Tractable infinite dimensional models: theory and applications

Sara Svaluto-Ferro

UNIVERSITÀ di VERONA

Winter School

Soesterberg, January 2025

Overview of the course

(i) Affine and Polynomial processes

(ii) Measure-valued diffusions

  2.1 HJM drift condition

  2.2 Examples

(iii) Signatures

  3.1 Something as... Taylor

  3.2 Something as... Stone Weierstrass

Part I: Affine and Polynomial processes

(based on Lecture Notes of Martin Larsson)

Introduction

How to work with $dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x_0$?

A model is only useful if statements can be made about its properties.

Feynman–Kac formula: Under suitable conditions and for any $T \geq 0$ it holds

$$\mathbb{E}[f(X_T)] = u(0, x_0)$$

for a large class of functions f, where $u : [0, T] \times \mathbb{R} \to \mathbb{R}$ solves the partial differential equation

$$\begin{aligned} u_t(t,x) + b(x)u_x(t,x) + \frac{1}{2}\sigma(x)^2 u_{xx}(t,x) &= 0, \qquad (t,x) \in [0, T) \times \mathbb{R}, \\ u(T,x) &= f(x), \qquad x \in \mathbb{R}. \end{aligned}$$

A variety of numerical methods exist for solving such equations.

Monte-Carlo simulation: which in its basic form consists of generating a large number of independent replications $X_T^{(1)}, \ldots, X_T^{(n)}$ of X_T, and then using the law of large numbers to obtain

$$\mathbb{E}[f(X_T)] \approx \frac{1}{n}\sum_{i=1}^{n} f(X_T^{(i)}).$$

What about some particularly-tractable processes?

If the dimensionality of X grows large, or if $\mathbb{E}[f(X_T)]$ has to be computed a large number of times (e.g. for different functions f or different coefficients b and σ), such methods eventually become computationally taxing.

However, for some classes of stochastic processes some shortcuts are available. This is in particular the case for

- Affine jump-diffusions, for $f(x) = e^{ux}$ where u is constant, and
- Polynomial jump-diffusions, for $f(x)$ a polynomial in x.

The scope is surprisingly broad. In finance this leads to models for equities, interest rates, credit risk, optimal investment, economic equilibrium, etc.

The goal of the first part of this course is to understand the theory behind such processes.

Jump-diffusions

Jump-diffusions

... i.e. semimartingales whose characteristics are of a particularly nice form.

Definition

Let X be a d-dimensional special semimartingale. We say that X is a *(time-homogeneous) jump-diffusion* if its characteristics (B, C, μ^p) are of the form

$$B_t = \int_0^t b(X_s)\, ds, \quad C_t = \int_0^t a(X_s)\, ds, \quad \mu^p(dt, d\xi) = \nu(X_{t-}, d\xi)dt$$

for some measurable functions $b : \mathbb{R}^d \to \mathbb{R}^d$ and $a : \mathbb{R}^d \to \mathbb{R}^{d\times d}$, and a kernel $\nu(x, d\xi)$ from $\mathbb{R}^d$ into $\mathbb{R}^d$ such that
$a(x)$ is symmetric positive semidefinite,
$\nu(x, \{0\}) = 0$, and $\int_{\mathbb{R}^d} |\xi|^2 \wedge |\xi| \nu(x, d\xi) < \infty$ for all $x \in \mathbb{R}^d$.

We refer to (b, a, ν) as the *coefficients* of X.

Example: let X be a solution of the SDE

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + \gamma(X_t)d(N_t - \lambda t),$$

for a Brownian motion W and a Poisson process N with intensity λ. Then X is a jump-diffusion with coefficients $(b, \sigma^2, \lambda\delta_{\gamma(\cdot)})$.

Generator

What about...the drift of $(f(X_t))_{t\geq 0}$?

Definition

Let X be a jump-diffusion with coefficients (b, a, ν). The extended generator of X is the operator $\mathcal{G}$ defined by

$$\mathcal{G}f(x) = b(x)^\top \nabla f(x) + \frac{1}{2}\operatorname{Tr}(a(x)\nabla^2 f(x)) + \int_{\mathbb{R}^d} \Big(f(x+\xi) - f(x) - \xi^\top \nabla f(x) \Big)\, \nu(x, d\xi)$$

for any C^2 function f such that the integral is well-defined. For $d = 1$ it reads

$$\mathcal{G}f(x) = b(x)f'(x) + \frac{1}{2}a(x)f''(x) + \int_{\mathbb{R}} \left(f(x+\xi) - f(x) - \xi f'(x) \right) \nu(x, d\xi).$$

Intuition: Ito!! Let X be a jump-diffusion with coefficients (b, a, ν) and generator $\mathcal{G}$. Then the process M^f given by

$$M_t^f = f(X_t) - f(X_0) - \int_0^t \mathcal{G}f(X_s)ds$$

is well-defined and a local martingale for any sufficiently integrable C^2 function $f : \mathbb{R}^d \to \mathbb{R}$.

Polynomial jump diffusions

Polynomial operator

Definition

The operator $\mathcal{G}$ is called *polynomial on E* if it is well-defined on $\mathrm{Pol}(E)$ and

$$\mathcal{G}(\mathrm{Pol}_n(E)) \subseteq \mathrm{Pol}_n(E)$$

for each $n \in \mathbb{N}$. In this case, we call X a polynomial jump-diffusion on E.

Proposition

Assume $\mathcal{G}$ is well-defined on $\mathrm{Pol}(E)$. Then the following are equivalent:

(i) $\mathcal{G}$ is polynomial on E;

(ii) The coefficients (b, a, ν) satisfy

$$b \in \mathrm{Pol}_1(E),$$
$$a + \int_{\mathbb{R}^d} \xi\xi^\top \nu(\,\cdot\,, d\xi) \in \mathrm{Pol}_2(E),$$
$$\int_{\mathbb{R}^d} \xi^{\boldsymbol{\alpha}} \nu(\,\cdot\,, d\xi) \in \mathrm{Pol}_{|\boldsymbol{\alpha}|}(E),$$

for all $|\boldsymbol{\alpha}| \geq 3$.

Examples

Consider

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + \gamma(X_t)d(N_t - \lambda t),$$

for a Brownian motion W and a Poisson process N with intensity λ.

Then X is a jump-diffusion with coefficients $(b, \sigma^2, \lambda\delta_{\gamma(\cdot)})$.

$\Longrightarrow$ If $b, \sigma^2 + \lambda\gamma^2, \lambda\gamma^k$ is a polynomial of degree 1,2, and k respectively, then X is a polynomial process.

Examples include: Brownian motion $(b = 0, \sigma = 1, \gamma = 0)$, OU processes $(b(x) = \kappa(\theta - x), \sigma = 1, \gamma = 0)$, geometric Brownian motion $(b = 0, \sigma(x) = x, \gamma = 0)$, CIR diffusion process $(b = 0, \sigma(x) = \sqrt{x}, \gamma = 0)$, a constant process jumping to 0 at an exponential(λ)-time $(b = -\lambda x, \sigma(x) = 0, \gamma = -x)$,...

The moment formula

Our next goal is to establish the *moment formula*, which describes how to calculate conditional expectations of the form

$$\mathbb{E}[p(X_T) \mid \mathcal{F}_t]$$

where X is a polynomial jump-diffusion and p is a polynomial. This is the most important result about polynomial jump-diffusions!

The moment formula: first in 1d

Let X be a polynomial jump-diffusion with generator $\mathcal{G}$, fix $n \in \mathbb{N}$ and let $H(x) := (1, x, x^2, \ldots, x^n)$.
Every $p \in \mathrm{Pol}_n(\mathbb{R})$ has a coordinate representation with respect to such basis, and we denote its coordinate (column) vector by $\vec{p} \in \mathbb{R}^{1+n}$. Thus

$$p(x) = H(x)\vec{p} = 1p_0 + \ldots + x^n p_n \qquad x \in \mathbb{R}.$$

Since $\mathcal{G}$ is polynomial, $\mathcal{G}(x^k)(x) \in \mathrm{Pol}_n(\mathbb{R})$ and hence

$$\mathcal{G}(x^k)(x) = H(x)\vec{q}_k = 1q_{k,0} + \ldots + x^n q_{k,n},$$

for some $\vec{q}_k \in \mathbb{R}^{1+n}$.
By linearity of $\mathcal{G}$ we get

$$\begin{aligned}
\mathcal{G}p(x) &= \mathcal{G}(1)(x)p_0 + \ldots + \mathcal{G}(x^n)(x)p_n \\
&= (H(x)\vec{q}_0)p_0 + \ldots + (H(x)\vec{q}_n)p_n \\
&= H(x)(\vec{q}_0, \ldots, \vec{q}_n)\vec{p} \\
&= H(x)G\vec{p}.
\end{aligned}$$

The moment formula then states that

$$\mathbb{E}[p(X_T) \mid \mathcal{F}_t] = H(X_t)e^{(T-t)G}\,\vec{p}, \quad \text{for } t \leq T.$$

The moment formula

Fix $n \in \mathbb{N}$ and set $N = \dim \mathrm{Pol}_n(E)$ ($= \binom{n+d}{d}$ if $E = \mathbb{R}^d$, but may be smaller in general). Choose $h_1, \dots, h_N \in \mathrm{Pol}_n(\mathbb{R}^d)$ such that

$$h_1|_E, \dots, h_N|_E \text{ form a basis for } \mathrm{Pol}_n(E).$$

Define the (row) vector valued function

$$H : \mathbb{R}^d \to \mathbb{R}^N, \quad H(x) = (h_1(x), \dots, h_N(x)).$$

For each $p \in \mathrm{Pol}_n(E)$ define $\vec{p} \in \mathbb{R}^N$ such that

$$p(x) = H(x)\vec{p}, \quad x \in E.$$

Let $\mathcal{G}$ be polynomial and thus map $\mathrm{Pol}_n(E)$ linearly to itself. Choosing $\vec{q}_k \in \mathbb{R}^N$ such that $\mathcal{G}h_k(x) := H(x)\vec{q}_k$ we get

$$\mathcal{G}p(x) = H(x)(\vec{q}_1, \dots, \vec{q}_N)\vec{p} =: H(x)G\vec{p}, \quad x \in E,$$

Theorem

Let X be an E-valued polynomial process with generator $\mathcal{G}$. Then for any $p \in \mathrm{Pol}_n(E)$ with coordinate vector $\vec{p} \in \mathbb{R}^{1+N}$, the moment formula holds,

$$\mathbb{E}[p(X_T) \mid \mathcal{F}_t] = H(X_t)e^{(T-t)G}\,\vec{p}, \quad \text{for } t \le T.$$

Example

Example

Consider the one-dimensional polynomial diffusion,

$$dX_t = (b + \beta X_t)dt + \sqrt{a + \alpha X_t + A X_t^2}\, dW_t$$

for some real parameters b, β, a, α, A. Its generator is

$$\mathcal{G} f(x) = (b + \beta x) f'(x) + \frac{1}{2}(a + \alpha x + A x^2) f''(x),$$

and the corresponding matrix is given by

$$G = \begin{pmatrix} 0 & b & 2\frac{a}{2} & 0 & \cdots & 0 \\ 0 & \beta & 2\left(b + \frac{\alpha}{2}\right) & 3 \cdot 2\frac{a}{2} & 0 & \vdots \\ 0 & 0 & 2\left(\beta + \frac{A}{2}\right) & 3\left(b + 2\frac{\alpha}{2}\right) & \ddots & 0 \\ 0 & 0 & 0 & 3\left(\beta + 2\frac{A}{2}\right) & \ddots & n(n-1)\frac{a}{2} \\ \vdots & & & 0 & \ddots & n\left(b + (n-1)\frac{\alpha}{2}\right) \\ 0 & & \cdots & & 0 & n\left(\beta + (n-1)\frac{A}{2}\right) \end{pmatrix}.$$

Idea of the proof

Fix (for simplicity) $X_0 = x_0 \in \mathbb{R}$. The proof is based on three results.

(i) $M_t^f = f(X_t) - f(X_0) - \int_0^t \mathcal{G}f(X_s)\,ds$, is a local martingale for $f \in \mathrm{Pol}(E)$.

(ii) Let $f \in \mathrm{Pol}(E)$ and define $\Gamma(f,f) = \mathcal{G}(f^2) - 2f\mathcal{G}f$. Then $\Gamma(f,f)(x) \geq 0$ for all $x \in E$, and

$$(M^f)^2 - \int_0^{\cdot} \Gamma(f,f)(X_s)ds$$

is a local martingale.

(iii) For any $k \in \mathbb{N}$ there exists a constant $C \in \mathbb{R}_+$ such that

$$\mathbb{E}[1 + |X_t|^{2k}] \leq (1 + |x_0|^{2k})\, e^{Ct}, \quad t \geq 0.$$

$\longrightarrow$ For each $f \in \mathrm{Pol}(E)$, the process $(M_t^f)_{t\geq 0}$, is a true martingale.

