Path-Signatures: Memory and Stationarity

Eduardo Abi Jaber

Based on joint works with

Paul Gassiat (Université Gustave Eiffel)
Louis-Amand Gérard (GEFIP)
Yuxing Huang (New York University)
Camille Illand (BNP)
Shaun Li (Morgan Stanley)
Xuyang Lin (New York University)
Dimitri Sotnikov (Ecole Polytechnique)

Winter School on Mathematical Finance - Soesterberg

January 21, 2026

Memory Matters

Memory-Aware modeling

Memory matters!

Many dynamical systems exhibit **path-dependent** behaviors such as

- long/short-memory effects
- lead-lag relationships
- multiple time scales

Memory-Aware modeling

Memory matters!

Many dynamical systems exhibit **path-dependent** behaviors such as

- long/short-memory effects
- lead-lag relationships
- multiple time scales

Main Question How to model these **path-dependent** (non-Markovian) effects in a mathematically tractable way?

Memory-Aware modeling

Big picture

- In a random environment: stochastic model
- Based on available information: $\mathcal{F}_t$
- Determine actions: hedging strategy, optimal investment/liquidation, etc.
- Evaluate rewards / quantities of interest: $\mathbb{E}[\xi \mid \mathcal{F}_t]$, option prices, etc.:

Memory-Aware modeling

Big picture

- In a random environment: stochastic model
- Based on available information: $\mathcal{F}_t$
- Determine actions: hedging strategy, optimal investment/liquidation, etc.
- Evaluate rewards / quantities of interest: $\mathbb{E}[\xi \mid \mathcal{F}_t]$, option prices, etc.:

Summarize then Linearize

1. Identify good **"Markovian"** variables to capture the available information such that the quantities of interest become

$$f\left(t, Y_t^1, \ldots, Y_t^n\right)$$

2. **Ideally**, variables that **linearize** the problem (semi-explicit formulas, fast computation). For example, a **polynomial** function of a factor Y is a **linear** function of the extended vector of monomials $(1, Y, Y^2, Y^3, \ldots, Y^m)$.

Main Question How to model these **path-dependent** (non-Markovian) effects in a mathematically tractable way?

Memory-Aware modeling

Main Question How to model these **path-dependent** (non-Markovian) effects in a mathematically tractable way?

Mathematical tools for memory-aware modeling

- Volterra processes
- Path-signatures

Named after the Italian mathematician **Vito Volterra** (1860–1940): mathematical biology and integral equations, one of the founders of functional analysis

Volterra processes are the continuous-time analogue of moving averages:

- Discrete-time intuition: weighted averages

$$\sum_{i=1}^{N} e^{-\lambda(t-t_i)} \Delta_i Z$$

- **Volterra processes** (continuous time)

$$X_t = \int_0^t K(t-s)\, dZ_s$$

- The kernel K encodes *memory* and persistence
- Timely in math finance

Vito Volterra

Introduced by the mathematician **Kuo Tsai Chen** (1923–1987): algebraic topology and analysis
Path signatures are sequence of **iterated integrals** associated with a path.

- For a path Z, its (truncated) signature is

$$\left(1,\ \int dZ,\ \int dZ \otimes dZ,\ \ldots\right)$$

- An algebraic object encoding the *entire path*
- **Analogue of polynomials on path space**

Central in rough paths theory (Lyons), controlled differential equations, machine learning on time series, and emerging financial applications.

Kuo Tsai Chen
https://sites.google.com/view/abijabereduardo/

Why do we love polynomials?

- **Universal approximators**
  - **Stone–Weierstrass** uniform approximation on compact sets.
  - **Taylor expansion** local approximation with explicit coefficients and quantitative error.
- **Linearization via lifting** for $u : [0, T] \times \mathbb{R} \to \mathbb{R}$:

$$u(t, x) \approx \sum_{i=0}^{M} \alpha_i(t)\, x^i = \langle \alpha(t), \boldsymbol{x} \rangle, \qquad \boldsymbol{x} := (1, x, x^2, \dots, x^M)$$

  Nonlinearity in x becomes linearity in the lifted state $\boldsymbol{x}$.
- **Products are linearized** thanks to **Cauchy's product**

$$\sum_{i=0}^{M} \alpha_i(t) x^i \sum_{j=0}^{N} \beta_j(t) x^j = \sum_{k=0}^{M+N} \gamma_k(t) x^k = \langle \gamma(t), \mathbf{x} \rangle, \quad \text{with} \quad \gamma_k(t) = \sum_{j=0}^{k} \alpha_{k-j}(t)\, \beta_j(t)$$

- **Universal approximators**
  - **Stone–Weierstrass** uniform approximation on compact sets.
  - **Taylor expansion** local approximation with explicit coefficients and quantitative error.
- **Linearization via lifting** for $u : [0, T] \times \mathbb{R} \to \mathbb{R}$:

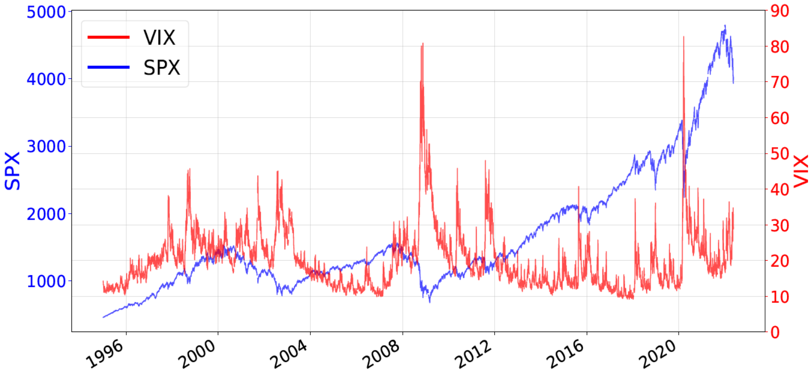
$$u(t,x) \approx \sum_{i=0}^{M} \alpha_i(t)\, x^i = \langle \alpha(t), \boldsymbol{x} \rangle, \qquad \boldsymbol{x} := (1, x, x^2, \dots, x^M)$$

  Nonlinearity in x becomes linearity in the lifted state $\boldsymbol{x}$.
- **Products are linearized** thanks to **Cauchy's product**

$$\sum_{i=0}^{M} \alpha_i(t) x^i \sum_{j=0}^{N} \beta_j(t) x^j = \sum_{k=0}^{M+N} \gamma_k(t) x^k = \langle \gamma(t), \mathbf{x} \rangle, \quad \text{with} \quad \gamma_k(t) = \sum_{j=0}^{k} \alpha_{k-j}(t)\, \beta_j(t)$$

Summarize then Linearize

Okay... but can polynomials be useful in finance ?

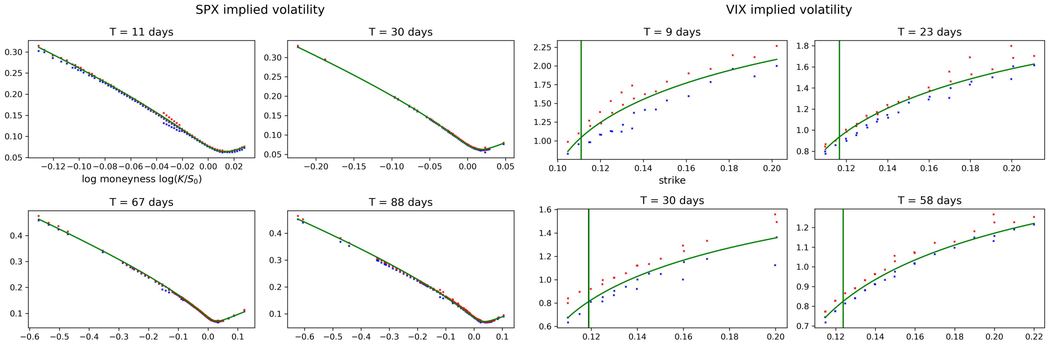


- **VIX**: the "**fear**" index that reflects market's expectations for volatility of the **S&P 500** over the next 30 days

- Related literature strongly agrees that *conventional (parametric) one-factor continuous Markovian stochastic volatility models* are not able to achieve a decent joint calibration
- Our main motivations can be stated as follows:

Can joint calibration be achieved by a ***simple*** *model?*

If so, can we do it in a ***tractable*** *way?*



Joint calibration of SPX IV, VIX IV and VIX futures on 23 October, 2017 using the exponential kernel K^{exp} with 6 parameters. The blue and red dots are bid/ask implied vol, green lines are model fit. Vertical bars represents VIX futures price.

Characteristics of iv: steeper SPX slopes for small T, upward slope for VIX, difficulty to match levels.

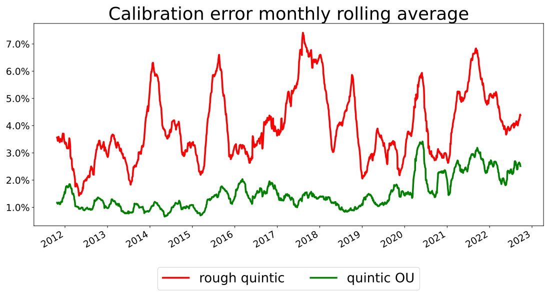
The Quintic Ornstein-Uhlenbeck Model

Ref

- *The quintic Ornstein-Uhlenbeck volatility model that jointly calibrates SPX & VIX smiles,* with **Camille Illand** and **Shaun Li**, Risk Magazine, Cutting Edge section (2023).
- *Joint SPX-VIX calibration with Gaussian polynomial volatility models: deep pricing with quantization hints* with **Camille Illand** and **Shaun Li**, Mathematical Finance (2025).

$$\frac{dS_t}{S_t} = \sigma_t dB_t, \quad S_0 > 0, \quad B = \rho W + \sqrt{1-\rho^2} W^{\perp},$$
$$\sigma_t = g_0(t) p(X_t), \quad p(x) = \alpha_0 + \alpha_1 x + \alpha_3 x^3 + \alpha_5 x^5,$$
$$X_t = \eta \int_0^t e^{-\lambda(t-s)} dW_s,$$

- non-negative coefficients $\alpha_0, \alpha_1, \alpha_3, \alpha_5 \geq 0$,
- p polynomial of degree five to reproduce upward slope of the VIX smile. Restricting α to be non-negative allows the sign of the leverage effect to be the same as ρ.
- input curve g_0 allowing the model to match term-structures observed on the market.



