UCLouvain

FACULTE DES SCIENCES

Faculté des sciences

Option pricing in a diffusive market with rough Hawkes jumps

Author: **Audren BALON**
Supervisor: **Donatien HAINAUT**
Reader: **Jérôme BARBARIN**
Academic year 2025–2026
Master [120] en sciences actuarielles

Abstract

Rough Hawkes processes have been extensively studied in recent years for modeling Poisson intensities. These processes introduce memory effects through self-exciting jumps, making them theoretically particularly suitable for capturing clustering phenomena observed in financial markets.

This thesis provides an overview of methods for handling non-Markovian yet affine processes, of which rough Hawkes models arise as a specific instance through the appropriate choice of rough kernel.

First, self-exciting jumps are incorporated into the classical diffusive dynamics of asset prices. We focus on unidimensional memory processes and compare them with the two-dimensional rough Hawkes process studied by Hainaut,[1].

Second, the empirical observation of volatility clustering motivates the introduction of self-exciting jumps into the Cox–Ingersoll–Ross (CIR) process governing the volatility in a Heston-type model. Market evidence suggests that jumps in volatility are associated with simultaneous jumps in the underlying asset, typically of opposite sign and smaller magnitude. We provide a mathematical analysis of this framework and derive expressions for the moment-generating function of the asset price in both settings.

Overall, the proposed modeling approach shows potential to reproduce the volatility smile after appropriate calibration.

Keywords Affine Volterra processes - Rough Hawkes Jumps - Heston model

Acknowledgments

I would like to thank Donatien Hainaut, whose guidance and discussions were consistently constructive, and whose teaching has been among the most valuable throughout my academic journey.

Beyond being a project, this thesis has been a rewarding opportunity to bring together the knowledge I have acquired throughout my studies, ranging from spectral methods and Galerkin approximations to financial modeling, in order to conduct a detailed analysis and explore non-Markovian processes.

Naturally, I would have greatly appreciated the opportunity to present a full calibration based on market-implied volatility. However, I hope that the depth and rigor of the analyses developed in this work are sufficiently meaningful to provide a solid foundation for future research to calibrate the rough Hawkes–Heston model introduced here and to assess its full practical potential.

Contents

1 Introduction to affine Volterra framework **11**
- 1.1 Step 1 : Spectral representation of the kernel . . . 12
- 1.2 Step 2 : Discretization of the infinite Markovian system . . . 20
- 1.3 Step 3 : Invariance of the Markovian generator of affine processes . . . 24
- 1.4 Step 4 : System of Ricatti-Volterra equations . . . 25

2 Rough Hawkes Process **27**
- 2.1 Procedural computation . . . 28
  - 2.1.1 Step 1 . . . 28
  - 2.1.2 Step 2 . . . 28
  - 2.1.3 Step 3 . . . 29
  - 2.1.4 Step 4 . . . 31
- 2.2 Calibration under real-world measure . . . 33
- 2.3 Change of measure . . . 35
- 2.4 DFFT algorithm . . . 39

3 Calibration and illustration: BTC returns **45**
- 3.1 Peak-over-threshold method . . . 45
- 3.2 Maximum log-likelihood estimation of parameters . . . 48
- 3.3 Comparison with other kernels . . . 50
  - 3.3.1 Classical Hawkes kernel (exponential kernel from Cauchy distribution) . . . 50
  - 3.3.2 Laplacian measure . . . 50
  - 3.3.3 Gaussian spectral measure . . . 51
  - 3.3.4 Logistic measure . . . 52
- 3.4 Comparison with two-dimensional rough Hawkes processes . . . 53
- 3.5 Call pricing . . . 57

4 Rough Hawkes-Heston models **61**
- 4.1 Rough Hawkes-Heston volatility model . . . 63
  - 4.1.1 Step 1 . . . 64
  - 4.1.2 Step 2 . . . 64
  - 4.1.3 Step 3 . . . 64
  - 4.1.4 Step 4 . . . 65
- 4.2 Mutual rough Hawkes jump in Heston volatility model . . . 68
  - 4.2.1 Step 1 . . . 68
  - 4.2.2 Step 2 . . . 69
  - 4.2.3 Step 3 . . . 69
  - 4.2.4 Step 4 . . . 69
- 4.3 Change of measure . . . 70

5 Illustrations : VIX **73**
5.1 Calibration under risk-neutral measure 73
5.2 Alternative calibration under $\mathbb{P}$ 75
5.3 Call pricing 77
5.4 Further considerations 82

6 Conclusion **83**

7 Appendix **85**
7.1 Dynamic of the volatility under CIR process 85

The choice of a model for pricing equities is a critical step in finance. Each proposed dynamic relies on specific assumptions whose validity must be assessed by comparing the model with observed market returns.

Historically, financial models have steadily increased in complexity.

The dynamics of asset returns are typically decomposed into a drift component and a diffusion component:

$$\frac{dS_t}{S_t} = \text{Mean term} + \text{Diffusion term}.$$

The elementary model is the Geometric Brownian Motion (GBM), defined by

$$\frac{dS_t}{S_t} = \mu\, dt + \sigma\, dW_t.$$

This specification assumes that returns are lognormal, or equivalently, that log-returns follow a normal distribution. This follows directly from Itô's lemma applied to $\log S_t$:

$$d\log S_t = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma\, dW_t \Rightarrow \underbrace{\log\left(\frac{S_t}{S_0}\right)}_{\text{log-return } X_t} = \left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W_t. \tag{1}$$

The latter expression (1) is expressed relatively to time zero. This can be extended to every nonzero time u because Brownian motion has stationary increments by definition:

$$\underbrace{\log\left(\frac{S_t}{S_u}\right)}_{\text{log-return } X_{t-u}} = \left(\mu - \frac{\sigma^2}{2}\right)(t-u) + \sigma \underbrace{(W_t - W_u)}_{\overset{d}{=} W_{t-u}}.$$

Consequently, the distribution of the log-returns is driven by the normally-distributed Brownian part W_t, whose mean is shifted by the drift of X_t :

$$W_t \sim \mathcal{N}(0,t) \quad \Rightarrow \quad X_t \sim \mathcal{N}\left(\left(\mu - \frac{\sigma^2}{2}\right) t,\, \sigma^2 t\right).$$

The normality of log-returns is surely a limitation of this model. Empirical observations of heavy tails, skewness and abrupt movements in financial returns have motivated the introduction of jump-diffusion models (JD) [2], where an additional jump component $(L_t)_{t\geq 0}$ is added:

$$\frac{dS_t}{S_t} = (\mu - C)\, dt + \sigma\, dW_t + dL_t,$$

with

$$L_t = \sum_{k=1}^{N_t} Y_k,$$

and where C is the compensator which ensures that jumps do not contribute to the mean value of the dynamic. Under the assumption that jump sizes are i.i.d.,

$$Y_k \sim Y,$$

and independent from the jump arrival process $(N_t)_{t\geq 0}$, modeled as an homogeneous Poisson process with intensity λ, i.e.

$$N_t \sim \text{Poi}(\lambda),$$

the compensator takes the following expression:

$$C = \mathbb{E}(L_t) = \mathbb{E}\left(\sum_{k=1}^{N_t} Y_k\right) = \underbrace{\mathbb{E}(N_t)}_{\lambda}\, \mathbb{E}(Y).$$

Similarly to GBM analysis, using the generalized Itô lemma, one obtains the dynamic of the log-returns (2), whose distribution is not gaussian anymore but influenced by jumps. This is the insight of the well-known theory of Levy processes [2].

$$\begin{aligned} d\log S_t &= \left(\mu - \frac{\sigma^2}{2} - \lambda\mathbb{E}(Y)\right) dt + \sigma\, dW_t + dL_t, \\ \log\left(\frac{S_t}{S_0}\right) &= \left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W_t - \lambda\, \mathbb{E}(Y) t + L_t. \end{aligned} \tag{2}$$

The purpose of this thesis is to explore more sophisticated processes, while relying on analytical tools similar to those used above. I will investigate, in particular:

- a generalization of the jump-arrival frequency via non-homogeneous Poisson processes,

  $$N_t \sim \text{Poi}(\lambda_t), \tag{3}$$

  with a focus on intensities exhibiting self-exciting jumps, modeled by (rough) Hawkes processes:

  $$\lambda_t = \lambda_0 + \zeta \int_0^t g(t-s) dN_s; \tag{4}$$

  where $g(-)$ is a memory kernel, i.e. a decreasing function such that

  1. $\lim_{t\to\infty} g(t) = 0$;
  2. $g(0) = 1$.

  I will introduce the general framework of memory kernels, and then define and consider more specifically the rough Hawkes kernel.

- a generalization of stochastic volatility models, such as the Heston model, extended to incorporate a rough Hawkes-type volatility dynamics :

  $$\begin{cases} dX_t = (\mu - \frac{V_t}{2})dt + \sqrt{V_t} dW_t^X \\ dV_t = \kappa(\theta - V_t)dt + \sigma\sqrt{V_t} dW_t^V + dL_t - \lambda_t \mathbb{E}(Y) dt \\ \lambda_t = \lambda_0 + \zeta \int_0^t g(t-s) dN_s \\ \langle dW_t^X, dW_t^V \rangle = \rho dt \end{cases}.$$

In both directions, this thesis will address in detail rough Hawkes processes and their comparison with classical Hawkes processes, exhibiting exponentially decreasing kernels. Memory is introduced either in the jump intensity λ_t or in the stochastic volatility V_t.