$\longrightarrow$ Setting $F(T) := \mathbb{E}[H(X_T)|\mathcal{F}_t]$ we get that

$$0 = F(T) - F(t) - \int_0^T F(s)Gds \quad \Rightarrow \quad F(T) = F(t)e^{(T-t)G},$$

and thus multiplying by $\vec{p}$:

$$\mathbb{E}[p(X_T) \mid \mathcal{F}_t] = H(X_T)e^{(T-t)G}\vec{p}.$$

Affine jump-diffusions

Affine Jump-diffusions

We now turn to affine jump-diffusions. Recall that we have fixed a state space $E \subseteq \mathbb{R}^d$ and an E-valued jump-diffusion X with coefficients (b, a, ν) and generator

$$\mathcal{G}f(x) = b(x)^\top \nabla f(x) + \frac{1}{2} \operatorname{Tr}(a(x)\nabla^2 f(x)) + \int_{\mathbb{R}^d} \Big(f(x+\xi) - f(x) - \xi^\top \nabla f(x) \Big) \, \nu(x, d\xi).$$

Definition

The operator $\mathcal{G}$ is called *affine on E* if there exist functions $R_0, \ldots, R_d$ from $\mathrm{i}\mathbb{R}^d$ to $\mathbb{C}$ such that

$$\mathcal{G}e^{u^\top x} = \left(R_0(u) + \sum_{i=1}^{d} R_i(u) x_i \right) e^{u^\top x}$$

holds for all $x \in E$ and $u \in \mathrm{i}\mathbb{R}^d$. In this case, we call X an affine jump-diffusion on E.

Characterization of affine jump-diffusions

Proposition

The following are equivalent:

(i) $\mathcal{G}$ is affine on E;

(ii) The coefficients (b, a, ν) are affine of the form

$$b(x) = b_0 + x_1 b_1 + \cdots + x_d b_d,$$
$$a(x) = a_0 + x_1 a_1 + \cdots + x_d a_d,$$
$$\nu(x, \cdot) = \nu_0 + x_1 \nu_1 + \cdots + x_d \nu_d$$

for all $x \in E$, for some matrices $a_i \in \mathbb{S}^d$, vectors $b_i \in \mathbb{R}^d$, and signed measures ν_i on $\mathbb{R}^d$ such that $\nu_i(\{0\}) = 0$ and $\int_{\mathbb{R}^d} |\xi| \wedge |\xi|^2 \, |\nu_i|(d\xi) < \infty$, $i = 0, \ldots, d$.

In this case, the functions $R_0, \ldots, R_d$ can be taken to be given by

$$R_i(u) = b_i^\top u + \frac{1}{2} u^\top a_i u + \int_{\mathbb{R}^d} \left(e^{u^\top \xi} - 1 - u^\top \xi \right) \nu_i(d\xi).$$

Affine $\Rightarrow$ Polynomial?

Corollary

If X is an affine jump-diffusion on E and $\mathcal{G}$ is well-defined on $\mathrm{Pol}(E)$, then X is a polynomial jump-diffusion on E.

Affine transform formula

Affine jump-diffusions on E not only satisfy the moment formula, subject to the generator being well-defined on $\mathrm{Pol}(E)$. Their characteristic functions are also analytically tractable.

Theorem

Assume X is an affine jump-diffusion on E. Fix $u \in \mathbb{C}^d$ such that $\mathrm{Re}\, u^\top x \leq 0$ and $T > 0$. Let $\phi : [0, T] \to \mathbb{C}$ and $\psi = (\psi_1, \ldots, \psi_d) : [0, T] \to \mathbb{C}^d$ be functions that solve the generalized Riccati equations

$$\begin{aligned} \phi'(\tau) &= R_0(\psi(\tau)), \qquad & \phi(0) &= 0, \\ \psi_i'(\tau) &= R_i(\psi(\tau)), \qquad & \psi_i(0) &= u_i, \quad i = 0, \ldots, d, \end{aligned}$$

for $\tau \in [0, T]$, where

$$R_i(u) = b_i^\top u + \frac{1}{2} u^\top a_i u + \int_{\mathbb{R}^d} \left(e^{u^\top \xi} - 1 - u^\top \xi \right) \nu_i(d\xi).$$

If $\mathrm{Re}\, \phi(\tau) + \mathrm{Re}\, \psi(\tau)^\top x \leq 0$ for all $(\tau, x) \in [0, T] \times E$, then the affine transform formula holds,

$$\mathbb{E}[e^{u^\top X_T} \mid \mathcal{F}_t] = e^{\psi_0(T-t) + \psi(T-t)^\top X_t}, \quad t \leq T.$$

Careful

Compared to the moment formula, the proof of the affine transform formula looks rather short and simple. This is deceptive, because several questions are left unanswered:

(i) nothing is said about existence and uniqueness of solutions of the generalized Riccati equations;

(ii) even if existence and uniqueness is established abstractly, one still has to verify

$$\mathrm{Re}\,\phi(\tau) + \mathrm{Re}\,\psi(\tau)^\top x \leq 0 \quad \text{for all } (\tau, x) \in [0, T] \times E,$$

which can be difficult if the solution is not explicitly given; and

(iii) it is often of interest to obtain the affine transform formula for u with non-zero real part. The martingale property of M then becomes more difficult to verify.

Fourier pricing

Consider an affine jump-diffusion X on $E \subseteq \mathbb{R}^d$ and suppose that the logprice satisfies $\log S_t = Y_t := X_t^1$. To price European puts (and similarly for other options), we need to compute

$$\mathbb{E}_{\mathbb{Q}}[(K - S_T)^+ \mid \mathcal{F}_t] = \mathbb{E}_{\mathbb{Q}}[(K - e^{Y_T})^+ \mid \mathcal{F}_t].$$

Idea: We know how to compute quantities like $\mathbb{E}_{\mathbb{Q}}[e^{u^\top X_T} \mid \mathcal{F}_t]$ in a tractable way. Moreover, for $K > 0$ and $w > 0$ it holds

$$(K - e^y)^+ = \frac{1}{2\pi} \int_{\mathbb{R}} e^{(\mathrm{i}\lambda - w)y} \frac{K^{w+1-\mathrm{i}\lambda}}{(i\lambda - w)(\mathrm{i}\lambda - w - 1)} d\lambda$$

for all $y \in \mathbb{R}$. An application of Fubini solves the problem!

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}}[(K - S_T)^+ \mid \mathcal{F}_t] &= \mathbb{E}_{\mathbb{Q}}[(K - e^{Y_T})^+ \mid \mathcal{F}_t] \\
&= \mathbb{E}_{\mathbb{Q}}\left[\frac{1}{2\pi} \int_{\mathbb{R}} e^{(\mathrm{i}\lambda - w)Y_T} \frac{K^{w+1-\mathrm{i}\lambda}}{(i\lambda - w)(\mathrm{i}\lambda - w - 1)} d\lambda \mid \mathcal{F}_t\right] \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \mathbb{E}_{\mathbb{Q}}[e^{(\mathrm{i}\lambda - w)Y_T} \mid \mathcal{F}_t] \frac{K^{w+1-\mathrm{i}\lambda}}{(i\lambda - w)(\mathrm{i}\lambda - w - 1)} d\lambda.
\end{aligned}$$

Part II: Measure-valued processes for energy markets

(Based on joint work with C. Cuchiero, L. Di Persio and F. Guida)

Energy markets and their financial products

- We consider energy markets, in particular electricity and gas markets, whose essential products are based on futures contracts.

$\longrightarrow$ Futures contract: the price of delivery of electricity/gas over a future time interval $[\tau_1, \tau_2]$ is fixed in the present. Denoting by t the present, the corresponding price is denoted by $F(t, \tau_1, \tau_2)$.

- Idea: $F(t, \tau_1, \tau_2)$ can be written as a weighted integral of instantaneous forward prices $f(t, u)$ with delivery at one fixed time $u \in [\tau_1, \tau_2]$, i.e.

$$F(t, \tau_1, \tau_2) = \int_{\tau_1}^{\tau_2} w(u, \tau_1, \tau_2) f(t, u) du,$$

where $w(u, \tau_1, \tau_2)$ denotes some weight function (see Benth et al. ('08)).

$\longrightarrow$ The "present" is changing day by day, hence "prices" are changing day by day, leading to a function-valued stochastic process $(f(t, \cdot))_{t \in [0,T]}$.

- Why do we want to model this stochasticity? Prices of options written on $F(t, \tau_1, \tau_2)$ for different maturities t and different delivery periods $[\tau_1, \tau_2]$ are available! This means that quantities as

$$\mathbb{E}_{\mathbb{Q}}[(F(t, \tau_1, \tau_2) - K)^+]$$

can be red from the market, where $\mathbb{Q}$ denotes a so called equivalent martingale measure.

Solving couple of problems

Given: $(f(t,\cdot))_{t\in[0,T]}$ where $f(t,u)$ is the price at time t for the instantaneous delivery at time u and T is a fixed time horizon.

- First problem: The support of $f(t,\cdot)$ is $[t,T] \longrightarrow$ it changes over time!
  Solution: Musiela parametrization: $(f(t,t+\cdot))_{t\in[0,T]}$.
- Second problem: Are we assuming to much regularity? Prices can jump at predictable times (see Fontana et al. (2020)), e.g. due to maintenance works or predictable dates of political decisions.

  - Evolution of gas spot prices in the last 5 years (Eur/MWh).
  - The last spike (end of August 2022) corresponds to the announcement of an indefinite shutdown of Nordstream 1 by Gasprom.

  Solution: Rather than using $f(t,t+x)dx$, we can also use a measure μ_t on $[0,T]$. Future prices at time $t\in[0,\tau_1]$ then become

$$F(t,\tau_1,\tau_2) = \int_{(\tau_1-t,\tau_2-t]} w(t+x,\tau_1,\tau_2)d\mu_t(dx).$$

Part II.1: HJM drift condition

Other financial conditions: Absence of arbitrage

- Arbitrage: strategy permitting to (possibly) obtain something in exchange of nothing. A good model for pricing should not allow such strategies!

- By Cuchiero, Klein, Teichmann ('16) one can exclude arbitrage opportunities by requiring that there exists an equivalent measure $\mathbb{Q}$ such that all traded products

$$\{F(t,\tau_1,\tau_2)_{t\in[0,\tau_1]} \mid 0 \leq \tau_1 < \tau_2 \leq T\} \text{ are local } \mathbb{Q}\text{-martingales.} \tag{1.1}$$

- Question: How does this translate to the underlying measure-valued process $(\mu_t)_{t\geq 0}$?

Including no arbitrage conditions

- State space $M_+(E)$: nonnegative measures on $E = [0, T]$, equipped with the weak-topolgy.
- Notation: $\langle \phi, \mu \rangle = \int_E \phi(x)\mu(dx)$
- Following an Heath-Jarrow-Morton (HJM) approach such martingality can be guaranteed imposing conditions on the "drift" of $(\mu_t)_{t\in[0,T]}$.

Theorem

Let $(\mu_t)_{t\geq 0}$ be an $M_+(E)$-valued process such that the future prices are given by

$$F(t, \tau_1, \tau_2) = \int_{(\tau_1 - t, \tau_2 - t]} w(t + x; \tau_1, \tau_2)\mu_t(dx), \quad t \in [0, \tau_1]$$

for $w(\cdot; \tau_1, \tau_2) \in C^\infty(\mathbb{R})$. Then the market is *free of arbitrage* if there exists an equivalent measure $\mathbb{Q}$ such that

$$\mathbb{E}_{\mathbb{Q}}[\sup_{t\in[0,T]} \mu_t(E)] < \infty$$

and for all $\phi \in D = \{x \mapsto \phi|_E(x) : \phi \in C^\infty(\mathbb{R}) \text{ s.t. } \phi'(0) = 0\}$

$$\left(\langle \phi, \mu_t \rangle + \int_0^t \langle \phi', \mu_s \rangle ds\right)_{t\in[0,T]} \qquad \text{(HJM-cond)}$$

is a $\mathbb{Q}$-martingale.

Note that (HJM-cond) is a weak formulation of "$d\mu_t(dx) = \frac{d}{dx}\mu_t(dx)dt + dN_t(dx)$", where N denotes a measure-valued (local) martingale.

Mathematics: first in 1D

Consider a set of functions $D \subseteq C^2(\mathbb{R})$. A diffusion type operator $L : D \to C(\mathbb{R})$ is a linear operator admitting the representation

$$Lf(x) = b(x)f'(x) + \frac{1}{2}a(x)f''(x).$$

A real valued process $(X_t)_{t\in[0,T]}$ is called solution to the martingale problem for L if

$$\left(f(X_t) - f(X_0) - \int_0^t Lf(X_s)ds\right)_{t\in[0,T]}$$

is a local martingale for each $f \in D$.