Flexibility

- Remarkable joint fits of SPX-VIX volatility surfaces (**1 week to 3 months**), daily calibration across more than 10 years of data
- Consistently outperforms in **all market conditions** more complex (rough) models:

$$X_t = \int_0^t (t-s)^{H-1/2} dW_s, \quad H < 1/2.$$

fractional (and singular) kernels do not align well with market data on short term.

- Deep pricing with quantization hints method

$$\frac{dS_t}{S_t} = \sigma_t dB_t, \quad S_0 > 0, \quad B = \rho W + \sqrt{1-\rho^2} W^{\perp},$$
$$\sigma_t = g_0(t) p(X_t), \quad p(x) = \alpha_0 + \alpha_1 x + \alpha_3 x^3 + \alpha_5 x^5,$$
$$X_t = \eta \int_0^t e^{-\lambda(t-s)} dW_s,$$

Tractability

- Explicit expression for the VIX

$$\text{VIX}_T^2 := \frac{100^2}{\Delta} \int_T^{T+\Delta} \mathbb{E}[\sigma_u^2 \mid \mathcal{F}_T] du,$$

  with $\Delta = 30$ days, which is polynomial in the OU process X_T. $\Rightarrow$ Pricing VIX products by integration against Gaussian density
- Pricing of SPX products: Simulation of X is exact. Other alternatives than MC or PDE methods?

Faster and more accurate pricing of SPX derivatives?

Ref

- *Fourier-Laplace transforms in polynomial Ornstein-Uhlenbeck volatility models,* with **Shaun Li** and **Xuyang Lin**, Finance & Stochastics (2025).

$$\frac{dS_t}{S_t} = \sigma_t dB_t, \quad S_0 > 0, \quad B = \rho W + \sqrt{1-\rho^2} W^{\perp},$$

$$\sigma_t = g_0(t) p(X_t) \quad p(x) = \sum_{k=0}^{\infty} \alpha_k x^k,$$

$$dX_t = (a + bX_t)dt + cdW_t.$$

$$\frac{dS_t}{S_t} = \sigma_t dB_t, \quad S_0 > 0, \quad B = \rho W + \sqrt{1-\rho^2} W^{\perp},$$
$$\sigma_t = g_0(t) p(X_t) \quad p(x) = \sum_{k=0}^{\infty} \alpha_k x^k,$$
$$dX_t = (a + bX_t)dt + c dW_t.$$

- Stein-Stein/Schobel-Zhu model: $p(x) = x$;
- Bergomi model $p(x) = \exp(x)$;
- Quintic OU model: $p(x) = p_0 + \alpha_1 x + \alpha_3 x^3 + \alpha_5 x^5$.

More generally, the model is well-defined with $\int_0^T \mathbb{E}[\sigma_s^2] ds < \infty$, or with p falling into a class of power series which contains all polynomial and exponential functions.

$$\frac{dS_t}{S_t} = \sigma_t dB_t, \quad S_0 > 0, \quad B = \rho W + \sqrt{1-\rho^2} W^{\perp},$$

$$\sigma_t = g_0(t) p(X_t) \quad p(x) = \sum_{k=0}^{\infty} \alpha_k x^k,$$

$$dX_t = (a + bX_t)dt + c dW_t.$$

- For $p(x) = x$, the volatility is affine in X (Stein-Stein/Schobel-Zhu model) and the model is affine (Duffie, Filipovic, and Schachermayer (2002)) in $(1, X, X^2)$ in the sense that the characteristic function is given by

$$\mathbb{E}\left[\exp\left(u \log\frac{S_T}{S_t} + v\int_t^T \sigma_s^2 ds\right) \Bigg| \mathcal{F}_t\right] = \exp\left(\psi_0(t) + \psi_1(t) X_t + \psi_2(t) X_t^2\right)$$

where (ψ_0, ψ_1, ψ_2) solve a system of Riccati ODEs $\Rightarrow$ Fast pricing by Fourier inversion methods.

$$\frac{dS_t}{S_t} = \sigma_t dB_t, \quad S_0 > 0, \quad B = \rho W + \sqrt{1-\rho^2} W^{\perp},$$

$$\sigma_t = g_0(t) p(X_t) \quad p(x) = \sum_{k=0}^{\infty} \alpha_k x^k,$$

$$dX_t = (a + bX_t)dt + cdW_t.$$

- For general p, one expects the model to be affine in $(1, X, X^2, \ldots, X^n, \ldots)$ so that an Ansatz of the characteristic function:

$$\mathbb{E}\left[\exp\left(u \log \frac{S_T}{S_t} + v \int_t^T \sigma_s^2 ds\right) \middle| \mathcal{F}_t\right] = \exp\left(\sum_{k \geq 0} \psi_k(t) X_t^k\right)$$

This is in accordance with recent expansions of characteristic function in (Cuchiero, Svaluto-Ferro, and Teichmann, 2023; Friz, Gatheral, and Radoičić, 2022)

$$
\begin{aligned}
\psi_k'(t) &= \left(g(t) + \frac{f(t)}{2}(f(t)-1)\right) g_0^2(T-t)(\alpha * \alpha)_k \\
&\quad + bk\psi_k(t) + a(k+1)\psi_{k+1}(t) + \frac{c^2(k+2)(k+1)}{2}\psi_{k+2}(t) \\
&\quad + \frac{c^2}{2}(\widetilde{\psi}(t) * \widetilde{\psi}(t))_k + \rho f(t) g_0(T-t) c(\alpha * \widetilde{\psi}(t))_k.
\end{aligned}
$$

with $\widetilde{\psi}_k = (k+1)\psi_{k+1}$ and $(u * v)_k = \sum_{i=0}^k u_i v_{k-i}$.

Fourier pricing in polynomial OU model

Riccati ODE

$$\psi_k'(t) = \left(g(t) + \frac{f(t)}{2}(f(t)-1)\right) g_0^2(T-t)(\alpha * \alpha)_k$$
$$+ bk\psi_k(t) + a(k+1)\psi_{k+1}(t) + \frac{c^2(k+2)(k+1)}{2}\psi_{k+2}(t)$$
$$+ \frac{c^2}{2}(\widetilde{\psi}(t) * \widetilde{\psi}(t))_k + \rho f(t) g_0(T-t) c(\alpha * \widetilde{\psi}(t))_k.$$

with $\widetilde{\psi}_k = (k+1)\psi_{k+1}$ and $(u * v)_k = \sum_{i=0}^k u_i v_{k-i}$.

Joint characteristic functional

Assume that there exists a solution $(\psi_k)_{\geq 0}$ to the infinite dimensional Riccati equation such that that $\sum_k |\psi_k(t)| x^k$ has infinite radius of convergence and $\mathfrak{Re}(\sum_k \psi_k(t) x^k) \leq 0$. Then,

$$\mathbb{E}\left[\exp\left(\int_t^T f(T-s) d\log S_s + \int_t^T g(T-s)\sigma_s^2 ds\right) \Bigg| X_t\right] = \exp\left(\sum_{k\geq 0} \psi_k(T-t) X_t^k\right).$$

$$\psi_k'(t) = \left(g(t) + \frac{f(t)}{2}(f(t)-1)\right)g_0^2(T-t)(\alpha * \alpha)_k$$
$$+ bk\psi_k(t) + a(k+1)\psi_{k+1}(t) + \frac{c^2(k+2)(k+1)}{2}\psi_{k+2}(t)$$
$$+ \frac{c^2}{2}(\widetilde{\psi}(t) * \widetilde{\psi}(t))_k + \rho f(t) g_0(T-t)c(\alpha * \widetilde{\psi}(t))_k.$$

with $\widetilde{\psi}_k = \psi_{k+1}(k+1)$ and $(u * v)_k = \sum_{i=0}^k u_i v_{k-i}$.

Fourier pricing in polynomial OU model

Riccati ODE

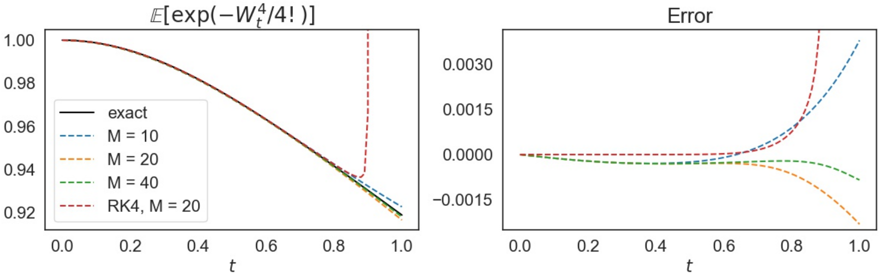
$$\psi_k'(t) = \Big(g(t) + \frac{f(t)}{2}(f(t)-1)\Big) g_0^2(T-t)(\alpha * \alpha)_k$$
$$+ bk\psi_k(t) + a(k+1)\psi_{k+1}(t) + \frac{c^2(k+2)(k+1)}{2}\psi_{k+2}(t)$$
$$+ \frac{c^2}{2}(\widetilde{\psi}(t) * \widetilde{\psi}(t))_k + \rho f(t) g_0(T-t) c(\alpha * \widetilde{\psi}(t))_k.$$

with $\widetilde{\psi}_k = \psi_{k+1}(k+1)$ and $(u * v)_k = \sum_{i=0}^{k} u_i v_{k-i}$.