The first chapter reviews the necessary theoretical background. In particular, introducing memory in the intensity or volatility renders the system (λ_t, L_t) non-Markovian. This creates challenges for tractability, as the moment generating function of the system cannot be computed directly. To retrieve Markov property, we rely on the approach developed by Hainaut [1], which aggregates all temporal dependence for times $s < t$ into auxiliary state variables, thereby embedding the system into an infinite-dimensional Markov process.

Chapter 2 develops this methodology for the first research direction: modeling the jump intensity with a rough Hawkes process.

Chapter 3 illustrates the implementation of these techniques in option pricing. We perform calibration of model parameters under the real-world probability measure $\mathbb{P}$ and compare pricing results obtained under rough Hawkes dynamics versus other kernels. We will compare the results obtained with the one-dimensional rough Hawkes process studied in Chapter 2 and those from the mutually excited bidimensional rough Hawkes process studied in [1].

Chapter 4 addresses the second research direction: the rough Hawkes–Heston volatility model and its mathematical development. We apply the same methodology than in Chapter 2.

Finally, Chapter 5 draws big lines of the methodology that could be adopt for calibration of the rough Hawkes-Heston volatility model. Namely, the most natural method is based on the replication of market implied volatilities by the model. The large number of parameters the rough Hawkes-Heston model introduces may bring indeterminacy and instabilities in the minimization. Therefore, I present a pseudo-calibration framework under $\mathbb{P}$ based on the conditional chi-squared distribution of the CIR process modeling volatility.

Chapter 1

Introduction to affine Volterra framework

In order to perform the computations in this Chapter, we define a market model $\mathcal{M}$ embedded with

- a risk-free rate, assumed flat;
- asset log-returns X_t, a bank account numeraire $B_t := \exp(-rt)$;
- a jump process L_t defined by the frequency process N_t (3) and random severity drawn from a specific distribution J;
- a filtered probability space
$$(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}) \text{ where}$$
  - Ω is the universe, the set of all scenarios;
  - a tribe $\mathcal{F}$, also known as a σ-algebra;
  - a natural filtration $\mathcal{F}_t$ embedding all information in the system up to time t;
  - $\mathbb{F}$ describes the increasing set of filtrations $(\mathcal{F}_t)_{t=0}^{T}$, such that $\mathcal{F}_t \subset \mathcal{F}_{t+1}$;
  - $\mathbb{P}$ is the probability measure in the real-world.

An option with maturity T can be mathematically priced through the expectation of the discounted payoff under some risk neutral measure. In general, computing expectations requires the expression of the probability density function (pdf). It can be obtained as the Fourier transform of the characteristic function, up to a constant factor[1]:

$$\begin{aligned} \mathcal{M}_X(z) = \mathbb{E}(\exp(izX_t) \mid \mathcal{F}_0) &= \int_{\mathbb{R}} \exp(izx) f(x)\, dx, \\ \Leftrightarrow \qquad f(x) = \frac{1}{2\pi}\mathcal{F}(\mathcal{M}_X(z)) &= \frac{1}{2\pi}\int_{\mathbb{R}} \exp(-izx)\mathcal{M}_X(z)\, dz. \end{aligned} \tag{1.1}$$

If the system (L_t, N_t, λ_t) is Markovian, it is possible to compute the moment generating function of log-returns X_t. In this thesis, I focus on this mgf for pricing purpose.

[1] The factor 2π arises from the normalization convention used for the Fourier basis in the frequency domain. Depending on the author and the chosen convention, different definitions of the Fourier transform can be found in the literature. In particular, some formulations split the scaling symmetrically by introducing a factor $\frac{1}{\sqrt{2\pi}}$ in both the forward and inverse Fourier transforms. Also: notice that $\mathbb{E}(\exp(-izX_t) \mid \mathcal{F}_0)$ can directly be expressed as the Fourier transform of $f(x)$ without any discussion about this scaling parameter.

However, it is possible to retrieve the moment generating function of any state in the Makovian system.The mgf relates to the characteristic function via

$$\mathbb{E}(\exp(wX_t) \mid \mathcal{F}_0)\Big|_{w=iz} = \mathbb{E}(\exp(wX_t) \mid X_0)\Big|_{w=iz}.$$

At inception, $\mathcal{F}_0 = \{X_0\}$. Moreover, because the system is Markovian, the moment generating function admits a closed-form expression:

$$\mathbb{E}(\exp(wX_t) \mid X_0)\Big|_{w=iz} = \exp(F(w,t,X_0))\Big|_{w=iz}.$$

The Lévy processes presented in the introduction are a particular Markovian case, for which the closed-form expression of the exponent is given by the Lévy–Khintchine representation:

$$F(w,t,X_0) = \psi(w)\,t,$$

where

$$\psi(w) = \mu w + \frac{1}{2}\sigma^2 w^2 + \int_{\mathbb{R}} \left(\exp(wx) - 1 - wx\mathbb{I}_{|x|<1}\right)\,\nu(dx).$$

In the non-Markovian case, the moment generating function does not admit a closed-form expression. One way to recover these Markovian properties is to introduce an infinite number of auxiliary states in the state representation of the system, thereby defining an infinite-dimensional Markovian system, following the procedure developed by Hainaut in [3].

The mathematical developments in this thesis rely on a common theoretical framework to compute this exponent $F(w,t,X_0)$ for the infinite-dimensional Markovian process. The formal tools used are those of affine Volterra processes. We emphasize intuition alongside this formalism. The steps used to "Markovianize" the system are:

1. Express the kernel involved in the memory process via its spectral representation, in order to define the auxiliary states $(Y_t)_{\forall t}$ that yield the infinite-dimensional Markovian representation of the system $(X_t, N_t, \lambda_t, (Y_t)_{\forall t})$.

2. Approximate the infinite-dimensional Markovian system by an n-dimensional discrete model. This allows working in a finite-dimensional space using classical Markov theory results. The discretization is performed so that the infinite-dimensional model is recovered in the limit.

3. Observe that the generator of the Markov process preserves the class of affine exponential functions. This allows us to determine *a priori* the general form of the moment generating function.

4. Solve the Kolmogorov equation induced by this generator, which reduces to a system of "generalized" Ricatti equations.

1.1 Step 1 : Spectral representation of the kernel

We endow our computations with a measurable space equipped by the Borelian measure $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$, a filtration $\mathcal{F}_t = \sigma((X_s, N_s, \lambda_s); s \leq t)$ and a positive measure

$$\mu : \mathcal{B}(\mathbb{R}^+) \to [0, \infty].$$

For the latter of this thesis from now, we consider that the measure μ is differentiable so that there exists a density $f(-)$ such that $\mu(dx) = f(x)dx$ for all increments dx. This is the case for all following kernels $g(-)$ considered in literature by Hainaut [4]:

1. Laplace distribution :

$$f(u) = \frac{1}{2}\left(\alpha \exp(-\alpha u)\,\delta_{u\geq 0} + \alpha \exp(\alpha u)\,\delta_{u<0}\right) \tag{1.2}$$

$$g(h) = \frac{\alpha^2}{\alpha^2 + h^2} \tag{1.3}$$

2. Gaussian spectral measure :

$$f(u) = \sqrt{\frac{\alpha}{\pi}} \exp(-\alpha u^2) \tag{1.4}$$

$$g(h) = \exp\left(-\frac{h^2}{4\alpha}\right) \tag{1.5}$$

3. Logistic distribution :

$$f(u) = \frac{\exp\left(-\frac{u}{\alpha}\right)}{\alpha\left(1 + \exp\left(-\frac{u}{\alpha}\right)\right)^2} \tag{1.6}$$

$$g(h) = \begin{cases} 1, & h = 0, \\ \dfrac{\pi\alpha h}{\sinh(\pi\alpha h)}, & h > 0 \end{cases} \tag{1.7}$$

4. Cauchy distribution (exponential kernel) :

$$f(u) = \frac{1}{\pi}\frac{\alpha}{\alpha^2 + u^2} \tag{1.8}$$

$$g(h) = \exp(-\alpha|h|) \tag{1.9}$$

5. Rough Hawkes kernel :

$$f(u) = \frac{u^{-\alpha}}{\Gamma(\alpha)\Gamma(1-\alpha)} \tag{1.10}$$

$$g(h) = \frac{h^{\alpha-1}}{\Gamma(\alpha)}, \qquad \alpha \in \left(\frac{1}{2}, 1\right) \tag{1.11}$$

When considering a kernel function $g(-)$, in most case, it admits the following spectral representation :

$$g(h) = \int_{\mathbb{R}} \exp(ihu)\mu(du) = \int_{\mathbb{R}} \exp(ihu) f(u) du. \tag{1.12}$$