Markovian setting for measure-valued diffusions

- Consider cylindrical functions $M(E) \to \mathbb{R}$ lying in

$$F^D = \Big\{\nu \mapsto f(\nu) = \Phi(\langle g_1, \nu\rangle, \dots, \langle g_m, \nu\rangle) : \Phi \in C^\infty(\mathbb{R}^m), g_k \in D,\ m \in \mathbb{N}_0\Big\},$$

  with $D = \{x \mapsto \phi|_E(x) : \phi \in C^\infty(\mathbb{R}) \text{ s.t. } \phi'(0) = 0\}$ (being a dense linear subspace of $C(E)$) and define $F_c^D := F^D(M_+(E)) \cap C_c(M_+(E))$.

- Directional derivatives: a function $f : M(E) \to \mathbb{R}$ is called differentiable at ν in direction δ_x for $x \in E$ if

$$\partial_x f(\nu) := \lim_{\varepsilon \to 0} \frac{f(\nu + \varepsilon\delta_x) - f(\nu)}{\varepsilon}$$

  exists. We write $\partial f(\nu)$ for the map $x \mapsto \partial_x f(\nu)$ and $\partial^k f(\nu)$ for the higher order derivatives from E^k to $\mathbb{R}$.

- For $f \in F^D$, we have for instance

$$\partial f(\nu) = \sum_{i=1}^m \partial_i \Phi(\langle g_1, \nu\rangle, \dots, \langle g_m, \nu\rangle) g_i,$$

  which is thus a function in D.

Diffusion-type operators and martingale problems

Definition

- A linear operator $L : F_c^D \to C_0(M_+(E))$ is called diffusion-type operator if it admits a representation

$$Lf(\nu) = B(\partial f(\nu), \nu) + \frac{1}{2} Q(\partial^2 f(\nu), \nu)$$

  for some operators $B : D \times M_+(E) \to \mathbb{R}$ and $Q : D \otimes D \times M_+(E) \to \mathbb{R}$ such that $B(\cdot, \nu)$ and $Q(\cdot, \nu)$ are linear for all $\nu \in M_+(E)$.

- An $M_+(E)$-valued process $(\mu_t)_{t \in [0,T]}$ with continuous trajectories is called solution to the martingale problem for L if

$$N_t^f = f(\mu_t) - f(\mu_0) - \int_0^t Lf(\mu_s) ds, \quad t \in [0, T]$$

  defines a local martingale for every f in F_c^D.

Existence result

Note that $M_+(E)$ is a locally compact and separable space, hence Martingale problem existence results in form of the positive maximum principle can be applied.

Theorem

Let L be diffusion-type operator such that $\nu \mapsto Q(\partial^2 f(\nu), \nu) \in C_0(M_+(E))$ for all $f \in F_c^D$. Suppose that the drift part B is given by

$$B(\partial f(\nu), \nu) = -\langle \frac{d}{dx} \partial f(\nu), \nu \rangle.$$

If the diffusion part Q satisfies the positive maximum principle, i.e. it holds that

$$f \in F_c^D, \, \nu^* \in M_+(E), \sup_{M_+(E)} f = f(\nu^*) \geq 0 \quad \text{implies} \quad Q(\partial^2 f(\nu^*), \nu^*) \leq 0,$$

then there exists an $M_+(E)$-valued solution to the martingale problem for L which satisfies the HJM condition.

Remark: The operator corresponding to the drift part $-\langle \frac{d}{dx} \partial f(\nu), \nu \rangle$ satisfies the positive maximum principle due to the choice of D.

Part II.2: Examples

Towards tractable examples

How can these models be used? Are there particular convenient choices of the parameters (B and Q) in this sense?

- As B is fully specified due to the HJM condition the only freedom consists in choosing the covariance structure Q.
- An analogue to neural SPDEs is to parametrize the map $M_+(E) \to \mathbb{R} : \nu \mapsto Q(g, \nu)$ for $g \in D \otimes D$ via neural networks taking measures as inputs

  $\longrightarrow$ See e.g. Benth et al. ('21); Acciaio et al. ('22), C., Schmocker and Teichmann ('22) for neural networks with infinite dimensional inputs.
- Alternative: let $(\mu_t)_{t\in[0,T]}$ lie in the class of polynomial and affine measure-valued diffusions.

Polynomial diffusions

Definition

A linear operator L is called $M_+(E)$-polynomial if it maps cylindrical polynomials P^D (i.e. $p(\langle \phi_1, \nu\rangle, \dots, \langle \phi_n, \nu\rangle)$ for a polynomial p) to polynomials (i.e. a slightly larger class than cylindrical polynomials) of same or lower degree.

Why is it nice? Because each quantity of the form

$$\mathbb{E}_{\mathbb{Q}}[p(\langle \phi_1, \mu_t\rangle, \dots, \langle \phi_n, \mu_t\rangle)]$$

for a polynomial p of degree k can be computed solving a system of $k+1$ linear-PDEs.

Polynomial specification

Theorem

Let L be a diffusion operator given by $Lf(\nu) = -\langle \frac{d}{dx}\partial f(\nu), \nu\rangle + \frac{1}{2}Q(\partial^2 f(\nu), \nu)$.

- Then L is $M_+(E)$-polynomial if and only if $\nu \mapsto Q(g, \nu)$ is quadratic for every $g \in D \otimes D$, i.e.

$$Q(g, \nu) = Q_0(g) + \langle Q_1(g), \nu\rangle + \langle Q_2(g), \nu^2\rangle$$

  for linear operators Q_0, Q_1, Q_2.

- Moreover, if $Q_0 = 0$ and if
  - $Q_1(g)(x) = \alpha(x)g(x, x)$ with $\alpha \in C_+(E)$ and
  - $Q_2(g)(x, y) = \frac{1}{2}(\pi(x, y)g(x, x) + \pi(y, x)g(y, y) + 2\beta(x, y)g(x, y))$,

    for some functions π and β satisfying certain admissibility conditions,

  then there is an $M_+(E)$-valued solution to the martingale problem for L.

Observation: The map $M_+(E) \to \mathbb{R} : \nu \mapsto Q(g, \nu)$ reduces to a quadratic function whose coefficients are parametrized by functions α, β, π which in turn can be parametrized by usual neural networks.

Example: choosing $Q_1 = 0$ we get a Black-Scholes-type measure valued model.

Affine measure-valued HJM-models: an example

Let $(\mu_t)_{t\in[0,T]}$ be a solution of the martingale problem for

$$Lf(\nu) = -\langle \frac{d}{dx}\partial f(\nu), \nu\rangle + \frac{1}{2}\langle \alpha(\text{diag}\partial^2 f(\nu)), \nu\rangle,$$

$\alpha \in C_+(E)$.

- This corresponds to a variant of the Dawson-Watanabe superprocess. Its diffusion part is analogous to the one-dimensional Feller diffusion.
- This is an example of an $M_+(E)$ measure valued affine process.
- Why is it nice? Its Fourier-Laplace transform for some function $u \in C(E, \mathbb{R}_- + \mathrm{i}\mathbb{R})$ is given by

$$\mathbb{E}[\exp(\langle u, \mu_t\rangle)] = \exp(\langle \psi(t,u), \mu_0\rangle),$$

where ψ solves the following nonlinear Riccati PDE

$$\partial_t \psi(t,u)(x) = -\frac{d}{dx}\psi(t,u)(x) + \frac{1}{2}\alpha(x)\psi(t,u)^2(x), \quad \psi(0,u) = u.$$

- Somehow surprisingly, this can be solved explicitly

$$\psi(t,u)(x) = \frac{u\big((x-t)^+\big)}{1 - u\big((x-t)^+\big)\int_0^t \frac{1}{2}\alpha\big((x-s)^+\big)ds}.$$

Computational aspects: how to choose α?

Recall that from the market we can read many quantities of the form

$$\mathbb{E}_{\mathbb{Q}}[(F(t,\tau_1,\tau_2)-K)^+] \qquad \text{for} \qquad F(t,\tau_1,\tau_2)=\langle w(t+\cdot,\tau_1,\tau_2)1_{(\tau_1,\tau_2]}(t+\cdot),\mu_t\rangle.$$

One then applies Fourier pricing techniques.

- Recall that $(x-K)^+=\int_{\mathbb{R}}\exp((C+\mathrm{i}\lambda)x)\widehat{f}_K(\lambda)d\lambda$.
- Apply Fubini and reduce the problem to compute

$$\mathbb{E}_{\mathbb{Q}}[\exp((C+\mathrm{i}\lambda)\langle w(t+\cdot,\tau_1,\tau_2)1_{(\tau_1,\tau_2]}(t+\cdot),\mu_t\rangle)]$$

  for each λ.
- Recall that we have an explicit representation of such quantities:

$$\exp(\langle\psi^\alpha(t,u^\lambda),\mu_0\rangle \qquad \text{for} \qquad u^\lambda(x)=(C+\mathrm{i}\lambda)w(t+x,\tau_1,\tau_2)1_{(\tau_1,\tau_2]}(t+x).$$

- Insert back in the integral and obtain the desired quantity.

Observation: The computed expectation depends on the map $\alpha: E\to\mathbb{R}_+$. Idea: one can now parametrize α with a neural network ad find the parameters that fit the most the prices on the market!

Calibration to market data

- We calibrate the model to call option prices for one specific maturity and one delivery period.
- We use EEX German Power data extracted from

  `https://www.eex.com/en/market-data/power/options`

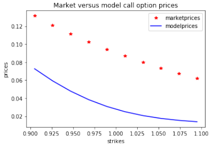
  at March 22, 2022 and calibrate to call options with maturity April 26, 2022 and delivery period one month, starting on May 1, 2022.
- To provide just a calibration example, we use a simple L^2-criterion, i.e. we minimize

$$\sum_K |\pi_{\mathsf{mkt}}(K) - \pi_0(K)|^2,$$

  where $\pi_{\mathsf{mkt}}(K)$ denotes the market call option price with strike K and $\pi_0(K)$ the model call option price obtained via Fourier pricing.

Calibration to market data

- When starting from appropriately intialized parameters for the neural networks such that the the curves look before training for instance like



the target curve can then be reached in around 150-300 gradient step iterations and thus yields the following fast and accurate calibration results

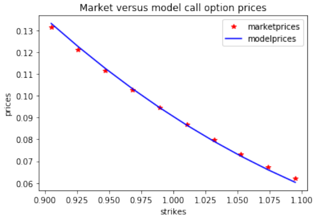


Figura: Market versus model call option prices for options expiring on April 26, 2022, written on forward contracts with delivery during May.

Conclusion

- Measure valued processes for energy markets to model future prices.
- Mathematically convenient and tractable, in particular in the affine and polynomial class.
- The "parameters" to calibrate are typically functions of the spatial variable.
- The universal approximation theorem suggests to parametrize them as neural networks.

$\Rightarrow$ Neural measure-valued processes

$\Rightarrow$ Potentially highly parametric infinite dimensional stochastic models that can be calibrated to market data.

- Outlook: Extend to time-inhomogenous measure-valued polynomial diffusions to account for seasonality and calibrate slice-wise for each maturity.

Part III: Signatures

Signatures...why?

Because the (time extended) signature of a continuous semimartingale uniquely determines its path...

...and because every polynomial on the signature has a linear representative.

$\longrightarrow$ If $S_T = F((X_t)_{t\in[0,T]})$ for some continuous map F, then

$$S_T \approx L(\widehat{\mathbb{X}}_T)$$

for some linear map L, where $\widehat{\mathbb{X}}$ denotes the signature of $t \mapsto (t, X_t)$.

$\longrightarrow$ Linear regressions, affine and polynomial technology, and other useful machinery can be applied!

Signature: definition and properties

Signature of a 1 dimensional path of finite variation

The signature $(\mathbb{X}_t)_{t\in[0,T]}$ of a 1-dimensional path $(X_t)_{t\in[0,T]}$ of finite variation is defined as

$$\mathbb{X}_t = \Big(1, \int_0^t 1 dX_{t_1}, \int_0^t \int_0^{t_1} 1 dX_{t_2} dX_{t_1}, \int_0^t \int_0^{t_1} \int_0^{t_2} 1 dX_{t_3} dX_{t_2} dX_{t_1}, \ldots\Big),$$

where the integrals are all Rieman-Stiltjes integrals.

- State space: extended tensor algebra $T((\mathbb{R})) = \{(a_0, a_1, a_2, \ldots) : a_i \in \mathbb{R}\}$.
- Notation: we use $\langle e_\emptyset, \mathbb{X}_t\rangle := 1$ and denote the element of $\mathbb{X}_t$ corresponding to k iterated integrals with respect to X as

$$\langle \underbrace{e_1 \otimes \cdots \otimes e_1}_{k \text{ times}}, \mathbb{X}_t\rangle \qquad \text{or} \qquad \langle e_1^{\otimes k}, \mathbb{X}_t\rangle \qquad \text{or} \qquad \langle e_I, \mathbb{X}_t\rangle \text{ for } I := (\underbrace{1, \ldots, 1}_{k \text{ times}})$$

- Observation: the k-th term of the signature, is given by the $(k-1)$-th term of the signature integrated from 0 to t:

$$\int_0^t \langle e_1^{\otimes(k-1)}, \mathbb{X}_s\rangle dX_s = \langle e_1^{\otimes k}, \mathbb{X}_t\rangle$$

- Attention: the signature of $(X_t)_{t\in[0,T]}$ and $(X_t + c)_{t\in[0,T]}$ for $c \in \mathbb{R}$ coincide!