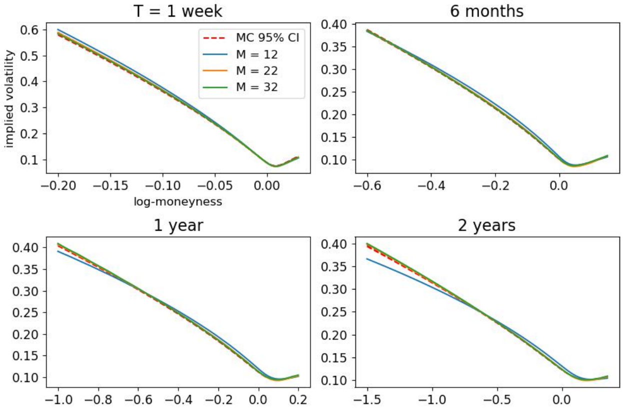
- **Theoretically** Not-standard infinite dimensional Riccati equations, no existence theory/result in the literature. Can be related to analyticity of solutions to PDE...
- **Numerically** Truncate system to some N but **stiff** system:
  - the calibrated coefficients (b, c) are large in general and the coefficients $(k+1)(k+2)$ become very large with the dimension k.
  - Standard Euler Schemes/ Explicit Runge-Kutta method are not enough.

Numerical solution of

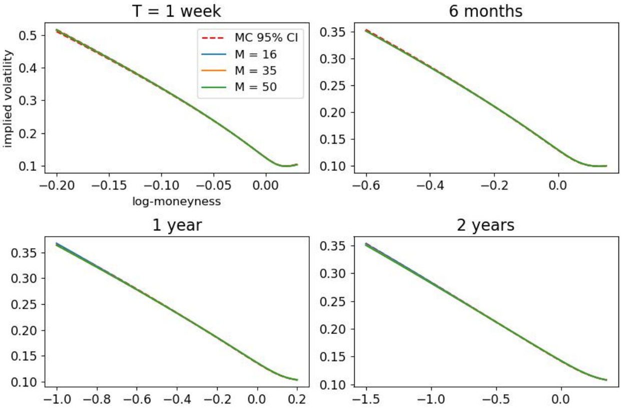
$$\mathbb{E}\left[\exp(-\frac{W_t^4}{4!})\right]$$



Fourier pricing for the Quintic OU model



Fourier pricing for the one factor Bergomi model



How to incorporate more memory in polynomials?

Signatures!

Signatures

We already considered the sequence of monomials

$$\left(1, W_t, \frac{1}{2}W_t^2, \frac{1}{3!}W_t^3, \ldots\right)$$

to linearize the pricing problem for the Quintic model. The sequence is called the **signature** of Brownian motion W and can be re-written as a sequence of iterated Stratonovich integrals

$$\left(1, \int_0^t dW_s, \int_0^t \int_0^{s_2} dW_{s_1} \circ dW_{s_2}, \int_0^t \int_0^{s_3} \int_0^{s_2} dW_{s_1} \circ dW_{s_2} \circ dW_{s_3}, \ldots\right).$$

Signatures

We already considered the sequence of monomials

$$\left(1, W_t, \frac{1}{2}W_t^2, \frac{1}{3!}W_t^3, \dots\right)$$

to linearize the pricing problem for the Quintic model. The sequence is called the **signature** of Brownian motion W and can be re-written as a sequence of iterated Stratonovich integrals

$$\left(1, \int_0^t dW_s, \int_0^t \int_0^{s_2} dW_{s_1} \circ dW_{s_2}, \int_0^t \int_0^{s_3} \int_0^{s_2} dW_{s_1} \circ dW_{s_2} \circ dW_{s_3}, \dots\right).$$

It can be extended to a **path-dependent setting**, in particular with the time augmented process $\widehat{W} : s \mapsto (s, W_s)$ to get

$$\widehat{\mathbb{W}}_t = \left(1, t, W_t, \frac{1}{2}t^2, \int_0^t s dW_s, \int_0^t W_s ds, \frac{1}{2}W_t^2, \dots, \dots\right)$$

The signature of a path is defined as the sequence of **iterated integrals** of the path. It can be seen as the analogue of **polynomials on path spaces**.

Universality of linear combinations of signatures on path-space:

- Universal approximation

$$f(t,(W_u)_{u\leq t}) \approx \langle \ell_t, \widehat{\mathbb{W}}_t \rangle$$

- **Universal representation?**

$$f(t,(W_u)_{u\leq t}) = \langle \ell_t, \widehat{\mathbb{W}}_t \rangle$$

Linear combinations can have either

- finitely many non-zero terms: finite polynomials on path space
- infinitely many non-zero terms: power series on path space

Signatures

1 will denote the first coordinate of $\widehat{W}$, i.e. time, and **2** the second, the Brownian motion.

$$\left(\varnothing, \begin{pmatrix} \mathbf{1} \\ \mathbf{2} \end{pmatrix}, \begin{pmatrix} \mathbf{11} & \mathbf{12} \\ \mathbf{21} & \mathbf{22} \end{pmatrix}, \begin{pmatrix} \mathbf{111} & \mathbf{121} & \\ \mathbf{211} & \mathbf{221} & \\ & \mathbf{112} & \mathbf{122} \\ & \mathbf{212} & \mathbf{222} \end{pmatrix}, \dots \right)$$

Example

- $\langle \mathbf{12}, \widehat{\mathbb{W}}_t \rangle = \int_0^t \int_0^s du dW_s,$
- $\langle \mathbf{21}, \widehat{\mathbb{W}}_t \rangle = \int_0^t \int_0^s dW_u ds,$
- $\langle \mathbf{212}, \widehat{\mathbb{W}}_t \rangle = \int_0^t \int_0^s \int_0^u dW_r \circ du \circ dW_s,$
- $\langle 2 \cdot \mathbf{212} - 3 \cdot \mathbf{12}, \widehat{\mathbb{W}}_t \rangle = \int_0^t \left(2 \int_0^s W_u du - 3s\right) \circ dW_s,$
- etc.

If (e_1, e_2) basis of $\mathbb{R}^2$, then for $(i_1, \ldots, i_2) \in \{1, 2\}^n$, we write $e_{i_1} \otimes e_{i_2} \otimes \ldots e_{i_n}$ as

$$\mathbf{i_1} \cdots \mathbf{i_n}$$

$$V_n := \{\mathbf{i_1} \cdots \mathbf{i_n} : \mathbf{i_k} \in \{1, 2\} \text{ for } k = 1, 2, \ldots, n\}.$$

Denote by $\varnothing$ the empty word, and $V_0 := \{\varnothing\}$ basis for $(\mathbb{R}^2)^{\otimes 0} = \mathbb{R}$.

If (e_1, e_2) basis of $\mathbb{R}^2$, then for $(i_1, \ldots, i_2) \in \{1,2\}^n$, we write $e_{i_1} \otimes e_{i_2} \otimes \ldots e_{i_n}$ as

$$\mathbf{i_1} \cdots \mathbf{i_n}$$

$$V_n := \{\mathbf{i_1} \cdots \mathbf{i_n} : \mathbf{i_k} \in \{1,2\} \text{ for } k = 1, 2, \ldots, n\}.$$

Denote by $\varnothing$ the empty word, and $V_0 := \{\varnothing\}$ basis for $(\mathbb{R}^2)^{\otimes 0} = \mathbb{R}$. It follows that $V := \cup_{n \geq 0} V_n$ represents the standard basis of $T((\mathbb{R}^2))$. In particular, every $\ell \in T((\mathbb{R}^2))$, can be decomposed as

$$\ell = \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} \ell^{\mathbf{v}} \mathbf{v},$$

where $\ell^{\mathbf{v}}$ is the real coefficient of ℓ at coordinate $\mathbf{v}$.

If (e_1, e_2) basis of $\mathbb{R}^2$, then for $(i_1, \ldots, i_2) \in \{1,2\}^n$, we write $e_{i_1} \otimes e_{i_2} \otimes \ldots e_{i_n}$ as

$$\mathbf{i_1} \cdots \mathbf{i_n}$$

$$V_n := \{\mathbf{i_1} \cdots \mathbf{i_n} : \mathbf{i_k} \in \{1,2\} \text{ for } k = 1, 2, \ldots, n\}.$$

Denote by $\varnothing$ the empty word, and $V_0 := \{\varnothing\}$ basis for $(\mathbb{R}^2)^{\otimes 0} = \mathbb{R}$. It follows that $V := \cup_{n \geq 0} V_n$ represents the standard basis of $T((\mathbb{R}^2))$. In particular, every $\ell \in T((\mathbb{R}^2))$, can be decomposed as

$$\ell = \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} \ell^{\mathbf{v}} \mathbf{v},$$

where $\ell^{\mathbf{v}}$ is the real coefficient of ℓ at coordinate $\mathbf{v}$.

$$\langle \ell, \widehat{\mathbb{W}}_t \rangle := \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} \ell^{\mathbf{v}} \widehat{\mathbb{W}}_t^{\mathbf{v}}.$$

for admissible elements ℓ in

$$\mathcal{A} := \left\{ \ell \in T((\mathbb{R}^2)) : \|\ell\|_t^{\mathcal{A}} < \infty \text{ for all } t \in [0, T] \text{ a.s.} \right\}, \quad \|\ell\|_t^{\mathcal{A}} := \sum_{n=0}^{\infty} \left| \sum_{\mathbf{v} \in V_n} \ell^{\mathbf{v}} \widehat{\mathbb{W}}_t^{\mathbf{v}} \right|, \quad t \geq 0,$$

Object	Interpretation	1-dim	2-dim
X	Path	W_t	$\widehat{W}_t = (t, W_t)$
$\mathbb{X}_t$	Signature	$(1, W_t, \frac{W_t^2}{2!}, \cdots)$	$\widehat{\mathbb{W}}_t := \left(1, \begin{pmatrix} t \\ W_t \end{pmatrix}, \cdots\right)$
$\langle \ell, \mathbb{X}_t \rangle$	Linear combination (possibly infinite)	$\sum_n \ell_n \frac{W_t^n}{n!}$ (power series)	$\sum_n \sum_{\mathbf{w} \in V^n} \ell^{\mathbf{w}} \widehat{\mathbb{W}}_t^{\mathbf{w}}$
$\langle \ell, \mathbb{X}_t \rangle \langle h, \mathbb{X}_t \rangle$	Product is linearized	$\sum_n (\hat{\ell} * \hat{h})_n \frac{W_t^n}{n!}$ (Cauchy)	$\langle \ell \sqcup\!\sqcup h, \widehat{\mathbb{W}}_t \rangle$ (Shuffle)
$F \approx f$	Approximation power	$F(W_t) \approx \sum_n \hat{f}_n W_t^n$ (Stone-Weierstrass Theorem)	$F(t, (W_s)_{s \leq t}) \approx \langle f, \widehat{\mathbb{W}}_t \rangle$ (Stone-Weierstrass Theorem)

Definition (Shuffle product)

For words $\mathbf{v}, \mathbf{w} \in V$ and letters $\mathbf{i}, \mathbf{j}$:
$(\mathbf{vi}) \sqcup\!\sqcup (\mathbf{wj}) = (\mathbf{v} \sqcup\!\sqcup (\mathbf{wj}))\mathbf{i} + ((\mathbf{vi}) \sqcup\!\sqcup \mathbf{w})\mathbf{j}, \qquad \mathbf{w} \sqcup\!\sqcup \varnothing = \varnothing \sqcup\!\sqcup \mathbf{w} = \mathbf{w}.$
Extended by linearity to the elements $\ell = \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} \ell^{\mathbf{v}} \mathbf{v}$.