Actually, this representation exists if and only if $\frac{1}{2\pi}g(h)$ has a Fourier transform, noted $f(u)$. In fact, the previous integral corresponds to the inverse Fourier transform of g up to a constant 2π. This factor corresponds to the scaling applied to the Fourier basis functions:

$$g(h) = \frac{1}{2\pi}\int_{\mathbb{R}} \exp(ihu)(2\pi f(u))du = 2\pi\mathcal{F}^{-1}(f(u)) \Leftrightarrow f(u) = \frac{1}{2\pi}\mathcal{F}(g(h)). \tag{1.13}$$

Since the Fourier transform is a linear operator, $\frac{1}{2\pi}g(h)$ has a transform if and only if $g(h)$ has one. This is the case if $g \in L^1(\mathbb{R})$:

$$\int_{\mathbb{R}} |g(h)|dh < \infty.$$

We incorporate this representation in the intensity dynamic λ_t:

$$\begin{aligned}
\lambda_t &= \lambda_0 + \zeta \int_0^t g(t-s)dN_s \\
&= \lambda_0 + \zeta \int_0^t \int_{\mathbb{R}} \exp(iu(t-s))\mu(du)dN_s \\
&= \lambda_0 + \zeta \int_{\mathbb{R}} \int_0^t \exp(iu(t-s))dN_s\mu(du) \quad \text{(Fubini's theorem)} \\
&= \lambda_0 + \zeta \int_{\mathbb{R}} Y_t(u)\mu(du)
\end{aligned}$$

We introduce a continuous variable $Y_t(u) \quad \forall u \in \mathbb{R}$, corresponding to an infinite number of states $\{Y_t(u)\}_{u\in\mathbb{R}}$ in the system. Therefore, the system

$$(L_t, N_t, \lambda_t, Y_t(u)) \quad \forall u \in \mathbb{R}$$

is an infinite Markovian system where

$$\begin{cases} d\lambda_t = \zeta \int_{\mathbb{R}} dY_t(u)\mu(du) \\ dY_t = iuY_t dt + dN_t \end{cases}.$$

We have embedded all the memory in the intensity of the Poisson process into an infinite number of new states to recover the "Markovianity" of the system. Considering these states, we have in some sense recovered the Markov properties, and we can compute the moment generating function using usual techniques after a n-dimensional discretization.

However, even when $g(-)$ does not have a Fourier transform, its Laplace transform $\mathcal{L}(-)$ can be considered :

$$g(h) = \int_{\mathbb{R}} \exp(zh)\mu(dz) = \int_{\mathbb{R}} \exp(zh)f(z)dz = \mathcal{L}^{-1}(f(z)) \tag{1.14}$$

where $z = v + iu$, $z \in \mathbb{C}$, $v \in \mathbb{R}$. In other words, a spectral representation can be retrieve even when $g \notin L^1(\mathbb{R})$, if the dampened version of the kernel $\exp(-vh)g(h) \in L^1(\mathbb{R})$.

$$\begin{aligned}
\exp(-vh)g(h) \in L^1(\mathbb{R}) &\Leftrightarrow \int_{\mathbb{R}} |\exp(-vh)g(h)|dh < \infty \\
&\Leftrightarrow \exp(-vh)g(h) \text{ has a Fourier transform} \\
&\Leftrightarrow g(h) \text{ has a Laplace transform.}
\end{aligned}$$

The remaining computations are exactly similar to the previous case :

$$\begin{aligned}
\lambda_t &= \lambda_0 + \zeta \int_0^t g(t-s)dN_s \\
&= \lambda_0 + \zeta \int_0^t \int_{\mathbb{R}} \exp(z(t-s))\mu(dz)dN_s \\
&= \lambda_0 + \zeta \int_{\mathbb{R}} \int_0^t \exp(z(t-s))dN_s\mu(dz) \quad \text{(Fubini's theorem)} \\
&= \lambda_0 + \zeta \int_{\mathbb{R}} Y_t(z)\mu(dz)
\end{aligned}$$

The system

$$(L_t, N_t, \lambda_t, Y_t(z)) \quad \forall z \in \mathbb{C}$$

is an infinite Markovian system described by the dynamics:

$$\begin{cases} d\lambda_t = \zeta \int_{\mathbb{R}} dY_t(u)\mu(du) \\ dY_t = (v + iu)Y_t dt + dN_t \end{cases}.$$

In case of rough Hawkes kernel (1.11), we can adopt another point of view for finding the infinite number of additional states. Either we can consider :

- As previously, as $g(-)$ has no Fourier transform, its dampened version $k(-)$ has got one:

$$k(h) = \exp(-vh)g(h) = \exp(-vh)\frac{h^{\alpha-1}}{\Gamma(\alpha)}.$$

  This is the insight of the Laplace transform. However, for $\alpha \in (\frac{1}{2}, 1)$, the kernel has one singularity at $h = 0$, which makes the bilateral Laplace transform considered in previous developments not well-defined. In that case, we simply consider the unilateral Laplace transform defined on $\mathbb{R}_+$ instead of $\mathbb{R}$.

- Otherwise, the reader can observe that $k(-)$ is a Sonine function (1.15). Using this property, I show how we can recover a Markovian representation of (L_t, N_t, λ_t).

Sonine functions

A function $k(-)$ is a Sonine function if there exists a function $l(-)$ such that

$$(l * k)(h) = 1 \quad \forall h \tag{1.15}$$

Sonine functions are mainly useful because the pair $(k(-), l(-))$ is the convolutional kernel of Sonine operators (1.22).

Sonine operators

Sonine operators for $\phi \in L^1(\mathbb{R}^+)$ are defined as

$$\begin{cases} (k * \phi)(t) = K\phi(t) \\ (l * \phi)(t) = L\phi(t) \end{cases}$$

such that

$$K^{-1}\phi(t) = \frac{d}{dt}(L\phi)(t) \tag{1.16}$$

Proof. We take the Laplace transform $\mathcal{L}(-)$ of both sides in Equation (1.15), and we define $L(s) := \mathcal{L}(l(t))$ and $K(s) := \mathcal{L}(k(t))$. This results in

$$\underbrace{L(s)}_{\mathcal{L}(l(t))} \times \underbrace{K(s)}_{\mathcal{L}(k(t))} = \frac{1}{s}. \tag{1.17}$$

Especially, if we define $\psi(t) = \frac{d}{dt}(L\phi(t))$, $\Phi(s) := \mathcal{L}(\phi(t))$ and $K(s) := \mathcal{L}(k(t))$, the Laplace transform of the first Sonine operator can be computed as

$$\begin{aligned} \mathcal{L}(K\psi(t)) &= \mathcal{L}((k * \psi)_t) \\ &= \mathcal{L}(k(t))\mathcal{L}(\psi(t)) \\ &= K(s)\Psi(s). \end{aligned} \tag{1.18}$$

Using the derivative properties of $\psi(t) = \frac{d}{dt}(L\phi(t))$, we compute the Laplace transform of ψ:

$$\begin{aligned}\Psi(s) &= sL(s)\Phi(s) \\ &= K^{-1}\Phi(s) \quad \text{in view of (1.17)}.\end{aligned} \tag{1.19}$$

Combining Equations (1.18) and (1.19),

$$\mathcal{L}(K\psi(t)) = K(s)K^{-1}\Phi(s) = \Phi(s) = \mathcal{L}(\phi(t)). \tag{1.20}$$

By uniqueness of Laplace transform in a certain region of convergence, $K\psi(t) = \phi(t)$. We insert the definition of $\psi(t)$, and we conclude :

$$K\psi(t) = K\frac{d}{dt}(L\phi(t)) = \phi(t) \Rightarrow K^{-1}\phi(t) = \frac{d}{dt}(L\phi(t)). \tag{1.21}$$

□

In the specific rough Hawkes case (1.11), closed-form expressions can be computed for $l(-)$. It is well-known that $K(s) = \frac{1}{(s+v)^\alpha}$ from basic Laplace transform tables. By Equation (1.17):

$$\begin{aligned}K(s)L(s) = \frac{1}{s} &\Rightarrow L(s) = \frac{(s+v)^\alpha}{s} \\ &\Rightarrow l(h) = \mathcal{L}^{-1}(L(s)) = v^\alpha + \frac{\alpha}{\Gamma(1-\alpha)}\int_h^\infty \frac{\exp(-vs)}{s^{1+\alpha}}ds.\end{aligned} \tag{1.22}$$

The resulting Markovian system is recovered after embedding all memory present in the first Sonine operator K, applied to a measure ϕ:

$$K\phi(t) = (k*\phi)(t) = \int_0^t k(t-s)\phi(s)ds = \int_0^t \exp(-v(t-s))\frac{(t-s)^{\alpha-1}}{\Gamma(\alpha)}\phi(s)ds \tag{1.23}$$

By the convolution definition, $K\phi(t)$ can be viewed as a dampened version of the Riemann-Liouville integral

$$I^\alpha\phi = \int_0^t \frac{(t-s)^{\alpha-1}}{\Gamma(\alpha)}\phi(s)ds,$$

exactly the same way we defined the unilateral Laplace transform as a dampened Fourier transform.