Signature of a 1 dimensional path of finite variation

The signature $(\mathbb{X}_t)_{t\in[0,T]}$ of a 1-dimensional path $(X_t)_{t\in[0,T]}$ of finite variation is the path taking values in $T((\mathbb{R}))$ given by

$$\mathbb{X}_t = (\langle e_\emptyset, \mathbb{X}_t\rangle, \langle e_1, \mathbb{X}_t\rangle, \langle e_1^{\otimes 2}, \mathbb{X}_t\rangle, \langle e_1^{\otimes 3}, \mathbb{X}_t\rangle, \ldots),$$

for $\langle e_\emptyset, \mathbb{X}_t\rangle = 1$ and $\langle e_1^{\otimes k}, \mathbb{X}_t\rangle = \int_0^t \langle e_1^{\otimes (k-1)}, \mathbb{X}_s\rangle dX_s$.

Signature of a d dimensional path of finite variation

The signature $(\mathbb{X}_t)_{t\in[0,T]}$ of an $\mathbb{R}^d$-valued path $(X_t^1, \dots, X_t^d)_{t\in[0,T]}$ of finite variation is defined as

$$\mathbb{X}_t = (1, \int_0^t 1dX_{t_1}^1, \dots, \int_0^t 1dX_{t_1}^d, \int_0^t \int_0^{t_1} 1dX_{t_2}^1 dX_{t_1}^1, \int_0^t \int_0^{t_1} 1dX_{t_2}^1 dX_{t_1}^2 \dots),$$

where the integrals are all Rieman-Stiltjes integrals.

- State space: extended tensor algebra

$$T((\mathbb{R}^d)) = \{(a_0, a_1, \dots) : a_i \in \underbrace{(\mathbb{R}^d)^{\otimes i}}_{\cong\mathbb{R}^{d^i}\text{,i.e. 1 dim }\forall\text{ iterated integral of deep }i} \}.$$

- Notation: we denote the element of $\mathbb{X}_t$ corresponding to the i_1-th element of X, integrated wrt the i_2-th component of X,..., integrated wrt the i_n-component of X as

$$\langle e_{i_1} \otimes \cdots \otimes e_{i_n}, \mathbb{X}_t \rangle \qquad \text{or} \qquad \langle e_I, \mathbb{X}_t \rangle \text{ for } I = (i_1, \dots, i_n).$$

⇒ Example: $X_t = (Y_t, Z_t)$, $\langle e_1 \otimes e_2, \mathbb{X}_t \rangle = \int_0^t \int_0^s 1dY_r dZ_s = \int_0^t (Y_s - Y_0) dZ_s$.

- Observation: signature terms can be defined recursively: for $I = (i_1, \dots, i_n)$ we have

$$\langle e_I, \mathbb{X}_t \rangle = \int_0^t \langle e_{i_1} \otimes \cdots \otimes e_{i_{n-1}}, \mathbb{X}_s \rangle dX_s^{i_n} = \langle e_{i_1} \otimes \cdots \otimes e_{i_n}, \mathbb{X}_s \rangle.$$

Signature of a d dimensional path of finite variation

The signature $(\mathbb{X}_t)_{t\in[0,T]}$ of a d-dimensional path $(X_t)_{t\in[0,T]}$ of finite variation is the path taking values in $T((\mathbb{R}^d))$ given by

$$\mathbb{X}_t = (\langle e_\emptyset, \mathbb{X}_t\rangle, \langle e_1, \mathbb{X}_t\rangle, \ldots, \langle e_d, \mathbb{X}_t\rangle, \langle e_1^{\otimes 2}, \mathbb{X}_t\rangle, \langle e_1 \otimes e_2, \mathbb{X}_t\rangle, \ldots),$$

for $\langle e_\emptyset, \mathbb{X}_t\rangle = 1$ and

$$\langle e_I, \mathbb{X}_t\rangle = \int_0^t \langle e_{i_1} \otimes \cdots \otimes e_{i_{n-1}}, \mathbb{X}_s\rangle dX_s^{i_n},$$

for $I = (i_1, \ldots i_n)$.

Signature of a d dimensional continuous semimartingale

The signature $(\mathbb{X}_t)_{t\in[0,T]}$ of a d-dimensional continuous semimartingale $(X_t)_{t\in[0,T]}$ is the process taking values in $T((\mathbb{R}^d))$ given by

$$\mathbb{X}_t = (\langle e_\emptyset, \mathbb{X}_t\rangle, \langle e_1, \mathbb{X}_t\rangle, \ldots, \langle e_d, \mathbb{X}_t\rangle, \langle e_1^{\otimes 2}, \mathbb{X}_t\rangle, \langle e_1 \otimes e_2, \mathbb{X}_t\rangle, \ldots),$$

for $\langle e_\emptyset, \mathbb{X}_t\rangle = 1$ and

$$\langle e_I, \mathbb{X}_t\rangle = \int_0^t \langle e_{i_1} \otimes \cdots \otimes e_{i_{n-1}}, \mathbb{X}_s\rangle \circ dX_s^{i_n},$$

where $I = (i_1, \ldots i_n)$ and $\circ$ (for now) denotes the Stratonivoch integral:

$$\int_0^t Y_t \circ dZ_t = \int_0^t Y_t dZ_t + \frac{1}{2}[Y, Z]_t.$$

The shuffle property or the integration by parts formula

Stratonovich (and Riemann-Stiltjes) integrals satisfy the integration by parts formula:

$$\int_0^t Y_s \circ dZ_s = Y_t Z_t - Z_0 Y_0 - \int_0^t Z_s \circ dY_s.$$

Setting $Y_t = \langle e_I, \mathbb{X}_t \rangle$ and $Z_t = \langle e_J, \mathbb{X}_t \rangle$ this yields

$$\begin{aligned}
\langle e_I, \mathbb{X}_t \rangle \langle e_J, \mathbb{X}_t \rangle &= \int_0^t \langle e_I, \mathbb{X}_t \rangle \circ d\langle e_J, \mathbb{X}_t \rangle + \int_0^t \langle e_J, \mathbb{X}_t \rangle \circ d\langle e_I, \mathbb{X}_t \rangle \\
&= \int_0^t \langle e_I, \mathbb{X}_t \rangle \langle e_{J'}, \mathbb{X}_t \rangle \circ dX_t^{j_m} + \int_0^t \langle e_J, \mathbb{X}_t \rangle \langle e_{I'}, \mathbb{X}_t \rangle \circ dX_t^{i_n},
\end{aligned}$$

for $e_I = e_{I'} \otimes e_{i_n}$ and $e_J = e_{J'} \otimes e_{j_m}$.

Defining $e_I \sqcup\!\sqcup\, e_\emptyset = e_\emptyset \sqcup\!\sqcup\, e_I = e_I$ and then recursively

$$e_I \sqcup\!\sqcup\, e_J := (e_I \sqcup\!\sqcup\, e_{J'}) \otimes e_{j_m} + (e_J \sqcup\!\sqcup\, e_{I'}) \otimes e_{i_n}$$

we get

$$\langle e_I, \mathbb{X}_t \rangle \langle e_J, \mathbb{X}_t \rangle = \underbrace{\langle e_I \sqcup\!\sqcup\, e_J, \mathbb{X}_t \rangle}_{\text{linear combination of } \mathbb{X}_t\text{'s elements!}}.$$

Every polynomial in the signature has a linear representation!

Examples examples...

Set for simplicity that $X_0 = 0$.

$$\begin{aligned}
\langle e_1, \mathbb{X}_t\rangle^2 &= (X_t)^2 \overset{\text{Itô}}{=} 2\int_0^t X_s dX_s + [X]_t = 2\int_0^t X_s \circ dX_s = 2\langle e_1 \otimes e_1, \mathbb{X}_t\rangle \\
&\Rightarrow \quad e_1 \sqcup\!\sqcup\, e_1 = 2 e_1 \otimes e_1 \\
\langle e_1, \mathbb{X}_t\rangle\langle e_2, \mathbb{X}_t\rangle &= X_t^1 X_t^2 \overset{\text{Itô}}{=} \int_0^t X_s^1 dX_s^2 + \int_0^t X_s^2 dX_s^1 + [X^1, X^2]_t \\
&= \int_0^t X_s^1 \circ dX_s^2 + \int_0^t X_s^2 \circ dX_s^1 = \langle e_1 \otimes e_2, \mathbb{X}_t\rangle + \langle e_2 \otimes e_1, \mathbb{X}_t\rangle \\
&\Rightarrow \quad e_1 \sqcup\!\sqcup\, e_2 = e_1 \otimes e_2 + e_2 \otimes e_1 \\
\langle e_1, \mathbb{X}_t\rangle^k &= k!\langle e_1^{\otimes k}, \mathbb{X}_t\rangle \\
&\Rightarrow \quad e_1 \sqcup\!\sqcup \cdots \sqcup\!\sqcup\, e_1 = k! e_1^{\otimes k} \\
&\Rightarrow \quad \langle e_1^{\otimes k}, \mathbb{X}_t\rangle = \frac{(X_t)^k}{k!} \\
&\Rightarrow \quad \text{If } X \text{ is 1-dimensional: } \mathbb{X}_t = \left(1, X_t, \frac{(X_t)^2}{2!}, \frac{(X_t)^3}{3!}, \ldots\right)
\end{aligned}$$

Examples, examples,...

Example

Set $X_t = t$. Then

$$\mathbb{X}_t = (1, t, \frac{t^2}{2}, \frac{t^3}{6}, \ldots, \frac{t^k}{k!}, \ldots).$$

Example

Let X be a one dimensional continuous semimartingale with $X_0 = 0$. Then

$$\mathbb{X}_t = (1, X_t, \frac{X_t^2}{2}, \frac{X_t^3}{6}, \ldots, \frac{X_t^k}{k!}, \ldots).$$

Example

Consider $\widehat{X}_t = (t, X_t)$, where X is a one dimensional continuous semimartingale with $X_0 = 0$. Then

$$\widehat{\mathbb{X}}_t = (1, t, X_t, \frac{t^2}{2}, \int_0^t s dX_s, \int_0^t X_s ds, \frac{X_t^2}{2}, \frac{t^3}{6}, \ldots).$$

The half-shuffle property

Setting

$$e_I \widetilde{\sqcup\!\sqcup} e_J := (e_I \sqcup\!\sqcup e_{J'}) \otimes e_{j_m}$$

where $e_J = e_{J'} \otimes e_{j_m}$ we get

$$\int_0^t \langle e_I, \mathbb{X}_s \rangle \circ d\langle e_J, \mathbb{X}_s \rangle = \langle e_I \widetilde{\sqcup\!\sqcup} e_J, \mathbb{X}_t \rangle.$$

The signature of $\langle e_I, \mathbb{X} \rangle$ can be written as linear combination of the signature of $\mathbb{X}$.

Uniqueness of the time extended signature

...namely: the value $\widehat{\mathbb{X}}_T$ of the signature $\widehat{\mathbb{X}}$ of $\widehat{X}_t := (t, X_t)$ at time T uniquely determines the trajectories of $(X_t - X_0)_{t \in [0,T]}$.

Why? For each k and i

$$\int_0^T (X_s^i - X_0^i) \frac{s^k}{k!} ds$$

can be written as (finite) linear combination of $\widehat{\mathbb{X}}_T$'s components!

Welcome back Markovianity :).

The Chen relation

Set

$$\mathbb{X}_{s,t} = (\langle e_\emptyset, \mathbb{X}_{s,t}\rangle, \langle e_1, \mathbb{X}_{s,t}\rangle, \ldots, \langle e_d, \mathbb{X}_{s,t}\rangle, \langle e_1^{\otimes 2}, \mathbb{X}_{s,t}\rangle, \langle e_1 \otimes e_2, \mathbb{X}_{s,t}\rangle, \ldots),$$

for $\langle e_\emptyset, \mathbb{X}_{s,t}\rangle = 1$ and

$$\langle e_I, \mathbb{X}_{s,t}\rangle = \int_s^t \langle e_{i_1} \otimes \cdots \otimes e_{i_{n-1}}, \mathbb{X}_{s,r}\rangle \circ dX_r^{i_n},$$

where $I = (i_1, \ldots i_n)$ and $\circ$ denotes the Stratonivoch integral.

Lemma (Chen relation)

$$\langle e_I, \mathbb{X}_{0,t}\rangle = \sum_{e_{I_1} \otimes e_{I_2} = e_I} \langle e_{I_1}, \mathbb{X}_{0,s}\rangle \langle e_{I_2}, \mathbb{X}_{s,t}\rangle.$$

Examples examples

Chen relation: $\langle e_I, \mathbb{X}_{0,t}\rangle = \sum_{e_{I_1}\otimes e_{I_2}=e_I}\langle e_{I_1}, \mathbb{X}_{0,s}\rangle\langle e_{I_2}, \mathbb{X}_{s,t}\rangle$.