Shuffle property

If $\ell_1, \ell_2 \in \mathcal{A}$, then $\ell_1 \sqcup\!\sqcup \ell_2 \in \mathcal{A}$ and

$$\langle \ell_1, \widehat{\mathbb{W}}_t \rangle \langle \ell_2, \widehat{\mathbb{W}}_t \rangle = \langle \ell_1 \sqcup\!\sqcup \ell_2, \widehat{\mathbb{W}}_t \rangle.$$

Summarize then Linearize

Back to our representation question

$$X_t = X_0 + \int_0^t a(t,s,(X_u)_{u\le s})ds + \int_0^t b(t,s,(X_u)_{u\le s})dW_s$$
$$= \langle \ell_t, \widehat{\mathbb{W}}_t \rangle ?$$

Based on

- *Path-dependent processes from signatures* with **Louis-Amand Gérard** and **Yuxing Huang**, 2024.

(!) Not a new question

- at least in the Markovian setting in 1980s: Ben Arous (1989) Doss, Fliess, Kunita, Sussmann, Yamato ...
- intimately related related to convergence questions of stochastic Taylor expansion's (Azencott, Bismut, Baudoin, Malliavin, Platen, ...).

Ornstein-Uhlenbeck (OU) process X:

$$X_t = x + \int_0^t \kappa(\theta - X_t)dt + \eta \int_0^t dW_s,$$

1. Write SDE in Stratonovich form, which here is the same.

Ornstein-Uhlenbeck (OU) process X:

$$X_t = x + \int_0^t \kappa(\theta - X_t)dt + \eta \int_0^t dW_s,$$

1. Write SDE in Stratonovich form, which here is the same.
2. Algebraic equation for candidate ℓ such as $X_t = \langle \ell, \widehat{\mathbb{W}}_t \rangle$

$$\ell = x\varnothing + (\kappa\theta - \kappa\ell)\mathbf{1} + \eta\mathbf{2} = p + \ell q$$

with

$$p = (x\varnothing + \kappa\theta\mathbf{1} + \eta\mathbf{2}), \quad q = -\kappa\mathbf{1}.$$

Ornstein-Uhlenbeck (OU) process X:

$$X_t = x + \int_0^t \kappa(\theta - X_t)dt + \eta \int_0^t dW_s,$$

1. Write SDE in Stratonovich form, which here is the same.
2. Algebraic equation for candidate ℓ such as $X_t = \langle \ell, \widehat{\mathbb{W}}_t \rangle$

$$\ell = x\varnothing + (\kappa\theta - \kappa\ell)\mathbf{1} + \eta\mathbf{2} = p + \ell q$$

with

$$p = (x\varnothing + \kappa\theta\mathbf{1} + \eta\mathbf{2}), \quad q = -\kappa\mathbf{1}.$$

3. Solve the linear algebraic equation

$$\ell = p(\varnothing - q)^{-1} = p \sum_k q^{\otimes k}$$

Ornstein-Uhlenbeck (OU) process X:

$$X_t = x + \int_0^t \kappa(\theta - X_t)dt + \eta \int_0^t dW_s,$$

1. Write SDE in Stratonovich form, which here is the same.
2. Algebraic equation for candidate ℓ such as $X_t = \langle \ell, \widehat{\mathbb{W}}_t \rangle$

$$\ell = x\varnothing + (\kappa\theta - \kappa\ell)\mathbf{1} + \eta\mathbf{2} = p + \ell q$$

   with

$$p = (x\varnothing + \kappa\theta\mathbf{1} + \eta\mathbf{2}), \quad q = -\kappa\mathbf{1}.$$

3. Solve the linear algebraic equation

$$\ell = p(\varnothing - q)^{-1} = p\sum_k q^{\otimes k}$$

4. Prove that $\ell \in \mathcal{A}$ so that $\langle \ell, \widehat{\mathbb{W}}_t \rangle$ is well defined, apply Itô and prove that it solves the SDE ...

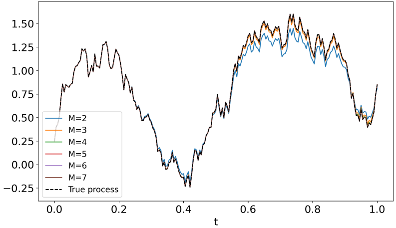
The Ornstein-Uhlenbeck (OU) process X:

$$dX_t = \kappa(\theta - X_t)dt + \eta dW_t, \quad X_0 = x \in \mathbb{R},$$

$$X_t = \langle \ell^{\mathrm{OU}}, \widehat{\mathbb{W}}_t \rangle, \quad \ell^{\mathrm{OU}} = (x\varnothing + \kappa\theta\mathbf{1} + \eta\mathbf{2})e^{\sqcup\!\sqcup -\kappa\mathbf{1}} \in \mathcal{A},$$

where

$$e^{\sqcup\!\sqcup \ell} := \varnothing + \ell + \frac{1}{2}\ell^{\sqcup\!\sqcup 2} + \cdots + \frac{1}{n!}\ell^{\sqcup\!\sqcup n} + \cdots$$



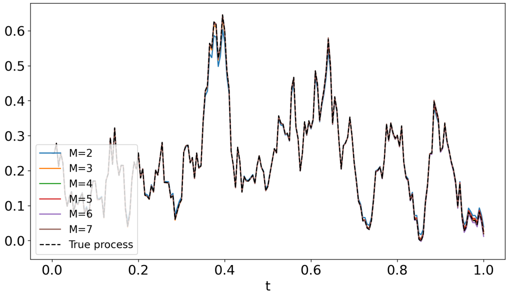
To be more explicit, up to order 3:

$$\ell^{\mathrm{OU}} = \left(x, \begin{pmatrix} -\kappa(x-\theta) \\ \eta \end{pmatrix}, \begin{pmatrix} \kappa^2(x-\theta) & 0 \\ -\kappa\eta & 0 \end{pmatrix}, \begin{pmatrix} -\kappa^3(x-\theta) & 0 \\ 0 & 0 \\ & \kappa^2\eta & 0 \\ & 0 & 0 \end{pmatrix}, \ldots \right).$$

The mean-reverting geometric Brownian motion

$$dY_t = \kappa(\theta - Y_t)dt + (\eta + \alpha Y_t)dW_t, \quad Y_0 = y \in \mathbb{R},$$

$$Y_t = \langle \ell^{\mathrm{mGBM}}, \widehat{\mathbb{W}}_t \rangle, \quad \ell^{\mathrm{mGBM}} = \left(y\varnothing + \left(\kappa\theta - \frac{\alpha\eta}{2} \right) \mathbf{1} + \eta \mathbf{2} \right) e^{\sqcup\!\sqcup \left(-\left(\kappa + \frac{\alpha^2}{2}\right)\mathbf{1} + \alpha \mathbf{2} \right)} \in \mathcal{A},$$



Certain linear Volterra equations:

$$X_t = X_0 + \int_0^t K_1(t-s)(a_0 + a_1 X_s)ds + \int_0^t K_2(t-s)(b_0 + b_1 X_s)dW_s.$$

Certain linear Volterra equations:

$$X_t = X_0 + \int_0^t K_1(t-s)(a_0 + a_1 X_s)ds + \int_0^t K_2(t-s)(b_0 + b_1 X_s)dW_s.$$

we will need the following structure on the kernels K_1 and K_2:

$$K_1(u) = \int_{[0,\infty)} e^{-xu}\mu_1(dx) \quad \text{and} \quad K_2(u) = \int_{[0,\infty)} e^{-xu}\mu_2(dx),$$

for finite measures μ_1 and μ_2 and such that

$$\int_{[0,\infty)} x^n \mu_1(dx) + \int_{[0,\infty)} x^n \mu_2(dx) < M^n, \quad n \in \mathbb{N},$$

for some constant $M > 0$. **(!)** Singular kernels such as the fractional kernel $t^{H-1/2}$ with $H < 1/2$ are excluded.