From the fundamental Gamma function identity, we further expand the integral (1.23):

$$(t-s)^{\alpha-1} = \int_0^\infty \exp(-\xi(t-s))\frac{\xi^{-\alpha}}{\Gamma(1-\alpha)}d\xi. \tag{1.24}$$

Finally, memory in (1.23) is embedded the same way we did using the spectral transforms:

$$K\phi(t) = \int_0^t \int_0^\infty \exp(-\xi(t-s))\frac{\xi^{-\alpha}}{\Gamma(1-\alpha)}d\xi\frac{\exp(-v(t-s))}{\Gamma(\alpha)}\phi(s)ds \tag{1.25}$$

$$= \int_0^t \int_0^\infty \exp(-(v+\xi)(t-s))\frac{\xi^{-\alpha}}{\Gamma(\alpha)\Gamma(1-\alpha)}d\xi\phi(s)ds \tag{1.26}$$

$$= \int_0^\infty \int_0^t \exp(-(t-s)(v+\xi))\phi(s)ds\frac{\xi^{-\alpha}}{\Gamma(\alpha)\Gamma(1-\alpha)}d\xi \quad \text{(Fubini's theorem)} \tag{1.27}$$

$$= \int_0^\infty Y_t(\xi)\underbrace{f(\xi)d\xi}_{\mu(d\xi)} \tag{1.28}$$

In our setting, we observe that $\phi(s)ds = dN_s$. Hence, the system has the following infinite Markovian formulation :

$$\begin{cases} d\lambda_t = \zeta(k * dN)(t) \\ \qquad = \zeta \displaystyle\int_0^t \exp(-v(t-s)) \frac{(t-s)^{\alpha-1}}{\Gamma(\alpha)} dN_s \\ \qquad = \zeta \displaystyle\int_0^\infty dY_t(\xi)\mu(d\xi) \\ dY_t = -(v+\xi)Y_t dt + dN_t \end{cases}$$

At the end of this section, we are able to embed all the memories of the process into an infinite number of state variables, leading to an infinite dimensional Markov process. The main ideas are summed up in the decision tree presented in Figure 1.1. However, all the necessary Markovian tools have been developed in a finite dimensional setting : that's why we introduce a finite-dimensional approximation of the system in the next section. This is very important that we model our approximation in a n-dimensional space for which the limit recovers the infinite dimensional Markovian model. Such convergence results are essential to ensure we will recover the moment generating function of initial system as a limit of the discretization's one.

We conclude this section with a consideration about the long-run intensity.

Long-term mean value

The long-term mean jump intensity defined with a memory kernel, driven by process (4), is

$$\frac{\lambda_0}{1-\zeta K(0)} \tag{1.29}$$

Proof. The spectral transforms we have introduced allow us to find a closed-form expression for the mean process

$$u(t) := \mathbb{E}(\lambda_t \mid \mathcal{F}_0) \tag{1.30}$$

In this expression, replace λ_t with its process definition (4), and apply Laplace transform:

$$\begin{aligned} u(t) &= \lambda_0 + \zeta \mathbb{E}((k * dN)_t \mid \mathcal{F}_0) \\ &= \lambda_0 + \zeta(k * \underbrace{\mathbb{E}(dN \mid \mathcal{F}_0)}_{u(t)})_t \\ \Rightarrow U(s) &= \frac{\lambda_0}{s} + \zeta K(s)U(s) \\ \Rightarrow U(s)[1-\zeta K(s)] &= \frac{\lambda_0}{s} \\ \Rightarrow U(s) &= \frac{\lambda_0}{s[1-\zeta K(s)]}. \end{aligned} \tag{1.31}$$

Therefore, we have the following asymptotical result:

$$u(t) = \mathcal{L}^{-1}(U(s)), \ \lim_{t\to\infty} u(t) = \lim_{s\to 0^+} sU(s) = \frac{\lambda_0}{1-\zeta K(0)} \tag{1.32}$$

□

This limit actually exists only if $1-\zeta K(s)$ is invertible and $u(t)$ has a limit. The first condition is satisfied if $\det(1-\zeta K(s)) \neq 0$. The second requires that $\frac{\lambda_0}{s[1-\zeta K(0)]}$ is

an accumulation point. Actually, this is the case because we can expand the expression into a converging serie:

$$\lim_{K\to\infty} \frac{\lambda_0}{s} \sum_{i=0}^{K} (\zeta K(0))^i = \frac{\lambda_0}{s[1-\zeta K(0)]} = \frac{\lambda_0}{s}(1 + \zeta K(0) + (\zeta K(0))^2 + \dots).$$

This is a well-known result from basis mathematical classes. The convergence is therefore ensured if $(\zeta K(0))^k$ tends to 0, or differently said if the spectral radius of $\zeta K(0)$ is lower than 1 :

$$\rho(\zeta K(0)) = \max_i |\lambda_i[\zeta K(0)]| < 1,$$

where $\{\lambda_i[\zeta K(0)] \mid i \in [n]\}$ is the set of all eigenvalues of $\zeta K(0)$.

Taking both conditions together, since $K(0)$ is scalar in our case, this simply results in the condition

$$\zeta K(0) < 1.$$

This can be directly generalized to higher dimensions systems. Intuitively, we can see the memory in the jump process as an expanding tree. When one jump occurs at rate $\zeta K(0)$, it will expand up to the generation k. If $(\zeta K(0))^k$ converges to zero for relatively small k, then the memory process is stable. It requires to have less than one children per parent node, meaning a reproduction rate lower than one.

In the specific case of dampened Hawkes kernel, the long-term mean is

$$K(s) = \frac{1}{(s+v)^\alpha} \Rightarrow \lim_{t\to\infty} u(t) = \frac{\lambda_0 v^\alpha}{v^\alpha - \zeta}, \tag{1.33}$$

and the convergence condition becomes $v^\alpha > \zeta$.

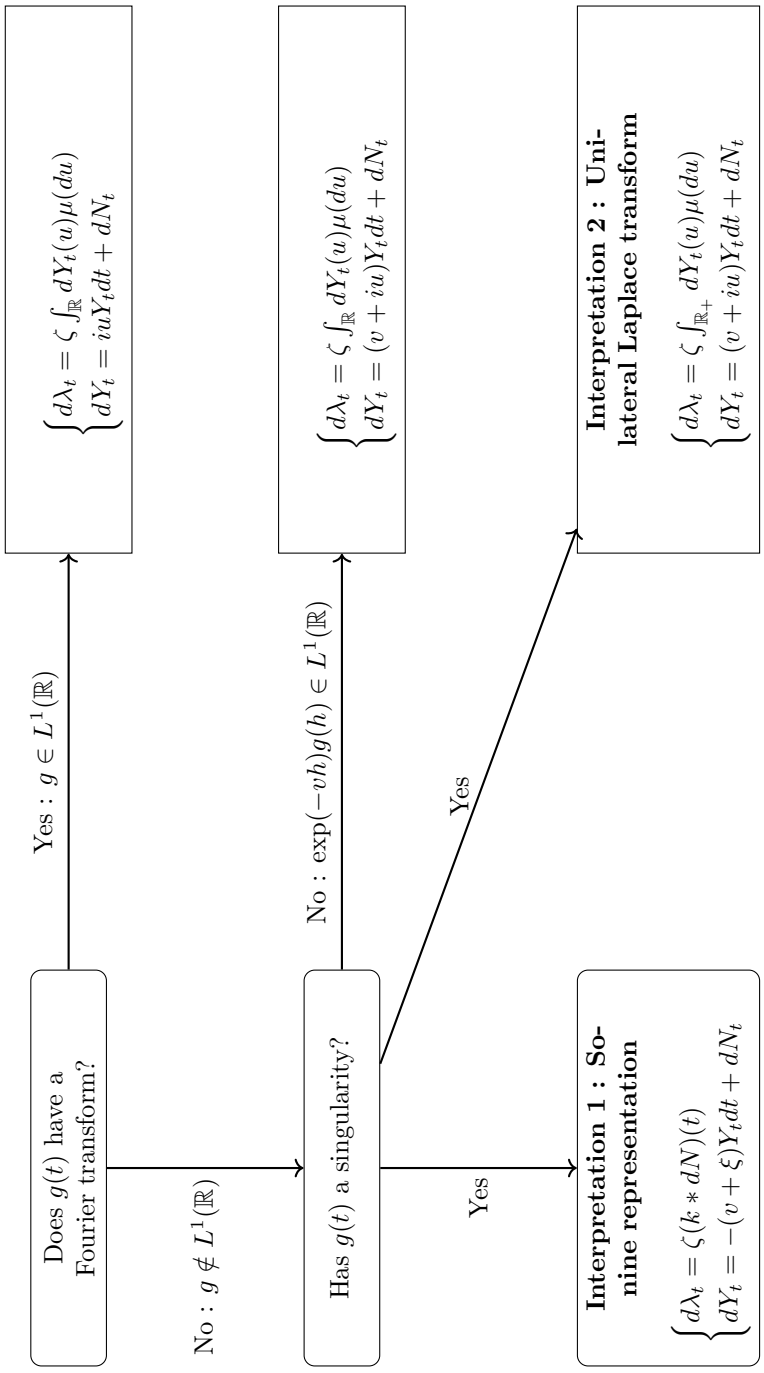

Figure 1.1: Decision tree depending on properties of $g(t)$.