For $I=(1)$ it reads

$$\underbrace{\langle e_{(1)}, \mathbb{X}_{0,t}\rangle}_{=X_t - X_0} = \underbrace{\langle e_{(1)}, \mathbb{X}_{0,s}\rangle\langle e_{\emptyset}, \mathbb{X}_{s,t}\rangle}_{=(X_s - X_0)} + \underbrace{\langle e_{\emptyset}, \mathbb{X}_{0,s}\rangle\langle e_{(1)}, \mathbb{X}_{s,t}\rangle}_{=(X_t - X_s)}.$$

For $I=(1,1)$ it reads

$$\underbrace{\langle e_{(1,1)}, \mathbb{X}_{0,t}\rangle}_{=\frac{(X_t - X_0)^2}{2}} = \underbrace{\langle e_{(1,1)}, \mathbb{X}_{0,s}\rangle\langle e_{\emptyset}, \mathbb{X}_{s,t}\rangle}_{=\frac{(X_s - X_0)^2}{2}} + \underbrace{\langle e_{(1)}, \mathbb{X}_{0,s}\rangle\langle e_{(1)}, \mathbb{X}_{s,t}\rangle}_{=(X_s - X_0)(X_t - X_s)} + \underbrace{\langle e_{\emptyset}, \mathbb{X}_{0,s}\rangle\langle e_{(1,1)}, \mathbb{X}_{s,t}\rangle}_{=\frac{(X_t - X_s)^2}{2}}.$$

It can also be used for:

$$\mathbb{E}[\langle e_I, \mathbb{X}_{0,t}\rangle|\mathcal{F}_s] = \sum_{e_{I_1}\otimes e_{I_2}=e_I}\langle e_{I_1}, \mathbb{X}_{0,s}\rangle\mathbb{E}[\langle e_{I_2}, \mathbb{X}_{s,t}\rangle|\mathcal{F}_s].$$

If X has independent increments this reduces to

$$\mathbb{E}[\langle e_I, \mathbb{X}_{0,t}\rangle|\mathcal{F}_s] = \sum_{e_{I_1}\otimes e_{I_2}=e_I}\langle e_{I_1}, \mathbb{X}_{0,s}\rangle\mathbb{E}[\langle e_{I_2}, \mathbb{X}_{s,t}\rangle].$$

Stone Weierstrass or the universal approximation theorem

Fix a continuous semimartingale X with $X_0 = 0$.

Let $(\widehat{\mathbb{X}}_t^2)_{t\in[0,T]}$ denote the signature of (t, X_t) *truncated* at level 2:

$$\widehat{\mathbb{X}}_t^2 = (1, t, X_t^1, \dots, X_t^d, \int_0^t s \circ ds, \dots, \int_0^t X_s^d \circ dX_s^d)$$

Then every quantity of the form

$$f\Big((\widehat{\mathbb{X}}_t^2)_{t\in[0,T]}\Big)$$

for some continuous map f can be almost surely approximated arbitrarily well on compact sets by objects of the form $\sum_{I\in\mathcal{I}} \lambda_I \langle e_I, \widehat{\mathbb{X}}_T \rangle$, where $\lambda_I \in \mathbb{R}$ and $\mathcal{I}$ contains a finite number of indices I.

Polynomiality: a and b have to be...linear?

Let X be the $\mathbb{R}$-valued process given by

$$dX_t = b(X_t)dt + \sqrt{a(X_t)}dW_t$$

for some polynomials a and b. Choose for simplicity $X_0 = 0$.

Lift X to its signature $\mathbb{X}_t = (1, X_t, \frac{X_t^2}{2}, \dots)$.

Observe that:

- a and b are linear maps in $\mathbb{X}$. Even more!
- $$d\frac{X_t^k}{k!} = \underbrace{\frac{X_t^{k-1}}{(k-1)!}b(X_t) + \frac{1}{2}\frac{X^{k-2}}{(k-2)!}a(X_t)}_{\text{linear in } \mathbb{X}_t!}dt + \sqrt{\underbrace{\left(\frac{X_t^{k-1}}{(k-1)!}\right)^2 a(X_t)}_{\text{linear in } \mathbb{X}_t!}}\,dW_t$$
  $$= L_{k,b}(\mathbb{X}_t)dt + \sqrt{L_{k,a}(\mathbb{X}_t)}dW_t,$$

  for some linear maps $L_{k,b}$ and $L_{k,a}$.

$\mathbb{X}$ is a candidate affine AND a polynomial process!

Even better: Polynomiality with polynomial diffusions

Let X be the $\mathbb{R}$-valued polynomial process given by

$$dX_t = b(X_t)dt + \sqrt{a(X_t)}dW_t$$

for some polynomials a and b of degree 2 and 1, respectively. Choose for simplicity $X_0 = 0$.

Lift X to its truncated signature $\mathbb{X}_t^{\leq n} = (1, X_t, \frac{X_t^2}{2}, \dots, \frac{X^n}{n!})$.

Observe that for each $k \leq n$

$$d\frac{X_t^k}{k!} = \underbrace{\frac{X_t^{k-1}}{(k-1)!}b(X_t) + \frac{1}{2}\frac{X^{k-2}}{(k-2)!}a(X_t)}_{\text{linear in } \mathbb{X}_t^{\leq n}!}dt + \sqrt{\underbrace{\left(\frac{X_t^{k-1}}{(k-1)!}\right)^2 a(X_t)}_{\text{quadratic in } \mathbb{X}_t^{\leq n}!}}\,dW_t$$

$$= L_{k,b}(\mathbb{X}_t^{\leq n})dt + \sqrt{Q_{k,a}(\mathbb{X}_t^{\leq n})}dW_t,$$

for some linear map $L_{k,b}$ and and some quadratic map $Q_{k,a}$.

$\mathbb{X}^{\leq n}$ is a finite dimensional polynomial process!

Some classics

A super-basic-but-still-useful example: let X be a vector of d correlated Brownian motions. Then the only non zero elements of the expected signature are given by

$$\mathbb{E}[\langle e_0^{\otimes k_0} \otimes e_{J_1} \otimes e_0^{\otimes k_1} \otimes e_{J_2} \otimes \cdots \otimes e_0^{\otimes k_m}, \mathbb{X}_t \rangle] = \frac{t^{\sum_{i=0}^m k_i + \sum_{i=1}^m h_i}}{\left(\sum_{i=0}^m k_i + \sum_{i=1}^m h_i\right)!} \left(\frac{1}{2}\right)^{\sum_{i=1}^m h_i} \prod_{i=1}^m \rho(J_i),$$

where $|J_i| = 2h_i$ and $\rho(J) := \prod_{k=1}^{|J|/2} \rho_{j_{2k-1}, j_{2k}}$.

Black Scholes: Fix

$$dX_t = \sigma X_t dB_t = \sqrt{X_0^2 \sigma^2 \langle e_{\emptyset}, \mathbb{X}_t \rangle + 2X_0 \sigma^2 \langle e_1, \mathbb{X}_t \rangle + 2\sigma^2 \langle e_1 \otimes e_1, \mathbb{X}_t \rangle} dB_t.$$

Then

$$\mathbb{E}[(\langle e_{\emptyset}, \mathbb{X}_t \rangle, \langle e_0, \mathbb{X}_t \rangle, \ldots, \langle e_1 \otimes e_1, \mathbb{X}_t \rangle)] = (1, 0, \ldots, 0) e^{tG^\top}$$

where

$$G^\top = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ X_0^2\sigma^2 & 0 & 2X_0\sigma^2 & 0 & 0 & 0 & 2\sigma^2 \end{pmatrix}.$$

Expected signature for polynomial processes

(i) Write the coefficients a and b in terms of $\mathbb{X}^{\leq 2}$. We need to identify $\mathbf{b}$ and $\mathbf{a}$ such that

$$b(X_t) = \langle \mathbf{b}, \mathbb{X}_t^{\leq 2} \rangle \qquad \text{and} \qquad a(X_t) = \langle \mathbf{a}, \mathbb{X}_t^{\leq 2} \rangle.$$

Example: for $a(X_t) = X_t$ we get

$$a(X_t) = (X_t - X_0) + X_0 = \langle e_1, \mathbb{X}_t^{\leq 2} \rangle + X_0 \langle e_\emptyset, \mathbb{X}_t^{\leq 2} \rangle = \langle X_0 e_\emptyset + e_1, \mathbb{X}_t^{\leq 2} \rangle.$$

(ii) Deduce the linear map L corresponding to the drift of $\langle e_I, \mathbb{X}_t \rangle$:

$$d\langle e_I, \mathbb{X}_t \rangle = \langle e_{I'}, \mathbb{X}_t \rangle \langle \mathbf{b}, \mathbb{X}_t^{\leq 2} \rangle + \frac{1}{2} \langle e_{I''}, \mathbb{X}_t \rangle \langle \mathbf{a}, \mathbb{X}_t^{\leq 2} \rangle,$$

where for $I = (i_1, \ldots, i_n)$ we set $I' = (i_1, \ldots, i_{n-1})$.

(iii) Construct its matrix representation with respect to the basis elements e_I, i.e.: find a matrix G such that

$$G vec(e_I) = vec(L e_I).$$

(iv) Conclude with the moment formula.

Theorem

For $T, t \geq 0$ and each $|I| \leq n$ it holds $\mathbb{E}[vec(\mathbb{X}_{T+t}^n) | \mathcal{F}_T] = e^{tG^\top} vec(\mathbb{X}_T^n)$, i.e.

$$\mathbb{E}[\langle e_I, \mathbb{X}_{T+t}^n \rangle | \mathcal{F}_T] = \sum_{|J| \leq n} (e^{tG^\top})_{IJ} \langle e_J, \mathbb{X}_T^n \rangle.$$

Summarizing: Nice properties of the signature

- Linearity: for each I, J there is a linear combination of indices $I ⧢ J$ such that
$$\langle e_I, \mathbb{X}_t\rangle\langle e_J, \mathbb{X}_t\rangle = \underbrace{\langle e_I ⧢ e_J, \mathbb{X}_t\rangle}_{\text{linear combination of } \mathbb{X}_t\text{'s elements!}}.$$
Every polynomial in the signature has a linear representation! Example:
$$\langle e_1, \mathbb{X}_t\rangle^2 = (X_t)^2 \overset{\text{Itô}}{=} 2\int_0^t X_s dX_s + [X]_t = 2\int_0^t X_s \circ dX_s = 2\langle e_1 \otimes e_1, \mathbb{X}_t\rangle.$$
- Uniqueness: the value of the signature of $\widehat{X}_t := (t, X_t)$ at time T uniquely determines the trajectories of $(X_t - X_0)_{t\in[0,T]}$.
Welcome back Markovianity :).
- Universal approximation theorem: For K compact, $f : K \to \mathbb{R}$ continuous, and $\varepsilon > 0$, there is a finite set $\mathcal{I}$ and $\lambda_I \in \mathbb{R}$ such that
$$|f(\widehat{\mathbb{X}}^2) - \sum_{I\in\mathcal{I}} \lambda_I \langle e_I, \widehat{\mathbb{X}}_T\rangle| 1_{\{\widehat{\mathbb{X}}^2\in K\}} < \varepsilon,$$
almost surely.
Door open for linear approximations!
- Signature of polynomial processes are polynomial processes.
The expected signature can is available in closed form!

Part III.1: Something as...Taylor

(joint work with F.Bandi and R.Renò)

An illustrative example

Let W^1 be the first component of a Brownian motion.