Certain linear Volterra equations:

$$Y_t = Y_0 + \int_0^t K_1(t-s)(a_0 + a_1 X_s)ds + \int_0^t K_2(t-s)(b_0 + b_1 X_s)dW_s.$$

Signature representation

$$Y_t = \langle \ell^{\mathrm{Vol}}, \widehat{\mathbb{W}}_t \rangle, \quad \ell^{\mathrm{Vol}} = p^{\mathrm{Vol}}(\varnothing - q)^{-1}$$

$$p := a_0 \mathbf{1} \int_{[0,\infty)} e^{\sqcup\!\sqcup - x\mathbf{1}} \mu_1(dx) + b_0(\mathbf{2} - \frac{1}{2} b_1 K_2(0)\mathbf{1}) \int_{[0,\infty)} e^{\sqcup\!\sqcup - x\mathbf{1}} \mu_2(dx) + Y_0 \varnothing,$$

$$q := a_1 \mathbf{1} \int_{[0,\infty)} e^{\sqcup\!\sqcup - x\mathbf{1}} \mu_1(dx) + b_1(\mathbf{2} - \frac{1}{2} b_1 K_2(0)\mathbf{1}) \int_{[0,\infty)} e^{\sqcup\!\sqcup - x\mathbf{1}} \mu_2(dx),$$

For more general kernels, possibly singular:

Approximation result

Let K_1, K_2 be locally square-integrable kernels. For $n \in \mathbb{N}$, let $\mu_1^n(dx) = \sum_{i=1}^n c_i \delta_{x_i}(dx)$ and $\mu_2^n(dx) = \sum_{i=1}^n d_i \delta_{y_i}(dx)$, for some $c_i, d_i, x_i, y_i \in \mathbb{R}$, and let K_1^n, K_2^n be the corresponding kernels and $Y_t^n = \langle \ell_n^{\text{Vol}}, \widehat{\mathbb{W}}_t \rangle$ as in previous theorem both with μ_1^n and μ_2^n instead of μ_1 and μ_2. Assume that

$$\int_0^T |K_1^n(s) - K_1(s)|^2 ds + \int_0^T |K_2^n(s) - K_2(s)|^2 ds \to 0, \quad \text{as } n \to \infty.$$

Then,

$$\sup_{t \in [0,T]} \mathbb{E}\left[\left| Y_t - \langle \ell_n^{\text{Vol}}, \widehat{\mathbb{W}}_t \rangle \right|^p \right] \to 0, \quad \text{as } n \to \infty, \quad p \in \mathbb{N}.$$

Gaussian Volterra processes including non-semimartingale processes such as the Riemann-Liouville fractional Brownian motion

$$W_t^H = \int_0^t (t-s)^{H-1/2} dW_s, \quad H \in (0,1).$$

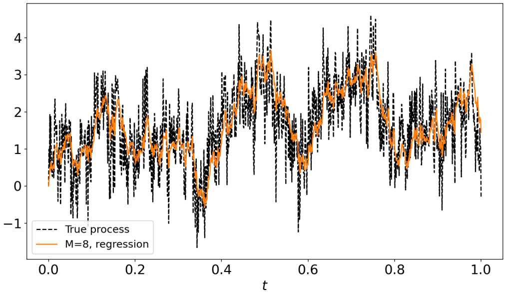
For instance, the (time-dependent) representation of W^H reads

$$W_t^H = \langle \ell_t^{\mathrm{RL}}, \widehat{\mathbb{W}}_t \rangle, \quad \ell_t^{\mathrm{RL}} = 1_{\{t>0\}} t^{H-\frac{1}{2}} \sum_{n=0}^{\infty} \frac{\left(\frac{1}{2}-H\right)^{\bar{n}}}{t^n} \mathbf{1}^{\otimes n}\mathbf{2},$$

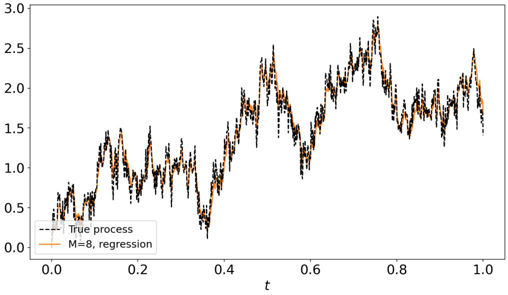
where $(\cdot)^{\bar{n}}$ is the rising factorial. This shows that $\langle \ell, \widehat{\mathbb{W}}_t \rangle$ is not always a semimartingale.

The Riemann-Liouville fractional Brownian motion

$$W_t^H = \int_0^t (t-s)^{H-1/2} dW_s, \quad H \in (0,1).$$



(a) $H = 0.1$

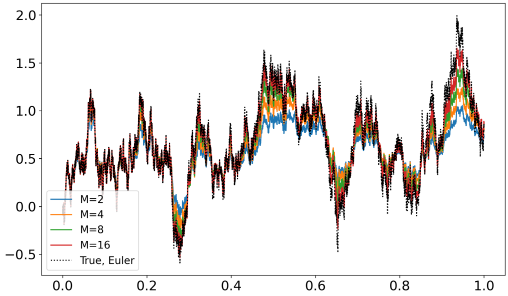


(b) $H = 0.3$

The shifted Riemann-Liouville fractional Brownian motion

$$W_t^{H,\eta} = \int_0^t (\eta + t - s)^{H-1/2} dW_s, \quad H \in (0,1).$$

$$\eta = \frac{1}{52}, \quad H = 0.1.$$



Towards Signature volatility models

$$\frac{dS_t}{S_t} = \Sigma_t dB_t, \quad \Sigma_t = \langle \sigma_t, \widehat{\mathbb{W}}_t \rangle, \quad B = \rho W + \sqrt{1-\rho^2} W^{\perp}.$$

Based on

- *Signature volatility models: pricing and hedging with Fourier* with **Louis-Amand Gérard**, SIAM Journal on Financial Mathematics, to appear (2025)

Refs

- Perez Arribas, Salvi, and Szpruch (2020)
- Cuchiero, Gazzani, Möller, and Svaluto-Ferro (2025).

Theorem

Let $f, g : [0, T] \to \mathbb{C}$ be measurable and bounded functions. Assume that there exists ψ solution to the following system of time-dependent Riccati equations

$$\dot{\psi}_t = \frac{1}{2}(\psi_t|_{\mathbf{2}})^{⧢2} + \rho f(t)(\sigma ⧢ \psi_t|_{\mathbf{2}}) + \frac{1}{2}\psi_t|_{\mathbf{22}} + \psi_t|_{\mathbf{1}} + \left(\frac{f(t)^2 - f(t)}{2} + g(t)\right)\sigma^{⧢2}, \quad \psi_0 = 0,$$

such that $\psi_t \in \mathcal{I}$ and $\mathfrak{Re}(\langle \psi_{T-t}, \widehat{\mathbb{W}}_t \rangle) \le 0$, then, the joint characteristic functional is given by

$$\mathbb{E}\left[\exp\left(\int_t^T f(T-s)d\log S_s + \int_t^T g(T-s)\sigma_s^2 ds\right)\middle|\mathcal{F}_t\right] = \exp\left(\langle \psi_{T-t}, \widehat{\mathbb{W}}_t \rangle\right).$$

- Similar representations for signature SDEs (Cuchiero, Svaluto-Ferro, and Teichmann, 2023).
- Related representations Friz, Gatheral, and Radoičić (2022); Lyons, Ni, and Tao (2024)

Volterra Bergomi model

$$dS_t = S_t \sigma_t dB_t$$

$$\sigma_t = \sigma_0 e^{\eta X_t - \frac{\eta^2}{2} \mathrm{Var}(X_t)}, \quad X_t = \int_0^t K(t-s) dW_s$$

Volterra Bergomi model

$$dS_t = S_t \sigma_t dB_t$$

$$\sigma_t = \sigma_0 e^{\eta X_t - \frac{\eta^2}{2}\mathrm{Var}(X_t)}, \quad X_t = \int_0^t K(t-s)dW_s$$

- n-factor Bergomi model (under-parametrized version with same BM)

$$K(t) = \sum_{i=1}^{n} c_i e^{-\lambda_i t}.$$

- Rough Bergomi of Bayer, Friz, and Gatheral (2016)

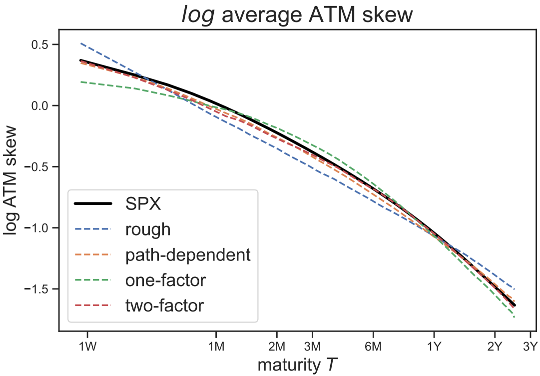
$$K(t) = ct^{H-1/2}, \quad H \in (0, 1/2)$$

- Shifted fractional kernel

$$K(t) = c(\eta + t)^{H-1/2}, \quad \eta > 0, \quad H \in \mathbb{R}.$$

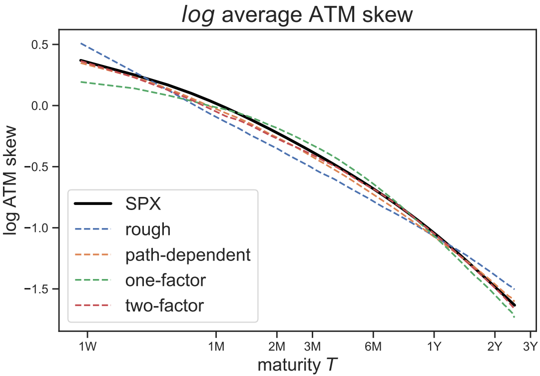
Log-plot SPX ATM skew is **concave**, **flattening behavior** at short maturities

$$T \to \partial_k \sigma_{iv}(T, k) \,|_{k=0}$$

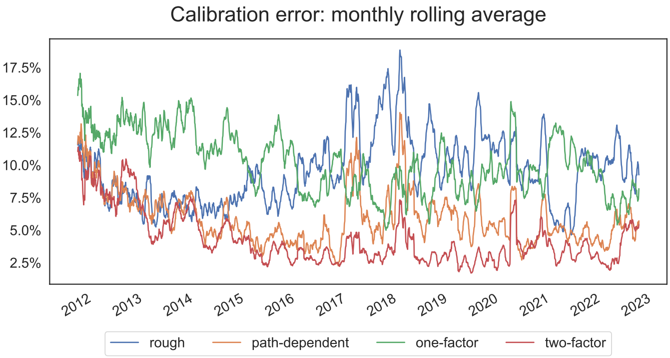


Log-plot SPX ATM skew is **concave**, **flattening behavior** at short maturities

$$T \to \partial_k \sigma_{iv}(T,k)\,|_{k=0}$$



- **Exponential kernel** $K(t) = ce^{-\lambda t}$ gets shape but lacks flexibility
- Two time scales captured via **double-exponential kernel** $K(t) = c_1 e^{-\lambda_1 t} + c_2 e^{-\lambda_2 t}$
- **Fractional kernel** $K(t) = ct^{H-1/2}$ ($H \in (0, 1/2)$) implies monofractal scaling (straight line) and blow up at 0, inconsistent with data
- **Shifted fractional kernel** $K(t) = c(a+t)^{H-1/2}$ with $a > 0$ breaks monofractality/roughness and decouples short and long term behaviour
- AJ and Li (2025); Bergomi (2015); Delemotte et al. (2023); Guyon and El Amrani (2022)