1.2 Step 2 : Discretization of the infinite Markovian system

The discretization essentially depends on the choice of unilateral or bilateral transform in the previous section. We have the continuous state variable $\{Y_t(u)\}_u$ for $u \in \mathbb{R}$ in the bilateral case and $u \in \mathbb{R}_+$ in the unilateral case. The discretization is detailled by Procedure1.2.

We reduces this infinite dimensional model to a finite-dimensional model with n (resp. $2n$ in the bilateral case) state variables.

Procedure for state space discretization in the unilateral case

We approximate the integral piecewise using an increasing partition of n points:

$$\mathcal{E}^{(n)} = \{\xi_0^{(n)}, \dots, \xi_{n-1}^{(n)}\} \subset \mathcal{E}^{(n+1)} = \{\xi_0^{(n+1)}, \dots, \xi_n^{(n+1)}\},$$

in such a way that the integral is recovered in the limit:

$$\int_{\mathbb{R}_+} Y_t(u)\,\mu(du) = \lim_{n\to\infty} \sum_{l=0}^{n-1} \int_{\xi_l^{(n)}}^{\xi_{l+1}^{(n)}} Y_t(u)\,\mu(du). \tag{1.34}$$

We approximate $Y_t(u)$ on the interval $[\xi_l^{(n)}, \xi_{l+1}^{(n)}]$ by its value at the midpoint, denoted $\widetilde{Y}_t(b_l)$, where

$$b_l = \frac{\xi_l^{(n)} + \xi_{l+1}^{(n)}}{2}. \tag{1.35}$$

Therefore, the approximation can be rewritten as

$$\sum_{l=0}^{n-1} \widetilde{Y}_t(b_l)\, w_l, \quad \text{where} \quad w_l = \int_{\xi_l^{(n)}}^{\xi_{l+1}^{(n)}} \mu(du). \tag{1.36}$$

Proof. This corresponds to the Galerkin procedure in the Hilbert space $L^2(\mu)$, defined as set of functions squared-integrable according to the measure $\mu(-)$

$$L^2(\mu) = \{f : \int_{\mathbb{R}_+} |\, f(u)\,|^2\, \mu(du) < \infty\}.$$

This space is endowed with the following inner product :

$$\langle f, g\rangle_{L^2} = \int_{\mathbb{R}_+} f(u)g(u)\mu(du).$$

Results in this setting are already well-known, but I remind some intuition in the following. The proof mainly consists in 3 parts :

1. Showing why taking the central value b_l of each interval is relevant;
2. Showing that Y_t can be recovered in the limit of its finite dimensional approximation $\tilde{Y}_t^{(n)}$;
3. Hence, showing the convergence result for the quadrature.

(1) The quadrature in question is equivalent to the approximation of the measure $\mu(-)$ and the state variable $Y_t(-)$ in the basis functions $\{\phi_l)\}_{l=0}^{n-1}$:

$$\mu(x) \approx \tilde{\mu}(x) = \sum_{l=0}^{n-1} w_l \phi_l(x),\ Y_t(x) \approx \tilde{Y}_t^{(n)}(l) = \sum_{l=0}^{n-1} c_l \phi_l(x).$$

The linear combination $\{c_l\}_{l=0}^{n-1}$ and $\{w_l\}_{l=0}^{n-1}$ are determined in the following discussions. Galerkin methods are used to approximate solutions of differential equations by projecting onto a finite-dimensional subspace spanned by chosen basis functions. We choose the approximated finite space

$$\mathcal{V}^{(n)} = \text{span}\langle \phi_0 = \delta_{[\xi_0^{(n)};\xi_1^{(n)})}, \dots, \phi_{n-1} = \delta_{[\xi_{n-1}^{(n)};\xi_n^{(n)}]} \rangle \subset L^2(\mu),$$

where the partitions $\mathcal{E}^{(n)}$ are chosen such that :

- $\lim_{n\to\infty} \xi_0^{(n)} = 0, \lim_{n\to\infty} \xi_n^{(n)} = \infty$. (the initial domain $\mathbb{R}_+$ is recovered in the limit)
- $\lim_{n\to\infty} \max_i \left|\xi_{i+1}^{(n)} - \xi_i^{(n)}\right| = 0$ (the partition becomes finer when n grows).

In consequence, $\lim_{n\to\infty} \mathcal{V}^{(n)} = L^2(\mu)$. The Hilbert space ensures that there exists an unique orthogonal projection onto each subspace $\mathcal{V}^{(n)}$:

$$\tilde{Y}_t^{(n)}(l) = \text{Proj}_{\mathcal{V}^{(n)}} Y_t = \arg\min_{v\in\mathcal{V}^{(n)}} \langle Y_t - v, Y_t - v\rangle_{L^2} = ||Y_t - v||_{L^2}^2.$$

By the orthogonal projection definition, the residual is orthogonal to the basis functions, i.e. $\forall k = 0, \dots, n-1$:

$$\begin{aligned} &\langle Y_t - \sum_{l=0}^{n-1} c_l \phi_l, \phi_k \rangle_{L^2} = 0 \\ \Leftrightarrow\ &\langle Y_t, \phi_k \rangle_{L^2} - \sum_{l=0}^{n-1} c_l \langle \phi_l, \phi_k \rangle_{L^2} = 0 \end{aligned} \tag{1.37}$$

This expression 1.37 is generally called the Galerkin condition. In order to develop this integral and isolate the factors $\{c_k\}_{k=0}^{n-1}$, we observe that the basis functions are orthogonal in $L^2(\mu)$:

$$\langle \phi_l, \phi_k \rangle_{L^2} = \int_{\mathbb{R}_+} \phi_l(u)\phi_k(u)\mu(du) = \begin{cases} w_k, & l = k, \\ 0, & l \neq k, \end{cases}$$

where

$$w_k = \int_{\xi_{k-1}^{(n)}}^{\xi_k^{(n)}} \mu(du).$$

Moreover, since $\phi_k = \delta_{[\xi_{k-1}^{(n)}, \xi_k^{(n)}]}$, we have

$$\langle Y_t, \phi_k \rangle = \int_{\xi_{k-1}^{(n)}}^{\xi_k^{(n)}} Y_t(u)\mu(du).$$

Therefore, the Galerkin condition implies

$$\langle Y_t, \phi_k \rangle - c_k w_k = 0 \text{ and, finally } c_k = \frac{1}{w_k} \int_{\xi_{k-1}^{(n)}}^{\xi_k^{(n)}} Y_t(u)\mu(du).$$

The mathematical expression of these c_k corresponds to the average value of Y_t on the interval $[\xi_{k-1}^{(n)}; \xi_k^{(n)}]$. This mean value is taken at the centre of each interval, whose image is the actual mean if considering a composite trapezoidal or piecewise constant approximation:

$$b_{k-1} = \frac{\xi_{k-1}^{(n)} + \xi_k^{(n)}}{2}.$$

Hence, Y_t is approximated on each interval $[\xi_{k-1}^{(n)}; \xi_k^{(n)}]$ by $\tilde{Y}_t(b_{k-1})$.

(2) The discretization is constructed from partitions $\{\xi_i^{(n)}\}_{i=0}^{n-1}$ satisfying

$$\max_i |\xi_{i+1}^{(n)} - \xi_i^{(n)}| \to 0.$$

Functions that are piecewise constant on these partitions generate finite-dimensional spaces $\mathcal{V}^{(n)}$ whose union is dense in $L^2(\mu)$:

$$\overline{\bigcup_{n\geq 1} \mathcal{V}^{(n)}} = L^2(\mu).$$

Indeed, piecewise constant functions are simple functions, and simple functions are dense in $L^2(\mu)$. Moreover, as the mesh size tends to zero, the cell averages defining the Galerkin projection converge to the original function in $L^2(\mu)$, which may also be justified using the Lebesgue differentiation theorem.

Therefore, for every $Y_t \in L^2(\mu)$,

$$\lim_{n\to\infty} \inf_{v\in\mathcal{V}^{(n)}} \|Y_t - v\|_{L^2(\mu)} = 0.$$

Since $\widetilde{Y}_t^{(n)}$ denotes the orthogonal projection of Y_t onto $\mathcal{V}^{(n)}$, it realizes this infimum, hence

$$\lim_{n\to\infty} \|Y_t - \widetilde{Y}_t^{(n)}\|_{L^2(\mu)} = 0.$$

Consequently, for m, n sufficiently large,

$$\|\widetilde{Y}_t^{(n)} - \widetilde{Y}_t^{(m)}\|_{L^2(\mu)} \leq \|\widetilde{Y}_t^{(n)} - Y_t\|_{L^2(\mu)} + \|\widetilde{Y}_t^{(m)} - Y_t\|_{L^2(\mu)} \to 0.$$

Thus, $(\widetilde{Y}_t^{(n)})_{n\geq 1}$ is a Cauchy sequence in $L^2(\mu)$. Since $L^2(\mu)$ is complete, the sequence converges in $L^2(\mu)$ to Y_t.