- Consider a stochastic process S admitting the representation

$$S_t = S_0 + \int_0^t c_0(s)ds + \int_0^t c_1(s)dW_s^1.$$

- Suppose that the processes c_0 and c_1 admit the same representation:

$$\begin{aligned} c_0(t) &= c_0(0) + \int_0^t c_{00}(s)ds + \int_0^t c_{10}(s)dW_s^1, \\ c_1(t) &= c_1(0) + \int_0^t c_{01}(s)ds + \int_0^t c_{11}(s)dW_s^1. \end{aligned}$$

- Then

$$\begin{aligned} S_t &= S_0 + \int_0^t \underbrace{c_0(s)}_{=c_{\mathbf{0}}(0)+\int_{\mathbf{0}}^s c_{\mathbf{00}}(r)dr+\int_{\mathbf{0}}^s c_{\mathbf{10}}(r)dW_r^{\mathbf{1}}} ds + \int_0^t \underbrace{c_1(s)}_{=c_{\mathbf{1}}(0)+\int_{\mathbf{0}}^s c_{\mathbf{01}}(r)dr+\int_{\mathbf{0}}^s c_{\mathbf{11}}(r)dW_r^{\mathbf{1}}} dW_s^1 \\ &= S_0 + c_0(0)t + c_1(0)W_t^1 + \underbrace{(\text{linear combination of double integrals})}_{=:\varepsilon_{\mathbf{1}}(t)} \end{aligned}$$

An illustrative example: a further step

$$
\begin{aligned}
S_t &= S_0 + \int_0^t \underbrace{c_0(s)}_{=c_0(0)+\int_0^s c_{00}(r)dr+\int_0^s c_{10}(r)dW_r^1} ds + \int_0^t \underbrace{c_1(s)}_{=c_1(0)+\int_0^s c_{01}(r)dr+\int_0^s c_{11}(r)dW_r^1} dW_s^1, \\
&= S_0 + c_0(0)t + c_1(0)W_t^1 \\
&\quad + \int_0^t \int_0^s c_{00}(r)drds + \int_0^t \int_0^s c_{10}(r)dW_r^1 ds \\
&\quad + \int_0^t \int_0^s c_{01}(r)drdW_s^1 + \int_0^t \int_0^s c_{11}(r)dW_r^1 dW_s^1.
\end{aligned}
$$

- Suppose that the processes c_{00}, c_{01}, c_{10}, and c_{11} admit the same representation:

$$
c_{ij}(s) = c_{ij}(0) + \int_0^s c_{0ij}(r)dr + \int_0^s c_{1ij}(r)dW_r^1.
$$

- Then

$$
\begin{aligned}
S_t &= S_0 + c_0(0)\int_0^t 1ds + c_1(0)\int_0^t 1dW_s^1 \\
&\quad + c_{00}(0)\int_0^t \int_0^s 1drds + c_{10}(0)\int_0^t \int_0^s 1dW_r^1 ds \\
&\quad + c_{01}(0)\int_0^t \int_0^s 1drdW_s^1 + c_{11}(0)\int_0^t \int_0^s 1dW_r^1 dW_s^1 \\
&\quad + (\text{linear combination of triple integrals}) \qquad =: \varepsilon_2(t),
\end{aligned}
$$

An illustrative example: as many steps as we want

Assuming that the procedure can be repeated till depth n and setting $\widehat{W}_t^0 = t$ and $\widehat{W}_t^1 = W_t$ we get

$$S_t = S_0 + \sum_{k=1}^{n} \sum_{(i_1,\dots,i_n)\in\{0,1\}^n} \int_0^t \int_0^{t_n} \cdots \int_0^{t_2} 1 d\widehat{W}_{t_1}^{i_1} \cdots d\widehat{W}_{t_n}^{i_n}$$
$$+(\text{linear combination of } n+1 \text{ iterated integrals}) \qquad =: \varepsilon_n(t)$$

Why is this nice?

- $\varepsilon_n(t)$ is an error term in cases of interests. Example: if $t \mapsto \mathbb{E}[c_{i_1,\dots,i_{n+1}}(t)^{2N}]$ is bounded on $[0,\delta]$, for all $m \leq 2N$ we get

$$\mathbb{E}[|\varepsilon_n(t)|^m] \leq Ct^{m(n+1)/2}.$$

- $S_t - \varepsilon_n(t)$ is linear map of a Markovian process.
- The red building blocks are signature's components: multiplying two of them we obtain a linear combination of them... and many other cool properties!

A bit of signature

For $\widehat{W}_t^0 = t$, $\widehat{W}_t^1 = W_t$, and $I = (i_1, \ldots, i_n)$ we write

$$\langle I, \widehat{\mathbb{W}}_t \rangle := \int_0^t \int_0^{t_n} \cdots \int_0^{t_2} 1 d\widehat{W}_{t_1}^{i_1} \cdots d\widehat{W}_{t_n}^{i_n}.$$

The sequence-valued process $\widehat{\mathbb{W}}$ of all such iterated integrals is called Itô-signature of $\widehat{W}$.

Given some coefficients c_I we also write

$$\langle \mathbf{c}, \widehat{\mathbb{W}}_t \rangle := c_\emptyset + \sum_I c_I \langle I, \widehat{\mathbb{W}}_t \rangle.$$

An illustrative example: as many steps as we want

Assuming that the procedure can be repeated till depth n and setting $\widehat{W}^0_t = t$ and $\widehat{W}^1_t = W_t$ we get

$$S_t = S_0 + \sum_{k=1}^{n} \sum_{(i_1,\dots,i_n)\in\{0,1\}^n} \int_0^t \int_0^{t_n} \cdots \int_0^{t_2} 1 d\widehat{W}^{i_1}_{t_1} \cdots d\widehat{W}^{i_n}_{t_n}$$
$$+(\text{linear combination of } n+1 \text{ iterated integrals}) \quad =: \varepsilon_n(t)$$

Why is this nice?

- $\varepsilon_n(t)$ is an error term in cases of interests. Example: if $t \mapsto \mathbb{E}[c_{i_1,\dots,i_{n+1}}(t)^{2N}]$ is bounded on $[0,\delta]$, for all $m \leq 2N$ we get

$$\mathbb{E}[|\varepsilon_n(t)|^m] \leq Ct^{m(n+1)/2}.$$

- $S_t - \varepsilon_n(t)$ is linear map of a Markovian process.
- The red building blocks are signature's components: multiplying two of them we obtain a linear combination of them. . . and many other cool properties!

Back to Markovianity

Suppose that S is n-times $\widehat{W}$ differentiable, i.e.

$$S_t = \langle c, \widehat{\mathbb{W}}_t \rangle + \varepsilon_n(t),$$

where c is a vector of coefficients $c_I \in \mathbb{R}$ and

$$\varepsilon_n(t) = \sum_{I=(i_1,\ldots,i_{n+1})} \int_0^t \int_0^{t_{n+1}} \cdots \int_0^{t_2} c_I(t_1) d\widehat{W}_{t_1}^{i_1} \cdots d\widehat{W}_{t_{n+1}}^{i_{n+1}},$$

for some stochastic process $t \mapsto c_I(t)$.

If $t \mapsto \mathbb{E}[c_{i_1,\ldots,i_{n+1}}(t)^{2N}]$ is bounded on $[0,\delta]$, for all $m \leq 2N$ we get

$$\mathbb{E}[|\varepsilon_n(t)|^m] \leq Ct^{m(n+1)/2}.$$

This implies that for each differentiable f with $\|f'\| < \infty$ it holds

$$\mathbb{E}[f(S_t)] = \mathbb{E}[f(\langle c, \widehat{\mathbb{W}}_t \rangle)] + o(t^{n/2}).$$

The same applies to moments $f(x) = x^k$ for $k \leq 2N$.

The red term is the expectation of a function applied to the Markov process $\widehat{\mathbb{W}}$!

An illustrative example: as many steps as we want

Assuming that the procedure can be repeated till depth n and setting $\widehat{W}_t^0 = t$ and $\widehat{W}_t^1 = W_t$ we get

$$S_t = S_0 + \sum_{k=1}^{n} \sum_{(i_1,\dots,i_n)\in\{0,1\}^n} \int_0^t \int_0^{t_n} \cdots \int_0^{t_2} 1 d\widehat{W}_{t_1}^{i_1} \cdots d\widehat{W}_{t_n}^{i_n}$$
$$+(\text{linear combination of } n+1 \text{ iterated integrals}) \qquad =: \varepsilon_n(t)$$

Why is this nice?

- $\varepsilon_n(t)$ is an error term in cases of interests. Example: if $t \mapsto \mathbb{E}[c_{i_1,\dots,i_{n+1}}(t)^{2N}]$ is bounded on $[0,\delta]$, for all $m \leq 2N$ we get

$$\mathbb{E}[|\varepsilon_n(t)|^m] \leq C t^{m(n+1)/2}.$$

- $S_t - \varepsilon_n(t)$ is linear map of a Markovian process.
- The red building blocks are signature's components: multiplying two of them we obtain a linear combination of them... and many other cool properties!

Markovianity is useful

Remark: Given a sufficiently integrable Markov process and a sufficiently differentiable map f

$$dY_t := b(Y_t)dt + \sqrt{a(Y_t)}dW_t,$$

by Itô formula it holds

$$f(Y_t) = \int_0^t \mathcal{G}f(Y_s)ds + \text{martingale}$$

for $\mathcal{G}f(y) = b(y)f'(y) + \frac{1}{2}a(y)f''(y)$.

This implies that

$$\frac{d^n}{dt^n}\mathbb{E}[f(Y_t)] = \frac{d^{n-1}}{dt^{n-1}}\mathbb{E}[\mathcal{G}f(Y_t)] = \frac{d^{n-2}}{dt^{n-2}}\mathbb{E}[(\mathcal{G}\mathcal{G}f)(Y_t)] \ldots = \mathbb{E}[(\underbrace{\mathcal{G}\cdots\mathcal{G}}_{n \text{ times}} f)(Y_t)].$$

By Taylor we can conclude that

$$\mathbb{E}[f(Y_t)] = f(Y_0) + \mathcal{G}f(Y_0)t + \mathcal{G}\mathcal{G}f(Y_0)\frac{t^2}{2} + \ldots + \underbrace{\mathcal{G}\cdots\mathcal{G}}_{n \text{ times}} f(Y_0)\frac{t^n}{n!} + o(t^n).$$

Markovianity is useful in our setting

Fix n even. For a sufficiently differentiable map f, we know that

$$\mathbb{E}[f(S_t)] = \mathbb{E}[f(\langle c, \widehat{\mathbb{W}}_t \rangle)] + o(t^{n/2}),$$

and $\widehat{\mathbb{W}}$ is Markov.

Setting $f_c := f(\langle c, \cdot \rangle)$ we thus obtain

$$\mathbb{E}[f(S_t)] = f_c(\widehat{\mathbb{W}}_0) + \mathcal{G} f_c(\widehat{\mathbb{W}}_0) t + \mathcal{G}\mathcal{G} f_c(\widehat{\mathbb{W}}_0) \frac{t^2}{2} + \ldots + \underbrace{\mathcal{G} \cdots \mathcal{G}}_{n/2 \text{ times}} f_c(\widehat{\mathbb{W}}_0) \frac{t^{n/2}}{(n/2)!} + o(t^{n/2}),$$

where $\mathcal{G}$ denotes the generator of $\widehat{\mathbb{W}}$.

Everything is explicit

One can show that $\mathcal{G}$ is mapping functions of the form $f(\langle c, \cdot\rangle)(\langle d, \cdot\rangle)$ to linear combination of functions of the form

$$f^{(k)}(\langle c, \cdot\rangle)(\langle \mathcal{G}_c^k(d), \cdot\rangle)$$

for some bilinear operator $\mathcal{G}^k$ with $k = 0, 1, 2$.

Theorem

Consider an n-times $\widehat{W}$-differentiable process $(S_t)_{t\in[0,T]}$ with expansion

$$S_t = \langle c, \widehat{\mathbb{W}}_t\rangle + \varepsilon_n(t).$$

Then for each $f \in C^{n+1}(\mathbb{R})$ it holds

$$\mathbb{E}[f(S_t)] = f(S_0) + \sum_{\ell=1}^{\lceil n/2\rceil} \frac{1}{\ell!}\left(\sum_{k_1,\ldots,k_\ell=0}^{2} f^{(k_1+\ldots+k_\ell)}(S_0)\mathcal{G}_{c,k_1,\ldots,k_\ell}(\emptyset)_\emptyset\right)t^\ell + o(t^{n/2}),$$

where $\mathcal{G}_{c,k_1,\ldots,k_n}(d) = \mathcal{G}_c^{k_n}(\cdots(\mathcal{G}_c^{k_1}(d)))$.

Good news: Bilinear maps are easy to code!
Given $f(S_0)$, $f'(S_0)\ldots$, $f^{(n+1)}(S_0)$ the coefficients can be computed by a computer.

Example

Suppose that S is 2 times $\widehat{W}$-differentiable with

$$\begin{aligned} S_t &= S_0 + \int_0^t \underbrace{c_0(s)}_{=c_0(0)+\int_0^s c_{00}(r)dr+\int_0^s c_{10}(r)dW^1_r} ds + \int_0^t \underbrace{c_1(s)}_{=c_1(0)+\int_0^s c_{01}(r)dr+\int_0^s c_{11}(r)dW^1_r} dW^1_s \\ &= S_0 + c_0(0)t + c_1(0)W^1_t + \ldots + c_{11}(0)\int_0^t W^1_s dW^1_s + \varepsilon_2(t). \end{aligned}$$

Then applying the theorem for $n = 2$ yields

$$\mathbb{E}[e^{iu(S_t - S_0)^2}] = 1 + iu(c_1(0))^2 t + o(t).$$

Generalizations

The result generalizes to

- $f(x) = x^k$ for $k \leq 2N$;
- a d-dimensional Brownian motion;
- the framework where the driver is given by e compound Poission processes and d Brownian motions.

Also, as the knowledge on the Brownian motion goes much beyond the generator of its signature, the door is open to many more applications.