- Daily calibration of SPX vol surface (maturities couple of days to 3 years) on more than 10 years
- Rough Bergomi model does not align with market data.
- Non-rough path-dependent Bergomi models aligns much better in **all market conditions**
- Deep pricing with quantization hints method

AJ & Li (2025). Volatility Models in Practice: Rough, Path-Dependent, or Markovian? Mathematical Finance

Shifted fractional Bergomi model

$$dS_t = S_t \sigma_t dB_t$$

$$\sigma_t = \sigma_0 e^{\eta W_t^{H,a} - \frac{\eta^2}{2} \mathrm{Var}(W_t^{H,a})}$$

with

$$W_t^{H,a} = \int_0^t (a + t - s)^{H-1/2} dW_s, \quad a > 0, \quad H \in \mathbb{R}.$$

Shifted fractional Bergomi model

$$dS_t = S_t \sigma_t dB_t$$
$$\sigma_t = \sigma_0 e^{\eta W_t^{H,a} - \frac{\eta^2}{2}\mathrm{Var}(W_t^{H,a})}$$

with

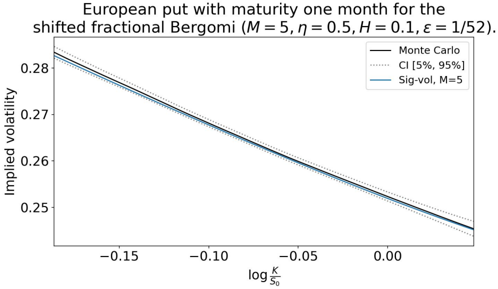
$$W_t^{H,a} = \int_0^t (a + t - s)^{H-1/2} dW_s, \quad a > 0, \quad H \in \mathbb{R}.$$

Recast into **signature volatility model** $\sigma_t = \langle \ell_t, \widehat{\mathbb{W}}_t \rangle$:

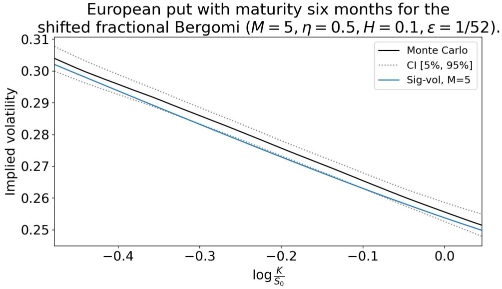
- $W_t^{H,a} = \langle \ell_t^{H,a}, \widehat{\mathbb{W}}_t \rangle$ with $\ell_t^{H,a} = 1_{\{t>0\}}(t+a)^{H-\frac{1}{2}} \sum_{n=0}^{\infty} \frac{\left(\frac{1}{2}-H\right)^{\bar{n}}}{(t+a)^n} \mathbf{1}^{\otimes n}\mathbf{2}$,
- using the **Shuffle product**

$$e^{\eta W_t^H} = \sum_{n\geq 0} \frac{\eta^n \langle \ell_t^{H,a}, \widehat{\mathbb{W}}_t \rangle^n}{n!} = \sum_{n\geq 0} \frac{\eta^n \langle (\ell_t^{H,a})^{\sqcup\!\sqcup n}, \widehat{\mathbb{W}}_t \rangle}{n!} = \langle e^{\sqcup\!\sqcup \eta \ell_t^{H,a}}, \widehat{\mathbb{W}}_t \rangle$$

Shifted fractional Bergomi model by **Fourier pricing** in signature volatility model:

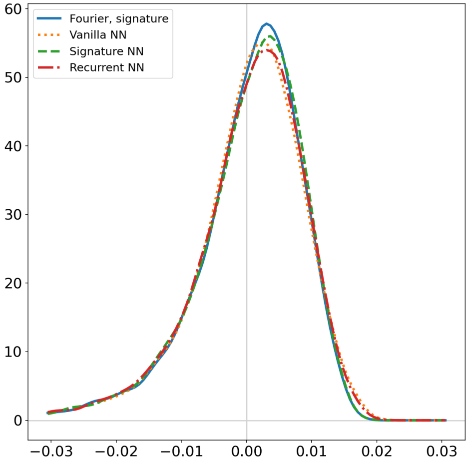


(c) 1 month



(d) 6 months

Shifted fractional Bergomi model by **Fourier quadratic hedging** in signature volatility model:



How to truncate?

$$dS_t^N = S_t^N \langle \sigma^N, \widehat{\mathbb{W}}_t \rangle dB_t, \quad B = \rho W + \sqrt{1-\rho^2} W^{\perp}.$$

Based on

- *Martingale property and moment explosions in signature volatility models* with **Paul Gassiat** and **Dimitri Sotnikov**

How to truncate?

$$dS_t^N = S_t^N \langle \sigma^N, \widehat{\mathbb{W}}_t \rangle dB_t, \quad B = \rho W + \sqrt{1-\rho^2} W^{\perp}.$$

Assume non-zero leading coefficient $\sigma^{\mathbf{2}^{\otimes N}} \neq 0$ in front of the word $\mathbf{2}^{\otimes N}$, (i.e. the term W_t^N.) and non-zero correlation ρ and $N \geq 2$:

How to truncate?

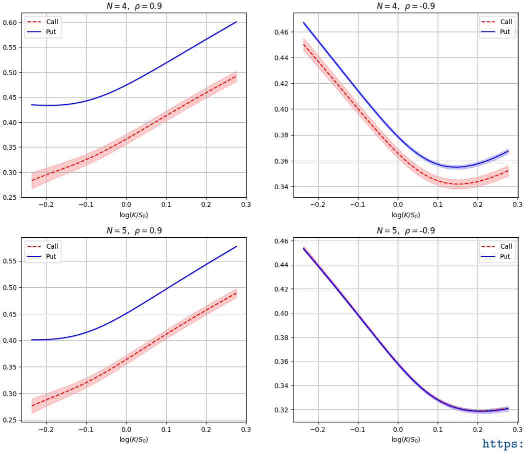
$$dS_t^N = S_t^N \langle \sigma^N, \widehat{\mathbb{W}}_t \rangle dB_t, \quad B = \rho W + \sqrt{1-\rho^2} W^{\perp}.$$

Assume non-zero leading coefficient $\sigma^{\mathbf{2}^{\otimes N}} \neq 0$ in front of the word $\mathbf{2}^{\otimes N}$, (i.e. the term W_t^N.) and non-zero correlation ρ and $N \geq 2$:

The price process S^N is a **true martingale** if and only if N is odd and $\rho\sigma^{\mathbf{2}^{\otimes N}} \leq 0$.

How to truncate?

The price process S^N is a **true martingale** if and only if N is odd and $\rho\sigma^{\mathbf{2}^{\otimes N}} \leq 0$.



Stationarity?

Two important properties when modeling **memory effects** of dynamical systems:

- **Time invariance** the output signal Y_t at time t depends only on the continuous input signal $(X_s)_{-\infty<s\leq t}$ up to time t, but not on the absolute time t.

$$Y_t = F((X_{t-s})_{s\geq 0}).$$

$\Rightarrow$ Postulates, in some sense, a **stationarity** in the relationship between input and output.

Two important properties when modeling **memory effects** of dynamical systems:

- **Time invariance** the output signal Y_t at time t depends only on the continuous input signal $(X_s)_{-\infty<s\leq t}$ up to time t, but not on the absolute time t.

  $$Y_t = F((X_{t-s})_{s\geq 0}).$$

  $\Rightarrow$ Postulates, in some sense, a **stationarity** in the relationship between input and output.
- **Fading memory** the influence of the distant past diminishes over time, i.e. Y_t gradually "forgets" the remote history of X.

Two important properties when modeling **memory effects** of dynamical systems:

- **Time invariance** the output signal Y_t at time t depends only on the continuous input signal $(X_s)_{-\infty<s\leq t}$ up to time t, but not on the absolute time t.

$$Y_t = F((X_{t-s})_{s\geq 0}).$$

$\Rightarrow$ Postulates, in some sense, a **stationarity** in the relationship between input and output.

- **Fading memory** the influence of the distant past diminishes over time, i.e. Y_t gradually "forgets" the remote history of X.
  - This idea dates back to the works of Volterra (1887) and Wiener (1958).

Two important properties when modeling **memory effects** of dynamical systems:

- **Time invariance** the output signal Y_t at time t depends only on the continuous input signal $(X_s)_{-\infty<s\leq t}$ up to time t, but not on the absolute time t.

$$Y_t = F((X_{t-s})_{s\geq 0}).$$

$\Rightarrow$ Postulates, in some sense, a **stationarity** in the relationship between input and output.

- **Fading memory** the influence of the distant past diminishes over time, i.e. Y_t gradually "forgets" the remote history of X.
  - This idea dates back to the works of Volterra (1887) and Wiener (1958).
  - Boyd and Chua (1985) formalize this concept by requiring the functional F to be continuous not with respect to the uniform topology, but with respect to a *weighted* uniform topology:

F has fading memory if its output remains close for input paths that are close in the recent past, even if they differ in the distant past.

- We are interested in modeling **time-invariant** dependence between two time series $(X_t, Y_t)_{t\in\mathbb{R}}$, i.e.

$$Y_t = F(X_{s\in(-\infty,t]}),$$

for some continuous function F.