(3) Notice that

$$\begin{aligned}
|\int_{\mathbb{R}_+} Y_t(u)\mu(du) - \int_{\mathbb{R}_+} \sum_{l=0}^{n-1} \tilde{Y}_t^{(n)}(b_l)\phi_l(x)\mu(du)| &= |\langle Y_t, 1\rangle_{L^2} - \langle \sum_{l=0}^{n-1} \tilde{Y}_t^{(n)}(b_l)\phi_l(x), 1\rangle_{L^2}| \\
&= |\langle Y_t - \sum_{l=0}^{n-1} \tilde{Y}_t^{(n)}(b_l)\phi_l(x), 1\rangle_{L^2}| \\
&\leq ||Y_t - \sum_{l=0}^{n-1} \tilde{Y}_t^{(n)}(b_l)\phi_l(x)||_{L^2}||1||_{L^2} \\
&= ||Y_t - \tilde{Y}_t^{(n)}(l)||_{L^2} \to 0
\end{aligned}$$

where the last inequality holds thanks to Cauchy-Schwartz inequality. □

In the specific case of Hawkes kernels, if we remind the expression for the measure density function $f(-)$ (1.11), we can compute

$$w_l = \int_{\xi_l^{(n)}}^{\xi_{l+1}^{(n)}} f(u)du = \frac{[\xi_{l+1}^{(n)}]^{1-\alpha} - [\xi_l^{(n)}]^{1-\alpha}}{(1-\alpha)\Gamma(1-\alpha)}.$$

Procedure for state space discretization in the bilateral case

Essentially, the technique is similar to two one-sided procedures applied to the subsets $l \in A^n = \{-n, \dots, -1\}$ and $l \in B^n = \{0, \dots, n-1\}$ such that

$$w_{-l} = -w_l, \quad b_{-l} = -b_l, \quad \xi_{-l}^{(n)} = -\xi_l^{(n)}.$$

We approximate the integral piecewise using an increasing partition of n points:

$$\mathcal{E}^{(n)} = \{\xi_l^{(n)}\}_{l \in A^n \cup B^n} \subset \mathcal{E}^{(n+1)} = \{\xi_l^{(n+1)}\}_{l \in A^{n+1} \cup B^{n+1}},$$

in such a way that the integral is recovered in the limit:

$$\int_{\mathbb{R}} Y_t(u)\,\mu(du) = \lim_{n\to\infty} \sum_{l=-n}^{n-1} \int_{\xi_l^{(n)}}^{\xi_{l+1}^{(n)}} Y_t(u)\,\mu(du). \tag{1.38}$$

We approximate $Y_t(u)$ on the interval $[\xi_l^{(n)}, \xi_{l+1}^{(n)}]$ by its value at the midpoint, denoted $\widetilde{Y}_t(b_l)$, where

$$b_l = \frac{\xi_l^{(n)} + \xi_{l+1}^{(n)}}{2}, \quad \forall l \in A^n \cup B^n. \tag{1.39}$$

Therefore, the approximation can be rewritten as

$$\sum_{l=-n}^{n-1} \widetilde{Y}_t(b_l)\, w_l, \quad \text{where} \quad w_l = \int_{\xi_l^{(n)}}^{\xi_{l+1}^{(n)}} \mu(du). \tag{1.40}$$

Proof. Similar to the proof of 1.2. □

In conclusion to this step, we have a finite dimensional Markov system $\tilde{Z}_t = (\tilde{X}_t, \tilde{\lambda}_t, \tilde{Y}_t(l))$. In the unilateral case :

$$\begin{cases} d\tilde{\lambda}_t = \sum_{l=0}^{n-1} \zeta w_l d\tilde{Y}_t(l) \\ d\tilde{Y}_t(l) = \alpha_l \tilde{Y}_t(l) + d\tilde{N}_t \end{cases} \tag{1.41}$$

In the bilateral case :

$$\begin{cases} d\tilde{\lambda}_t = \sum_{l=-n}^{n-1} \zeta w_l d\tilde{Y}_t(l) \\ d\tilde{Y}_t(l) = \alpha_l \tilde{Y}_t(l) + d\tilde{N}_t \end{cases} \tag{1.42}$$

In each case,

- $\alpha_l = ib_l$ if the kernel has a Fourier transform,
- $\alpha_l = v + ib_l$ if the kernel has a Laplace transform,
- $\alpha_l = -(v + b_l)$ if the kernel is a Sonine function.

1.3 Step 3 : Invariance of the Markovian generator of affine processes

For affine processes, it is possible to guess the generic expression of the moment generating function because the space of affine exponential is stable through the Markovian dynamic. We will show this intuition in this section. For this part of the thesis, we assume a generic finite-dimensional model

$$\{X_t^0, \dots, X_t^k, Y_t^0(l), \dots, Y_t^m(l)\}.$$

This setting has exactly the structure of an approximated model similar to the one we obtained in the previous section, after applying the discretization procedure to continuous variables $\{Y_t^0(u), \dots, Y_t^m(u)\}$. We make the arbitrary choice that the reader wants to compute the moment generating function of X_t^0,

$$u(t, X_t^0) = \mathbb{E}(\exp(-wX_T^0) \mid \mathcal{F}_t) = \mathbb{E}(\exp(-wX_T^0) \mid X_t).$$

In this thesis, our goal is to recover a closed-form expression for the moment generating function for the asset log-return X_t. We highlight here the fact that the procedure we present in this chapter is much more general. It allows recovering the moment generating function of any of the states in the Markovian system. For instance, in literature [3], Hainaut applied this technique to recover the mgf of λ_t.

We define the transition operator of the Markov setting :

$$(P_h u)(t, x) = \mathbb{E}(u(t + h, X_{t+h}^0) \mid X_t^0 = x). \tag{1.43}$$

The family of operator $\{P_h\}_{h \geq 0}$ is a semigroup, which means that :

- $(P_0 u)(t, x) = \mathbb{E}(u(t, X_t^0) \mid X_t^0 = x) = u(t, x)$;
- $(P_{h+v} u)(t, x) = \mathbb{E}(u(t + h + v, X_{t+h+v}^0) \mid X_t^0 = x) = (P_v P_h u)(t, x)$.

The generator of the Markov system is then defined as

$$\begin{aligned} \mathcal{A}u(t, x) &= \lim_{h \to 0^+} \frac{(P_h u)(t, x)) - u(t, x)}{h} \\ &= \lim_{h \to 0^+} \frac{\mathbb{E}(u(t + h, X_{t+h}^0) \mid X_t^0 = x) - u(t, x)}{h}. \end{aligned}$$

We obtain the backward equations using martingale property and first-order expansion of the right-hand in Equation 1.43 :

$$u(t + h, x) = u(t, x) + h\partial_t u(t, x).$$

Hence, Equation 1.43 becomes

$$(P_h u)(t, x) = \mathbb{E}(u(t, X_{t+h}^0) \mid X_t^0 = x) + h\mathbb{E}(\partial_t u(t, X_{t+h}^0) \mid X_t^0 = x); \tag{1.44}$$

The dynamic programming principle states that

$$\sup_u \{(P_h u)(t, x) + f(t, x)\} = u(t, x).$$

When no control is applied on the system, the reward function $f(t, x)$ is identically zero, and this principle reduces to the martingale property :

$$(P_h u)(t, x) = u(t, x).$$

The expected value of the function at time $t+h$ is equal to its current value at t. This property inevitably recalls the martingale requirement. The definition of martingale "no reward on average" takes all its sense considering the null reward function in the dynamic programming approach. Applying this principle, it follows that 1.44 gives the Backward Kolmogorov equation :

$$\begin{aligned}
&(P_h u)(t,x) = u(t,x) = \mathbb{E}(u(t, X^0_{t+h}) \mid X^0_t = x) + h\mathbb{E}(\partial_t u(t, X^0_{t+h}) \mid X^0_t = x) \\
&\Rightarrow \mathbb{E}(u(t, X^0_{t+h}) \mid X^0_t = x) - u(t,x) + h\mathbb{E}(\partial_t u(t, X^0_{t+h}) \mid X^0_t = x) = 0 \\
&\Rightarrow \frac{\mathbb{E}(u(t, X^0_{t+h}) \mid X^0_t = x) - u(t,x)}{h} + \mathbb{E}(\partial_t u(t, X^0_{t+h}) \mid X^0_t = x) = 0
\end{aligned} \tag{1.45}$$

Taking the limit for $h \to 0$,

The Backward Kolmogorov equation is:

$$\mathcal{A}u(t,x) + \partial_t u(t,x) = 0$$

We can easily check that the family of affine exponential is invariant through the operator $\mathcal{A}$:

$$u(t,x) = \exp(\Phi(t,T))$$

where

$$\Phi(t,T) = q_0(t,T) + \sum_{j=1}^{k} q_{X^j}(t,T)\tilde{X}^j_t + \sum_{i=0}^{m}\sum_{l} q_{Y^i_l}(t,T)\tilde{Y}^i_t(l) - w\tilde{X}^0_t$$

$$\Rightarrow \mathcal{A}u(t,x) = -\partial_t u(t,x) = -\partial_t \Phi(t,T) u(t,x). \tag{1.46}$$

The sum over l corresponds to the index discussed in the previous section. This mean that $l \in \{0, \dots, n-1\}$ for unilateral case and $l \in \{-n, \dots, 0, \dots, n-1\}$ otherwise. Through Equation 1.46, we observe that at each time step, the affine functions $u(t,x)$ acts as the instantaneous eigenvectors of the generator $\mathcal{A}$ with corresponding instantaneous eigenvalues $\lambda(t) = -\partial_t \Phi(t,T)$. The backward Kolmogorov equation induces a time-inhomogeneous eigenvalue problem for the infinitesimal generator $\mathcal{A}$. Hence, the affine exponential family is invariant under the action of $\mathcal{A}$ and behaves as an instantaneous eigenbasis of the generator. This structure is analogous to Sturm–Liouville spectral theory, where differential operators admit eigenfunctions forming a basis. Here, the role of the Sturm–Liouville eigenfunctions is played by time-dependent exponential-affine modes.