Example: Edgeworth expansion

$$\mathbb{E}\Big[\exp\Big(iu\frac{S_t - c_0 t}{c_1\sqrt{t}}\Big)\Big]e^{\frac{u^2}{2}}$$

$$= 1 + \Big[-\frac{c_{11}}{c_1}\frac{i}{2}u^3\Big]\sqrt{t}$$

$$+\frac{1}{2}\Big[-\Big(\frac{c_{01}}{c_1}+\frac{c_{10}}{c_1}\Big)u^2 + \Big(\frac{c_{11}}{c_1}\Big)^2\Big(-\frac{1}{2}u^2 + u^4 - \frac{1}{4}u^6\Big) + \Big(\frac{c_{21}}{c_1}\Big)^2\Big(\frac{1}{3}u^4 - \frac{1}{2}u^2\Big)$$

$$-\frac{c_{111}}{c_1}\frac{i}{6}u^3\Big]t + o(t).$$

Part III.2: Something as...Stone-Weierstrass

(Joint work with C.Cuchiero, G.Gazzani, J.Möller, J.Teichmann)

Outline

- The model
- Calibration to option prices and discussion of the performance
- What about including the VIX?
- Joint calibration of SPX and VIX options
- Conclusion and outlooks

The model

The model

Goal: provide a *good* model for a set of *traded assets* $S = (S^1, \dots, S^D)$.
$\rightarrow$*good* = universal, tractable, and easy to calibrate.

Main ingredient: the *market's primary (underlying) process* $\widehat{X}_t := (t, X_t)$.

Requested properties:

- The realizations of $\widehat{X}$ are available in form of time series data and/or the law of $\widehat{X}$ under the pricing measure is known.
- It is reasonable to assume that:
  - X is d-dimensional continuous semimartingale.
  - $\widehat{X}$ encodes all the randomness of S in a good way, meaning that the paths of S are continuous maps of the paths of $\widehat{X}$.

The model: $S_n(\ell)_t = (S_n^1(\ell^1)_t, \dots, S_n^D(\ell^D)_t)$, where

$$S_n^j(\ell^j)_t := \ell_\emptyset^j + \sum_{0<|I|\le n} \ell_I^j \langle e_I, \widehat{\mathbb{X}}_t \rangle,$$

- $\widehat{\mathbb{X}}$ is the signature of $\widehat{X}$,
- $n \in \mathbb{N}$ is the degree of truncation,
- $\ell_\emptyset^j, \ell_I^j \in \mathbb{R}$ are the deterministic coefficients to be found.

See also Perez Arribas, Salvi, Szpruch ('20).

In one sentence: the model

$$S^j_n(\ell^j)_t := \ell^j_{\emptyset} + \sum_{0<|I|\leq n} \ell^j_I \langle e_I, \widehat{\mathbb{X}}_t \rangle,$$

is a linear model whose parameters are ℓ^j_I and whose building blocks are

$$\langle e_I, \widehat{\mathbb{X}}_t \rangle = \int_0^t \int_0^{t_n} \cdots \int_0^{t_2} 1 \, d\widehat{X}^{i_1}_{t_1} \cdots d\widehat{X}^{i_n}_{t_n}$$

for some continuous semimartingale $\widehat{X} = (\widehat{X}^0, \widehat{X}^1, \ldots, \widehat{X}^d)$.

The model: $S_n(\ell)_t := \ell_\emptyset + \sum_{0<|I|\leq n} \ell_I \langle e_I, \widehat{\mathbb{X}}_t \rangle$ $(D = 1)$

Flexibility: From the UAT S can be approximated by $S_n(\ell)$.

Universality: Any classical model driven by Brownian motions can be arbitrarily well approximated. Extensions to Lévy driven models are possible (joint work with F. Primavera).

Classical requirements: No arbitrage can easily be guaranteed.

Tractability: Time extended signature of $S_n(\ell)$ can be written as map of $(\ell, \widehat{\mathbb{X}})$.
$\longrightarrow$ Knowing $\mathbb{E}_\mathbb{Q}[\widehat{\mathbb{X}}_t]$, computing an approximation of the price of (path-dependent) options reduces to evaluating a polynomial. Mathematically:

$$\mathbb{E}_\mathbb{Q}[F((S_n(\ell)_t)_{t\in[0,T]})] \approx P(\ell, \mathbb{E}_\mathbb{Q}[\widehat{\mathbb{X}}_t]),$$

for some some P such that $P(\,\cdot\,, \mathbb{E}_\mathbb{Q}[\widehat{\mathbb{X}}_T])$ is polynomial.

$\longrightarrow$ Formulas for the computations of $\mathbb{E}_\mathbb{Q}[\widehat{\mathbb{X}}_t]$ are available if X is a sufficiently regular Markov (or non Markov) diffusion.

Calibration to option prices

Calibration to option prices

Model: $S_{n+1}(\ell)_t := S_{n+1}(\ell)_0 + \ell_\emptyset \langle \tilde{e}_\emptyset, \widehat{\mathbb{X}}_t \rangle + \sum_{0<|I|\leq n} \ell_I \langle \tilde{e}_I, \widehat{\mathbb{X}}_t \rangle$. $(D = d = 1)$

Scenario: The following quantities are available:

- Prices of options on S.
- The law of the *market's primary (underlying) process* $\widehat{X}$ under the pricing measure $\mathbb{Q}$.

Cool idea: Since computing the approximated price of an (even path dependent) option with the proposed model reduces to evaluating a polynomial, calibration on (even path dependent) option prices could be done in a simple and efficient way.

$\longrightarrow$ ...cool but dangerous! The given approximation has to be good enough in each optimization's step!

Alternative idea: Use Monte Carlo pricing (with variance reduction). Note that there is no need of new simulations in the optimization procedure.

Calibration to option prices: procedure

Scenario: The following quantities are available:

- Prices $\pi_1, \dots, \pi_N$ of N options with payoffs

$$F_1((S_t)_{t\in[0,T_1]}), \dots, F_N((S_t)_{t\in[0,T_N]}).$$

Procedure:

- Look for ℓ matching the corresponding option prices, i.e. minimizing the expression

$$\sum_{i=1}^{N} w^i \left(P_i^{MC}(\ell) - \pi^i \right)^2,$$

for some weights w^i, where $P^{MC}(\ell)$ denotes the empirical mean of

$$F_i\Big((S_n(\ell)_t)_{t\in[0,T_i]} \Big).$$

Important observation: the linearity of the model makes this procedure very quick. Trajectories of $\widehat{\mathbb{X}}$ could be simulated just once in advance and stored. A coefficients update reduces to a scalar product.

Calibration to option prices: the Heston model

- Consider a Heston model (d=2, D=1):

$$dS_t = \mu S_t dt + S_t \sqrt{V_t} dB_t^{\mathbb{P}}$$
$$dV_t = \kappa(\theta - V_t)dt + \sigma\sqrt{V_t} dW_t^{\mathbb{P}},$$

- Goal: approximate S with $S_3(\ell^*)$, using two $\mathbb{Q}$-Brownian motions as primary underlying process ($\ell^* \in \mathbb{R}^{13}$).
- Test: Compute the implied volatility surface (using Monte Carlo) under $S_3(\ell^*)$ (red) and compare it with the Heston's one (blue).

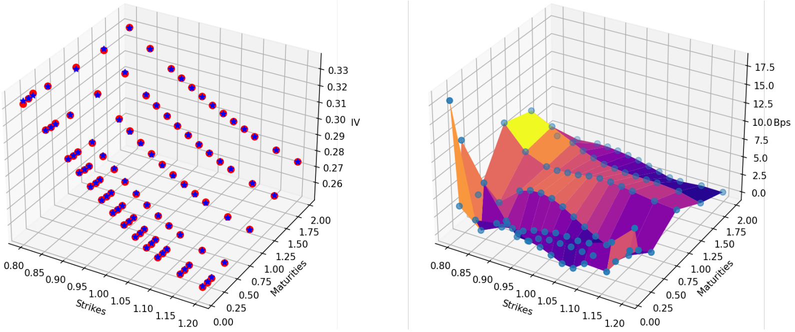


Figura: IVSs and corresponding absolute error (7 maturities from 30 days to 2 years).

Calibration to option prices: S&P 500 17.03.2021

- Let S be the stochastic process describing the price of S&P 500 starting at day 17.03.2021.
- Goal: approximate S with $S_4(\ell^*)$, using two $\mathbb{Q}$-Brownian motions as primary underlying process ($\ell^* \in \mathbb{R}^{121}$).
- Test: Compute the implied volatility surface (using Monte Carlo) under $S_4(\ell^*)$ and compare it with the market's one.

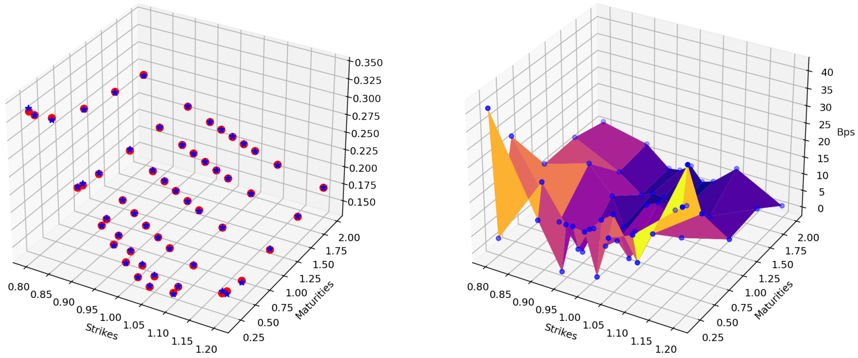


Figura: IVSs and corresponding absolute error (6 maturities within 60 days and 2 years).

Remarks on the previous example

- The result is obtained using a closer-to-sup-norm loss function:

$$\sum_{i=1}^{N} \alpha\varepsilon^{i}\left(P_i^{MC}(\ell) - \pi^{i}\right)^{p},$$

  where ε_i is the absolute error for the i-th price in a previous calibration and α and p are big.
- The calibrated model produces a reasonable implied volatility surface also for out of sample strikes and maturities.

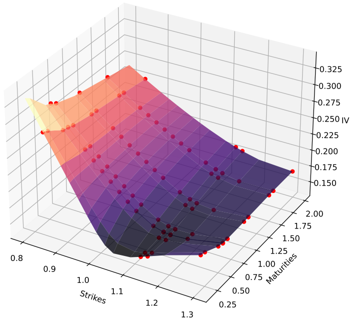


Figura: IVS of the calibrated model (6 maturities within 60 days and 2 years). Out of sample represented as red dots.

What about including the VIX?

The VIX Index

The CBOE Volatility Index (VIX) is a popular measure of the market's expected volatility on the S&P-500 Index, calculated and published by the Chicago Board Options Exchange (CBOE).

The current VIX index value quotes the expected annualized change in the SPX-500 over the following 30 days, based on options-based theory and current options-market data, more precisely

$$\mathsf{VIX}_T := \sqrt{\mathbb{E}\left[-\frac{2}{\Delta}\log\left(\frac{S_{T+\Delta}}{S_T}\right)|\mathcal{F}_T\right]},$$

where $\Delta = 30$ days and $S = (S_t)_{t\geq 0}$ denotes the S&P-500 index.

The model

We consider a model where the dynamics of the S&P-500 index and the corresponding volatility process are given by

$$\begin{aligned} dS_t &= S_t \sigma_t^S dB_t \\ \sigma_t^S &= \ell_\emptyset + \sum_{0<|I|\leq n} \ell_I \langle e_I, \widehat{\mathbb{X}}_t \rangle, \end{aligned}$$

where

- $B = (B_t)_{t\geq 0}$ is a one-dimensional Brownian motion
- $n \in \mathbb{N}$
- $X = (X^1, \dots, X^d)$ is a d-dimensional continuous semimartingale. Denoting $Z = (X, B)$, then the correlation matrix between X and B is given by

  $$\rho_{i,j} = \frac{[Z^i, Z^j]}{\sqrt{[Z^i]}\sqrt{[Z^j]}} \in [-1, 1],$$

  for all $i, j = 1, \dots, d+1$, where $[\cdot, \cdot]$ denotes the quadratic covariation.
- $\ell := \{\ell_I \in \mathbb{R} : |I| \leq n\}$ the collection of parameters of the model.

A reasonable choice for the primary process

Our choice for the primary process goes back to the good old polynomial diffusions: we assume that

$$dX_t = b(X_t)dt + \sqrt{a(X_t)}dW_t,$$

where

- b, a are polynomial of order one and two, respectively.
- $W = (W_t)_{t\geq 0}$ is a d-dimensional Brownian motion.

Why? Because they are a wide class and the truncated time extended signature

$$\widehat{\mathbb{X}}_t^{\leq N} := (1, \langle e_0, \widehat{\mathbb{X}}_t\rangle, \dots, \langle e_{(d,\dots,d)}, \widehat{\mathbb{X}}_t\rangle)$$

of $(X_t)_{t\geq 0}$ is a polynomial process too. We can thus compute its conditional moments by means of a matrix exponential.

But why do we care? Because the VIX index on S can be re-written as

$$\mathrm{VIX}_T = \sqrt{\frac{1}{\Delta}\mathbb{E}\left[\int_T^{T+\Delta}(\sigma_t^S)^2 dt \middle| \mathcal{F}_T\right]},$$

and $\int_T^{T+\Delta}(\sigma_t^S)^2 dt$ is a polynomial (of degree 1) in $\widehat{\mathbb{X}}_T^{2n+1}$ and $\widehat{\mathbb{X}}_{T+\Delta}^{2n+1}$.