- In practice, often only a **single realization** of a time series or dynamical system is observed (financial data), from the **infinitely distant past $-\infty$ up to t.**

Problem

Can we generalize the exponential moving average (EMA)

$$y_n = \sum_{k\geq 0} e^{-\lambda k} x_{n-k}$$

to a framework that:

- is **universal** and captures **nonlinear** features of the time series;
- Preserves the **Markov** property;
- Remains mathematically **tractable**?

The Exponentially Fading Memory Signature

Based on

- *Exponentially Fading Memory Signature* with **Dimitri Sotnikov**

Fading Memory Signature
Definition

- We consider an $\mathbb{R}^d$-valued increment continuous semimartingales X, i.e., for all $s \in \mathbb{R}$, the process $(X_{s+t} - X_s)_{t\geq 0}$ is a semimartingale in the usual sense with respect to $(\mathcal{F}_{s+t})_{t\geq 0}$.
- Vector $\boldsymbol{\lambda} = (\lambda^1, \ldots, \lambda^d) \in \mathbb{R}^d$ with positive entries.

- We consider an $\mathbb{R}^d$-valued increment continuous semimartingales X, i.e., for all $s \in \mathbb{R}$, the process $(X_{s+t} - X_s)_{t \geq 0}$ is a semimartingale in the usual sense with respect to $(\mathcal{F}_{s+t})_{t \geq 0}$.
- Vector $\boldsymbol{\lambda} = (\lambda^1, \ldots, \lambda^d) \in \mathbb{R}^d$ with positive entries.

Fading Memory Signature

The fading-memory $\boldsymbol{\lambda}$-signature of X is defined by the components

$$\mathbb{X}_t^{\boldsymbol{\lambda}, \mathbf{i_1 \cdots i_n}} := \int_{-\infty < u_1 < \cdots < u_n < t} e^{-\lambda^{i_1}(t-u_1)} dX_{u_1}^{i_1} \circ \cdots \circ e^{-\lambda^{i_n}(t-u_n)} dX_{u_n}^{i_n}, \quad \mathbf{i_k} \in \{\mathbf{1}, \ldots, \mathbf{d}\},$$

and on $[s, t]$

$$\mathbb{X}_{s,t}^{\boldsymbol{\lambda}, \mathbf{i_1 \cdots i_n}} := \int_{s < u_1 < \cdots < u_n < t} e^{-\lambda^{i_1}(t-u_1)} dX_{u_1}^{i_1} \circ \cdots \circ e^{-\lambda^{i_n}(t-u_n)} dX_{u_n}^{i_n}.$$

First two levels for the time-extended Brownian motion

The first elements of $\widehat{\mathbb{W}}_t^{\boldsymbol{\lambda}}$ are given by

$$\widehat{\mathbb{W}}_t^{\lambda,0} = 1, \quad \widehat{\mathbb{W}}_t^{\lambda,1} = \begin{pmatrix} \lambda^{-1} \\ Y_t \end{pmatrix}, \quad \widehat{\mathbb{W}}_t^{\lambda,2} = \begin{pmatrix} \frac{\lambda^{-2}}{2!} & \lambda^{-1} \int_{-\infty}^t e^{-2\lambda(t-s)} dW_s \\ \int_{-\infty}^t e^{-2\lambda(t-s)} Y_s ds & \frac{Y_t^2}{2!} \end{pmatrix},$$

where $Y = (Y_t)_{t\in\mathbb{R}}$ is a stationary Ornstein–Uhlenbeck process defined by

$$Y_t = \int_{-\infty}^t e^{-\lambda(t-s)} dW_s.$$

Moreover, $\widehat{\mathbb{W}}_t^{\lambda,\mathbf{2}^{\otimes n}} = \dfrac{Y_t^n}{n!}$.

First two levels for the time-extended Brownian motion

The first elements of $\widehat{\mathbb{W}}_t^{\boldsymbol{\lambda}}$ are given by

$$\widehat{\mathbb{W}}_t^{\lambda,0} = 1, \quad \widehat{\mathbb{W}}_t^{\lambda,1} = \begin{pmatrix} \lambda^{-1} \\ Y_t \end{pmatrix}, \quad \widehat{\mathbb{W}}_t^{\lambda,2} = \begin{pmatrix} \frac{\lambda^{-2}}{2!} & \lambda^{-1} \int_{-\infty}^t e^{-2\lambda(t-s)} dW_s \\ \int_{-\infty}^t e^{-2\lambda(t-s)} Y_s ds & \frac{Y_t^2}{2!} \end{pmatrix},$$

where $Y = (Y_t)_{t \in \mathbb{R}}$ is a stationary Ornstein–Uhlenbeck process defined by

$$Y_t = \int_{-\infty}^t e^{-\lambda(t-s)} dW_s.$$

Moreover, $\widehat{\mathbb{W}}_t^{\lambda,\mathbf{2}^{\otimes n}} = \dfrac{Y_t^n}{n!}$.

- Note that the first two levels are **stationary**!

First properties
Time Invariance and Stationarity

If we shift the inputs by h, we shift the output by h:

Time Invariance

If $X = (X_t)_{t\in\mathbb{R}}$ is a continuous semimartingale and $Y_\cdot = X_{\cdot+h}$ for some $h \in \mathbb{R}$, then

$$\mathbb{Y}^{\boldsymbol{\lambda}}_t = \mathbb{X}^{\boldsymbol{\lambda}}_{t+h}.$$

Stationarity

Suppose that $X = (X_t)$ is an $\mathbb{R}^d$-valued continuous semimartingale with **stationary increments**. Then, $\mathbb{X}^{\boldsymbol{\lambda}} = (\mathbb{X}^{\boldsymbol{\lambda}}_t)_{t\in\mathbb{R}}$ is a stationary $T((\mathbb{R}^d))$-valued process. In particular, for all $t \in \mathbb{R}$,

$$\mathcal{L}(\mathbb{X}^{\boldsymbol{\lambda}}_t) = \mathcal{L}(\mathbb{X}^{\boldsymbol{\lambda}}_0).$$

- In particular, this holds for $X_t = (t, W_t)$.

Applications to learning

Application 1: **Regression** and Prediction

We observe a signal $S_t = \sin(Z_t)$, where

$$dZ_t = -\mu Z_t dt + \nu dW_t,$$

Goal do a linear regression of S_t against

1. The standard time-augmented signature of $\widehat{W}_t = (t, W_t)$:

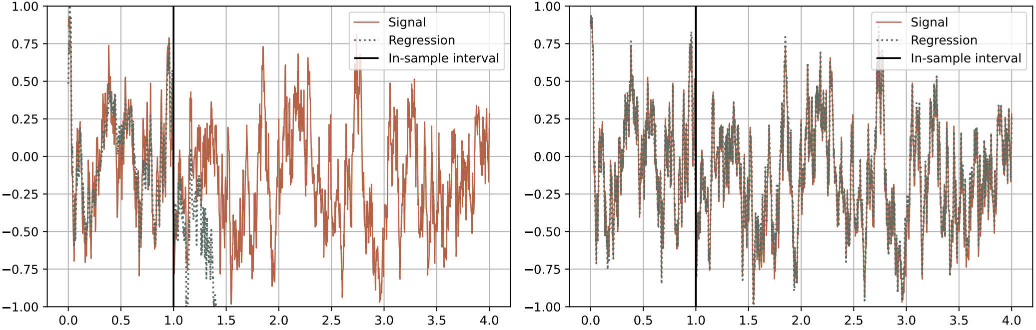
$$S_t \approx \langle \boldsymbol{\ell}, \widehat{\mathbb{W}}_t \rangle$$

2. The EFM-signature of $\widehat{W}_t = (t, W_t)$:

$$S_t = \left\langle \boldsymbol{\ell}, \widehat{\mathbb{W}}^{\lambda}_t \right\rangle$$

Estimate $\boldsymbol{\ell}$ observing W and the signal S.

Application: **Regression**



Signature (on the left) and Fading memory signature (on the right) regression. Signal parameters are $\mu = 25$, $\nu = 3$, truncation order is $N = 5$. Vertical bar separates in-sample and out-of-sample data.

Contact

- eduardo.abi-jaber@polytechnique.edu
- https://sites.google.com/view/abijabereduardo/

References I

E. AJ and S. Li. Volatility models in practice: Rough, path-dependent, or markovian? *Mathematical Finance*, 2025.

C. Bayer, P. Friz, and J. Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6): 887–904, 2016.

L. Bergomi. Smile dynamics IV. *Risk Magazine*, 2009.

L. Bergomi. *Stochastic volatility modeling*. CRC press, 2015.

F. Bourgey, J. Delemotte, and S. De Marco. Smile dynamics and rough volatility. *Available at SSRN*, 2024.

S. P. Boyd and L. O. Chua. Fading memory and the problem of approximating nonlinear operators with volterra series. 1985. URL https://api.semanticscholar.org/CorpusID:14862233.

C. Cuchiero, S. Svaluto-Ferro, and J. Teichmann. Signature sdes from an affine and polynomial perspective. *arXiv preprint arXiv:2302.01362*, 2023.

References II

C. Cuchiero, G. Gazzani, J. Möller, and S. Svaluto-Ferro. Joint calibration to spx and vix options with signature-based models. *Mathematical Finance*, 35(1):161–213, 2025.

J. Delemotte, S. D. Marco, and F. Segonne. Yet another analysis of the sp500 at-the-money skew: Crossover of different power-law behaviours. *Available at SSRN 4428407*, 2023.

D. Duffie, D. Filipovic, and W. Schachermayer. Affine processes and application in finance. Technical report, National Bureau of Economic Research, 2002.

P. K. Friz and J. Gatheral. Computing the ssr. *Quantitative Finance*, 25(5):701–710, 2025.

P. K. Friz, J. Gatheral, and R. Radoičić. Forests, cumulants, martingales. *The Annals of Probability*, 50(4):1418–1445, 2022.

J. Guyon and M. El Amrani. Does the term-structure of equity at-the-money skew really follow a power law? *Available at SSRN 4174538*, 2022.

T. Lyons, H. Ni, and J. Tao. A pde approach for solving the characteristic function of the generalised signature process. *arXiv preprint arXiv:2401.02393*, 2024.