1.4 Step 4 : System of Ricatti-Volterra equations

The last step is the resolution of a system of "generalized" Ricatti equations in order to compute a closed form expression given our parametrization of the moment generating function $q_0(t,T)$, $q_{X^j}(t,T)$, $q_{Y^i,l}(t,T)$.

Here is the classical framework. Consider $f(t, \{X^j_t\}_{j=0}^k, \{Y^i_t(l)\}_{i=0}^m \forall l)$ a function depending on all state variables :

1. Apply the generalized Itô's lemma to the function, in order to compute df
2. Compute the infinitesimal generator

$$\mathbb{E}(\frac{df}{dt} \mid \mathcal{F}_t).$$

3. Based on the infinitesimal generator, we can compute the generator $\mathcal{A}$.
4. Finally, we compare the expressions for $\partial_t u(t,x)$ and $\mathcal{A}u(t,x)$. Since

$$\mathcal{A}u(t,x) + \partial_t u(t,x) = 0,$$

we group the coefficients multiplying each state variable before zeroing them. This is the only way to retrieve the solution since the only way to reach the null vector is using the zero combination of the basis vectors in the Markovian space. This gives birth to $k+1$ ODE, in addition to the m equations resulting for each l.

The last general remark of this section is about the resolution of these ODE's. In fact, the Backward Komolgorov system cannot be solved in case of dampened rough kernel as we will see in the next chapter. In that case, we must consider forward Kolmogorov equations. Let $p(-,-)$ designate the probability density of u. Since we have endowed our computations in an Hilbert space $L^2(\mu)$, we can obtain the forward equations from the backward ones using the density:

The Forward Kolmogorov equation is:

$$\partial_t p(t,x) - \mathcal{A}^* p(t,x) = 0.$$

Proof. We consider the following expectation:

$$\mathbb{E}[u(t,\cdot)] = \langle u(t,\cdot), p(t,\cdot)\rangle = \int_{\mathbb{R}_+} u(t,x)p(t,x)\,dx.$$

Taking the temporal derivative,

$$\partial_t\langle u(t,\cdot), p(t,\cdot)\rangle = \langle \partial_t u(t,\cdot), p(t,\cdot)\rangle + \langle u(t,\cdot), \partial_t p(t,\cdot)\rangle.$$

By the backward Kolmogorov equation, we obtain:

$$\partial_t\langle u(t,\cdot), p(t,\cdot)\rangle = -\langle \mathcal{A}u(t,\cdot), p(t,\cdot)\rangle + \langle u(t,\cdot), \partial_t p(t,\cdot)\rangle.$$

Furthermore, by definition of the adjoint operator $\mathcal{A}^*$,$\langle \mathcal{A}u, p\rangle = \langle u, \mathcal{A}^* p\rangle$:

$$\partial_t\langle u(t,\cdot), p(t,\cdot)\rangle = \langle u(t,\cdot),, \partial_t p(t,\cdot) - \mathcal{A}^* p(t,\cdot)\rangle.$$

Remind that $u(t, X_t^0)$ is a martingale:

$$\mathbb{E}[u(t,X_t^0) \mid \mathcal{F}_0] = u(0,X_0^0),$$

which is constant in t. Therefore,

$$\partial_t \mathbb{E}[u(t,\cdot)] = \partial_t\langle u(t,\cdot), p(t,\cdot)\rangle = 0.$$

We thus obtain:

$$\langle u(t,\cdot), \partial_t p(t,\cdot) - \mathcal{A}^* p(t,\cdot)\rangle = 0.$$

Since this holds for all functions u, we conclude:

$$\partial_t p(t,x) = \mathcal{A}^* p(t,x).$$

□

Chapter 2

Rough Hawkes Process

For this first application, we consider the following jump-diffusion process, where memory is introduced in the jump frequency:

$$\frac{dS_t}{S_t} = \mu\, dt + \sigma\, dW_t + dL_t^{\text{comp}}, \tag{2.1}$$

where

$$L_t = \sum_{k=1}^{N_t} J_k, \quad J_k \sim J, \quad \text{and} \quad N_t \sim \text{Poi}(\lambda_t),$$

with intensity

$$\lambda_t = \lambda_0 + \zeta \int_0^t k(t-s)\, dN_s.$$

λ_t depends on the number of past jumps, whatever their severity. Another models could be consider, in which jumps are self-exciting through dL_s instead of dN_s. In that case, intensity is self-excited not only through past jump occurence but also magnitude of past jumps.

The compensated jump process is designed such as it has no contribution to the mean value of the process. In other words, to ensure zero mean of L_t^{comp} :

$$dL_t^{\text{comp}} = dL_t - \mathbb{E}(dL_t).$$

Since the jump process is a compounded process,

$$\mathbb{E}(dL_t) = \mathbb{E}(J) \times \mathbb{E}(dN_t) = \mathbb{E}(J) \times \lambda_t dt.$$

Therefore, the Equation (2.1) can be written as

$$\frac{dS_t}{S_t} = (\mu - \mathbb{E}(J)\lambda_t)dt + \sigma dW_t + dL_t.$$

The reader should recognize the contribution of the compensator $C = \mathbb{E}(J)\lambda_t$ in the drift of the dynamic, as argue in the introduction.
Alternatively, using the convolution representation, we can express the intensity as

$$\lambda_t = \lambda_0 + \zeta(k * dN)_t.$$

All things considered, the system that we will study throughout next two chapters is

$$\begin{cases} dX_t = (\mu - \frac{\sigma^2}{2} - \lambda_t \mathbb{E}(J))dt + \sigma dW_t + dL_t \\ \lambda_t = \lambda_0 + \zeta(k \circledast dN)_t \end{cases}, \tag{2.2}$$

where $k(-)$ is specifically chosen as the dampened rough Hawkes kernel :

$$k(h) = \exp(-vh)\frac{h^{\alpha-1}}{\Gamma(\alpha)} \quad \text{for } \alpha \in (1/2, 1), v \in \mathbb{R}.$$

2.1 Procedural computation

The aim of this first section is to compute the theoretical closed-form expression for the moment generating function of the asset log-returns $X_t = \log \frac{S_t}{S_0}$. The reader is referred to all the considerations introduced in the first chapter. All the section provides the complete proof for the following result :

Moment generating function of the log-return in the dynamic (2.2)

$$\mathbb{E}(\exp(-wX_s) \mid \mathcal{F}_0) = \exp(-s[w(\mu - \frac{\sigma^2}{2}) - \frac{\sigma^2 w^2}{2}] + q_\lambda(s)\lambda_0 - wX_0)$$

where $q_\lambda(s)$ is solution of the following differential equation :

$$\partial_s q_\lambda(s) = w\mathbb{E}(J) + \mathbb{E}(\exp(-wJ))\exp(\zeta(K\frac{dq_\lambda}{ds})(s)) - 1 \tag{2.3}$$

and K is the Sonine operator defined in Chapter 1 (1.22):

$$\begin{aligned}\Psi(s) = (K\frac{dq_\lambda}{ds})(s) = (k * \frac{dq_\lambda}{ds})(s) &= \int_0^s k(s-u)\frac{dq_\lambda}{du}(u)du \\ &= \int_0^s \exp(-v(s-u)) \times \frac{(s-u)^{\alpha-1}}{\Gamma(\alpha)}\frac{dq_\lambda}{du}(u)du, \quad v \in \mathbb{R}\end{aligned} \tag{2.4}$$

Proof. The closed-form formula is obtained via the procedure presented in Chapter 1.