VIX with signatures

Theorem

Under the proposed model

$$\mathrm{VIX}_T(\ell) = \sqrt{\frac{1}{\Delta}\ell^\top Q(T)\ell},$$

where,

$$\begin{aligned} Q_{IJ}(T) &= \mathbb{E}\big[\langle (e_I \sqcup\!\sqcup\, e_J) \otimes e_0, \widehat{\mathbb{X}}_{T+\Delta}\rangle | \mathcal{F}_T\big] - \langle (e_I \sqcup\!\sqcup\, e_J) \otimes e_0, \widehat{\mathbb{X}}_T\rangle \\ &= ((e_I \sqcup\!\sqcup\, e_J) \otimes e_0)^\top (e^{\Delta G^\top} - \mathrm{Id})\widehat{\mathbb{X}}_T^{2n+1}, \end{aligned}$$

with G being the matrix associated to the generator $\mathcal{G}$ of $\widehat{\mathbb{X}}^{\leq 2n+1}$.

Observe that we use the same notation for elements of the tensor algebra and their vectorisation.

VIX options

It is important to note that:

- The VIX is not a martingale.
- VIX options are therefore *written on future contracts.*

Recall that for a VIX-Future of maturity T, it's *settlement price* at time $t \in [0, T]$ is given by

$$F_t^T := \mathbb{E}\left[\mathrm{VIX}_T \; e^{r(t-T)} | \mathcal{F}_t\right].$$

In particular,

- At maturity it holds that $F_T^T = \mathrm{VIX}_T$.
- The spot price of a VIX-option depends on the maturity, since the spot price is $F_0^T = \mathbb{E}\left[\mathrm{VIX}_T e^{-rT}\right]$.
- During calibration, also the Futures' prices should be calibrated.

Calibration task

Let $\mathcal{T}$ be a set of maturities and $\mathcal{K}$ a set of strikes. Introduce

$$\pi_{\text{VIX}}^{\text{model}}(\ell, T, K) := \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} (\text{VIX}_T(\ell, \omega_i) - K)^+, \qquad F_{\text{VIX}}^{\text{model}}(\ell, T) := \frac{e^{-rT}}{N_{MC}} \sum_{i=1}^{N_{MC}} \text{VIX}_T(\ell, \omega_i).$$

Then the loss function reads as follows:

$$L_{\text{VIX}}(\ell) := \sum_{T \in \mathcal{T}, K \in \mathcal{K}} \mathcal{L}\left(\pi_{\text{VIX}}^{\text{model}}(\ell, T, K), \pi_{\text{VIX}}^{b,a}(T, K), \sigma_{\text{VIX}}^{b,a}(T, K), F_{\text{VIX}}^{\text{model}}(\ell, T), F_{\text{VIX}}^{mkt}(T)\right)$$

where $\pi_{\text{VIX}}^{b,a}(T, K), \sigma_{\text{VIX}}^{b,a}(T, K), F_{\text{VIX}}^{mkt}(T)$ denote market's option bid/ask prices, implied volatilities and market's futures' prices, respectively.

Our choice:

$$\mathcal{L}^{\beta}(\pi, \pi^{mkt,b,a}, \sigma^{mkt,b,a}, F, F^{mkt}) =$$

$$\left(\frac{\left(\beta \tilde{1}_{\{\pi \notin [\pi^{mkt,b}, \pi^{mkt,a}]\}} + (1-\beta)\right) \left|\pi - (\pi^{mkt,a} + \pi^{mkt,b})/2\right| + \left|\delta^{mkt} e^{-rT}(F - F^{mkt})\right|}{\upsilon^{mkt}(\sigma^{mkt,a} - \sigma^{mkt,b})} \right)^2,$$

where

- υ^{mkt} and δ^{mkt} denote the Vega and Delta of the option under the Black-Scholes model which depend on the maturity and on the strike price;
- F and F^{mkt} denote futures with maturity T;
- $\tilde{1}_{\{x \notin [y^b, y^a]\}} := s(y^b - x) + s(x - y^a)$ for $s(x) := \frac{1}{2}\tanh(100x) + \frac{1}{2}$ a smooth version of the indicator function.

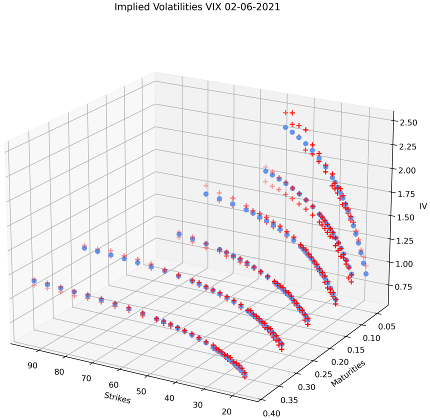
Primary process

For X we choose a 2-dimensional Ornstein-Uhlenbeck processes

$$dX_t^j = \kappa^j(\theta^j - X_t^j)dt + \sqrt{a(X_t)}dW_t, \qquad X_0 = x_0,$$

for $a_{ij}(X_t) = \sigma^i \sigma^j \rho_{ij}$, and W being a d-dimensional Brownian motion.

Implied volatility



$T_1 = 0.0383$	$T_2 = 0.0767$	$T_3 = 0.1342$	$T_4 = 0.2108$	$T_5 = 0.2875$	$T_6 = 0.3833$
(90%,250%)	(90%,250%)	(80%,310%)	(80%,300%)	(75%,395%)	(80%,405%)

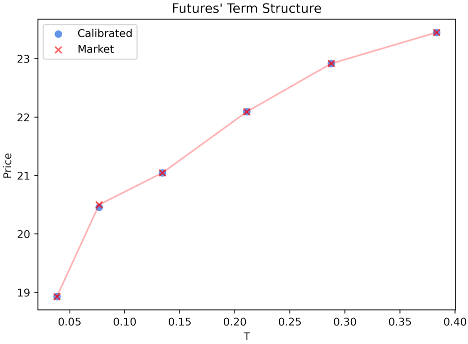
$$d = 2, \quad n = 3, \quad \kappa = (0.1, 25)^\top, \quad \theta = (0.1, 4)^\top, \quad \sigma = (0.7, 10)^\top, \quad \rho = \begin{pmatrix} 1 & -0.577 & 0.3 \\ \cdot & 1 & -0.6 \\ \cdot & \cdot & 1 \end{pmatrix}$$

Future prices

Relative absolute error between the market future prices and the calibrated ones:

$T_1 = 0.0383$	$T_2 = 0.0767$	$T_3 = 0.1342$
$\varepsilon_{T_1} = 7.0 \times 10^{-6}$	$\varepsilon_{T_2} = 2.1 \times 10^{-3}$	$\varepsilon_{T_3} = 1.3 \times 10^{-5}$

$T_4 = 0.2108$	$T_5 = 0.2875$	$T_6 = 0.3833$
$\varepsilon_{T_4} = 1.5 \times 10^{-4}$	$\varepsilon_{T_5} = 1.9 \times 10^{-6}$	$\varepsilon_{T_6} = 1.3 \times 10^{-6}$



Joint calibration of SPX and VIX options

SPX dynamics

Theorem

Under our model it holds

$$S_t(\ell) = S_0 \exp\left\{ -\frac{1}{2}\ell^\top Q^0(t)\ell + \sum_{|I|\le n} \ell_I \langle \tilde{e}_I{}^B, \widehat{\mathbb{Z}}_t \rangle \right\},$$

where $Z_t = (X_t, B_t)$ and $\tilde{e}_I{}^k$ are linear combinations of indices. The components of the matrix $Q^0(t)$ are given by

$$Q^0_{IJ}(t) = \langle (e_I \sqcup\mathchoice{}{}{}{}\sqcup e_J) \otimes e_0, \widehat{\mathbb{X}}_t \rangle.$$

Joint calibration of SPX and VIX options

Denote by $L_{\mathsf{SPX}}(\ell)$ the SPX loss function, where

$$L_{\mathsf{SPX}}(\ell) := \sum_{T\in\mathcal{T}, K\in\mathcal{K}} \mathcal{L}^{\beta}(\pi^{\mathsf{model}}_{\mathsf{SPX}}(\ell, T, K), \pi^{mkt,b}_{\mathsf{SPX}}(T, K), \pi^{mkt,a}_{\mathsf{SPX}}(T, K))$$

To achieve a joint calibration of the SPX/VIX options and VIX futures we minimize

$$L_{\mathsf{joint}}(\ell, \lambda) := \lambda L_{\mathsf{SPX}}(\ell) + (1-\lambda) L_{\mathsf{VIX}}(\ell),$$

for some $\lambda \in (0, 1)$.

Parameters specifications

Maturity and moneyness specifications:

$T_1^{\text{VIX}} = 0.0383$	$T_2^{\text{VIX}} = 0.0767$
(90%,220%)	(90%,220%)

$T_1^{\text{SPX}} = 0.0383$	$T_2^{\text{SPX}} = 0.1205$	$T_3^{\text{SPX}} = 0.1588$
(92%,105%)	(70%,105%)	(80%,120%)

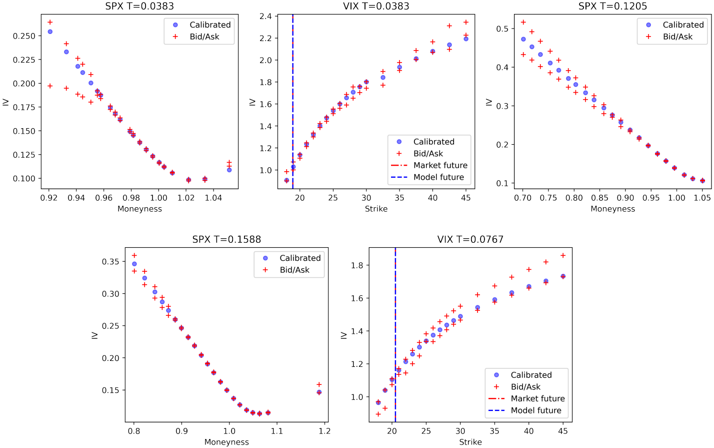
The shortest maturity considered is of 14 days for both SPX and VIX, then the second and third maturity of the SPX are 44 days and 58 days, respectively, and the second one for the VIX is 28 days. Moreover, we consider a high moneyness level (up to 220%) for VIX options, usually rather difficult to fit.

Models' parameters: $d = 3$, $n = 3$ (hence $\ell \in \mathbb{R}^{85}$), $\lambda = 0.35$, $\beta = 1$, and X being a three dimensional OU process with

$$\kappa = (0.1, 25, 10)^\top, \qquad \theta = (0.1, 4, 0.08)^\top, \qquad \sigma = (0.7, 10, 5)^\top,$$

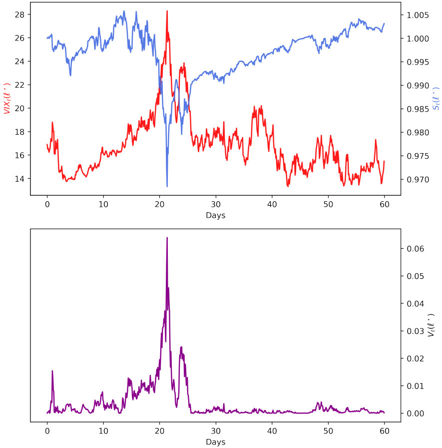
$$\rho = \begin{pmatrix} 1 & 0.213 & -0.576 & 0.329 \\ \cdot & 1 & -0.044 & -0.549 \\ \cdot & \cdot & 1 & -0.539 \\ \cdot & \cdot & \cdot & 1 \end{pmatrix}, \qquad X_0 = (1, 0.08, 2)^\top.$$

Implied volatility and future prices



Trajectories

Let $\ell^\star \in \mathbb{R}^{85}$ be the calibrated parameters, fix $T = 60$ days, and sample a trajectory for $(V_t(\ell^\star))_{t\in[0,T]}$ with $V_t = \sigma_t^S(\ell^\star)^2$, $(\mathrm{VIX}_t(\ell^\star))_{t\in[0,T]}$, and $(S_t(\ell^\star))_{t\in[0,T]}$.



Conclusions

Conclusions

- We saw that from a mathematical point of view signatures have some extremely interesting properties and deserve to be used in a modeling context.

⇒ $F((X_t)_{t \in [0,T]}) \approx L(\widehat{\mathbb{X}}_T)$ for some linear map L.

- We introduced a linear model based on the signature of an underlying process.

⇒ Flexible: classical models can be approximated arbitrarily well.

⇒ Tractable: since as soon as $\mathbb{E}_{\mathbb{Q}}[\widehat{\mathbb{X}}]$ is known, estimators for different quantities are available in closed form.

- We illustrated a calibration method and the corresponding performance on simulated and real data.
- We show how to extend this approach to obtain a tractable and flexible representation of prices and corresponding VIX.
- We illustrated a calibration method and the corresponding performance on real data.

Thank you for your attention!