References III

I. Perez Arribas, C. Salvi, and L. Szpruch. Sig-sdes model for quantitative finance. pages 1–8, 10 2020. doi: 10.1145/3383455.3422553.

V. Volterra. *Sopra le funzioni che dipendono da altre funzioni.* 1887.

N. Wiener. *Nonlinear Problems in Random Theory*. M.I.T. paperback series. MIT Press, 1958. ISBN 9780262730129. URL https://books.google.fr/books?id=HQBRAAAAMAAJ.

What about dynamical properties and longer maturities?

Ref

- *Capturing Smile Dynamics with the Quintic Volatility Model: SPX, Skew-Stickiness Ratio and VIX* with **Shaun Li** (2025).

The **Skew Stickiness Ratio** (SSR), $\mathcal{R}_{t,T}$ is defined in Bergomi (2009) as

$$\mathcal{R}_{t,T} := \frac{1}{\mathcal{S}_{t,T}} \frac{\partial_t \langle \widehat{\sigma}^T, \log S \rangle_t}{\partial_t \langle \log S \rangle_t},$$

- $\widehat{\sigma}_t^T$ denotes the ATM implied volatility
- and $\mathcal{S}_{t,T}$ is the ATM (forward) skew.

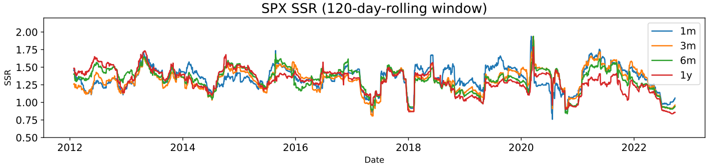
The SSR can be interpreted as the instantaneous change of the ATM implied volatility with respect to the instantaneous change of the log-price, normalised by the ATM skew.

The **Skew Stickiness Ratio** (SSR), $\mathcal{R}_{t,T}$ is defined in Bergomi (2009) as

$$\mathcal{R}_{t,T} := \frac{1}{\mathcal{S}_{t,T}} \frac{\partial_t \langle \widehat{\sigma}^T, \log S \rangle_t}{\partial_t \langle \log S \rangle_t},$$

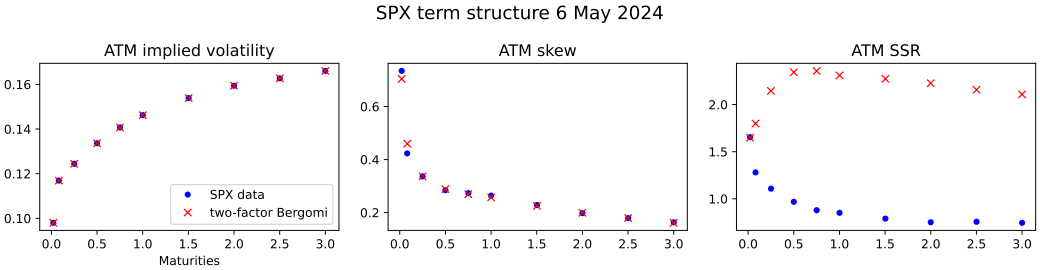
- $\widehat{\sigma}_t^T$ denotes the ATM implied volatility
- and $\mathcal{S}_{t,T}$ is the ATM (forward) skew.

The SSR can be interpreted as the instantaneous change of the ATM implied volatility with respect to the instantaneous change of the log-price, normalised by the ATM skew.



Empirical time series of the SSR from 2012 to 2022, computed using a 120-day rolling window.

Calibrating smiles and SSR is **notoriously difficult** for stochastic volatility models Bourgey, Delemotte, and De Marco (2024); Friz and Gatheral (2025). **Two factor Bergomi model:**



Calibrated two factor Bergomi model on term structures May 6, 2024.

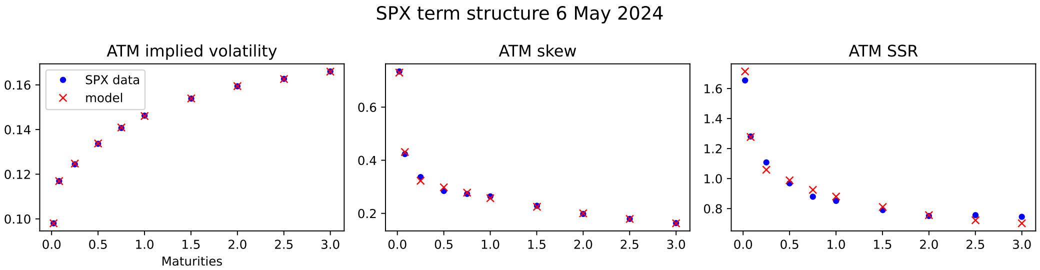
$$dS_t = S_t\sigma_t\left(\rho dW_t + \sqrt{1-\rho^2}dW_t^{\perp}\right), \quad S_0 > 0,$$

$$\sigma_t = g_0(t)p(Z_t), \quad p(z) = \sum_{k=0}^{5}\alpha_k z^k,$$

$$Z_t = \theta X_t + (1-\theta)Y_t,$$

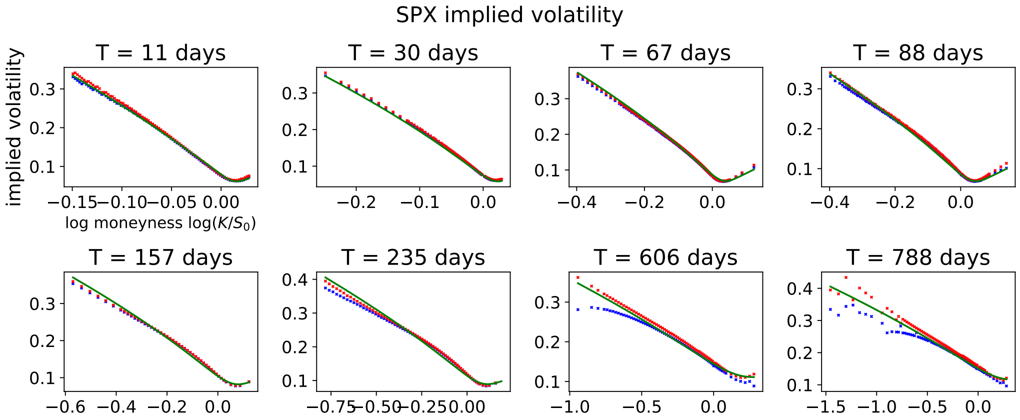
$$X_t = \int_0^t e^{-\lambda_x(t-s)}dW_s, \quad Y_t = \int_0^t e^{-\lambda_y(t-s)}dW_s,$$

The **two factor Quintic model** is able to achieve impressive fits of the term structures of ATM-vol, skew and SSR:



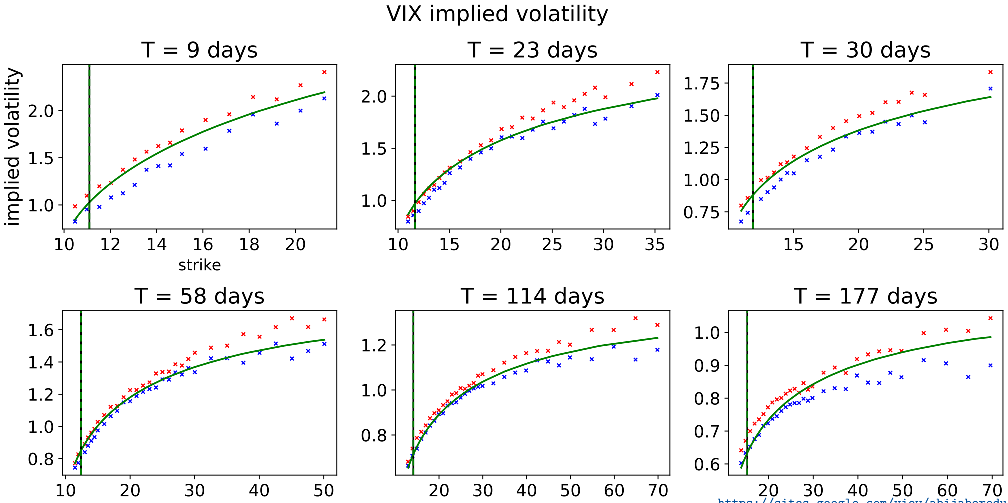
Calibrated two factor Quintic model on term structures May 6, 2024.

SPX and VIX:

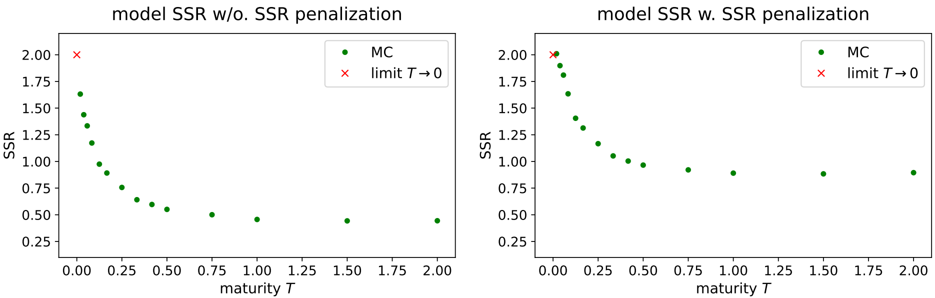


SPX & VIX smiles (bid/ask in blue/red dots) and VIX futures (vertical black lines) on 23 October 2017, jointly calibrated by the two-factor Quintic OU model (in green) with SSR penalisation.

SPX and VIX:


VIX implied volatility
T = 9 days
T = 23 days
T = 30 days
T = 58 days
T = 114 days
T = 177 days
implied volatility
strike

Penalisation for consistent values of SSR:



SSR of the two-factor Quintic OU model computed by finite difference and Monte Carlo. The left-hand side graph is the SSR of the two-factor Quintic OU model jointly calibrated to SPX and VIX smiles. The right-hand side graph is the SSR of the two-factor Quintic OU model jointly calibrated to SPX and VIX smiles, as well as the SSR.