frac{T}{2} + \frac{1}{2} \frac{1}{2} + \frac{1}{2} \frac{1}{2} T^2.
\]