2.1.1 Step 1

$k(-)$ is the dampened version of the rough Hawkes kernel $g(-)$ (1.11) :

$$g(h) = \frac{h^{\alpha-1}}{\Gamma(\alpha)} \quad \text{for } \alpha \in (1/2, 1).$$

This function $g(-)$ has a spectral representation via Laplace transform, corresponding to the Fourier transform of its dampened version $k(-)$, then we can embed all the memory of the system in an infinite number of new state variables, as argued in the first chapter. This memory is represented by the continuous process $\{Y_t(u)\}_{u\in\mathbb{R}_+}$. However, because the kernel has a pole at $h = 0$, we consider the unilateral transform. This is why $Y_t(u)$ is defined on $\mathbb{R}_+$.

2.1.2 Step 2

After the unilateral finite-dimensional approximation procedure (See the procedure in previous chapter 1.2), we obtain the following n-dimensional Markovian system :

$$\begin{cases} d\tilde{X}_t = (\mu - \frac{\sigma^2}{2} - \tilde{\lambda}_t\mathbb{E}(J)) + \sigma dW_t + d\tilde{L}_t \\ d\tilde{\lambda}_t = \zeta\sum_{l=0}^{n-1} w_l d\tilde{Y}_t(l) \\ d\tilde{Y}_t(l) = -(v + b_l)\tilde{Y}_t(l)dt + d\tilde{N}_t \end{cases} . \tag{2.5}$$

2.1.3 Step 3

We apply the generalized Itô's lemma to a function $f(t, \tilde{Z}_t)$ defined in the finite-dimensional Markovian system :

$$\begin{aligned} df(t,\widetilde{Z}_t) = {} & \partial_t f(t,\widetilde{Z}_t)\,dt + \partial_{\widetilde{X}} f(t,\widetilde{Z}_t)\,d\widetilde{X}_t + \partial_{\widetilde{\lambda}} f(t,\widetilde{Z}_t)\,d\widetilde{\lambda}_t \\ & + \sum_{l=0}^{n-1} \partial_{\widetilde{Y}(l)} f(t,\widetilde{Z}_t)\,d\widetilde{Y}_t(l) \\ & + \frac{\sigma^2}{2}\partial^2_{\widetilde{X}\widetilde{X}} f(t,\widetilde{Z}_t)\,dt \\ & + \Big[f\big(t,\widetilde{Z}_t + \Delta\widetilde{Z}_t\big) - f(t,\widetilde{Z}_t)\Big]\,d\widetilde{N}_t. \end{aligned} \tag{2.6}$$

where

- $\tilde{Z}_t = (\tilde{X}_t, \tilde{\lambda}_t, \tilde{Y}_t(l))$ is the vector of states in the system;
- $\Delta\tilde{Z}_t = (\Delta\tilde{X}_t, \Delta\tilde{\lambda}_t, \Delta\tilde{Y}_t(l)) = (+z, +\zeta\sum_{l=0}^{n-1} w_l, +1)$ is a vector of shocks following a jump.

The infinitesimal generator can be directly obtained via

$$\begin{aligned} \mathbb{E}(\frac{df}{dt} \mid \mathcal{F}_t) = {} & \partial_t f(t,\widetilde{Z}_t) + \partial_{\widetilde{X}} f(t,\widetilde{Z}_t)\,\mathbb{E}(\frac{d\widetilde{X}_t}{dt} \mid \mathcal{F}_t) + \partial_{\widetilde{\lambda}} f(t,\widetilde{Z}_t)\,\mathbb{E}(\frac{d\widetilde{\lambda}_t}{dt} \mid \mathcal{F}_t) \\ & + \sum_{l=0}^{n-1} \partial_{\widetilde{Y}(l)} f(t,\widetilde{Z}_t)\,\mathbb{E}(\frac{d\widetilde{Y}_t(l)}{dt} \mid \mathcal{F}_t) \\ & + \frac{\sigma^2}{2}\partial^2_{\widetilde{X}\widetilde{X}} f(t,\widetilde{Z}_t) \\ & + \tilde{\lambda}_t \int_{\mathbb{R}_+} \Big[f\big(t,\widetilde{Z}_t + \Delta\widetilde{Z}_t\big) - f(t,\widetilde{Z}_t)\Big]\,\tilde{\nu}(dz). \end{aligned} \tag{2.7}$$

$\tilde{\nu}(-)$ is the density of the jump severity distribution in the approximated finite-dimensional basis. The expectations are given by the dynamics :

1. $\mathbb{E}(\frac{d\widetilde{X}_t}{dt} \mid \mathcal{F}_t) = \mu - \frac{\sigma^2}{2} - \tilde{\lambda}_t\mathbb{E}(J)$;
2. $\mathbb{E}(\frac{d\widetilde{\lambda}_t}{dt} \mid \mathcal{F}_t) = -\zeta\sum_{l=0}^{n-1} w_l(v + b_l)\tilde{Y}_t(l)$;
3. $\mathbb{E}(\frac{d\widetilde{Y}_t(l)}{dt} \mid \mathcal{F}_t) = -(v + b_l)\tilde{Y}_t(l)$.

Now we have obtained the infinitesimal general, remind that the generator is $\mathcal{A}f(t, \tilde{Z}_t) = \mathbb{E}(\frac{df}{dt} \mid \mathcal{F}_t) - \partial_t f(t, \widetilde{Z}_t)$. We postulate an affine exponential structure of the form

$$u(t,x) = \exp\Big(q_0(t,T) + q_{\widetilde{\lambda}}(t,T)\,\widetilde{\lambda}_t + \sum_{l=0}^{n-1} q_l(t,T)\,\widetilde{Y}_t(l) - w\,x\Big). \tag{2.8}$$

The resolution of the backward Kolmogorov equation amounts to zeroing the infinitesimal generator:

$$\mathbb{E}(\frac{du}{dt} \mid \mathcal{F}_t) = \mathcal{A}u(t,x) + \partial_t u(t,x) = 0. \tag{2.9}$$

This is carried out by identifying terms and cancelling coefficients term by term. We first replace the affine parametrization (2.8) in the Kolmogorov equation (2.9). For the temporal derivative part, we obtain:

$$\partial_t u(t,x) = \Big[\partial_t q_0(t,T) + \partial_t q_{\widetilde{\lambda}}(t,T)\,\widetilde{\lambda}_t + \sum_{l=0}^{n-1} \partial_t q_l(t,T)\,\widetilde{Y}_t(l)\Big] u(t,x). \tag{2.10}$$

For the generator, we have:

$$\begin{aligned}
\mathcal{A}u(t,x) = {} & [-w\,u(t,x)]\Big[\mu - \frac{\sigma^2}{2} - \widetilde{\lambda}_t\,\mathbb{E}(J)\Big] \\
& + q_{\widetilde{\lambda}}(t,T)\,u(t,x)\Big[-\zeta \sum_{l=0}^{n-1} w_l(v+b_l)\,\widetilde{Y}_t(l)\Big] \\
& - \sum_{l=0}^{n-1} q_l(t,T)\,(v+b_l)\,u(t,x)\,\widetilde{Y}_t(l) \\
& + \frac{\sigma^2}{2} w^2\,u(t,x) \\
& + \widetilde{\lambda}_t \int_{\mathbb{R}_+} \Big[u\big(t,\widetilde{Z}_t + \Delta\widetilde{Z}_t\big) - u\big(t,\widetilde{Z}_t\big)\Big]\,\widetilde{\nu}(dz).
\end{aligned} \tag{2.11}$$

Firstly, we identify the constant terms (in blue) :

$$\partial_t q_0(t,T) - w(\mu - \frac{\sigma^2}{2}) + \frac{\sigma^2}{2} w^2 = 0 \tag{2.12}$$

$$\Rightarrow \partial_t q_0(t,T) = w(\mu - \frac{\sigma^2}{2}) - \frac{\sigma^2}{2} w^2 \tag{2.13}$$

Then, we identify the coefficients of $\tilde{\lambda}_t$ (in green):

$$\begin{aligned}
& \partial_t q_{\tilde{\lambda}}(t,T) + w\mathbb{E}(J) + \int_{\mathbb{R}_+} [u(t,\Delta\tilde{Z}_t) - 1]\tilde{\nu}(du) = 0 \\
\Rightarrow & \partial_t q_{\tilde{\lambda}}(t,T) + w\mathbb{E}(J) + \exp\Big(\sum_{l=0}^{n-1}(\zeta w_l q_{\tilde{\lambda}}(t,T) + q_l(t,T))\Big) \times \underbrace{\int_{\mathbb{R}_+} \exp(-wz)\tilde{\nu}(du)}_{\mathbb{E}(\exp(-wJ))} - 1 = 0
\end{aligned} \tag{2.14}$$

$$\Rightarrow \partial_t q_{\tilde{\lambda}}(t,T) = -w\mathbb{E}(J) - \exp\Big(\sum_{l=0}^{n-1}(\zeta w_l q_{\tilde{\lambda}}(t,T) + q_l(t,T))\mathbb{E}(\exp(-wJ)) + 1 \tag{2.15}$$

Thirdly, we identify the coefficient of each $\tilde{Y}_t(l)$:

$$\partial_t q_l(t,T) - (v+b_l) q_l(t,T) - \zeta w_l (v+b_l) q_{\tilde{\lambda}_t}(t,T) = 0 \tag{2.16}$$

$$\Rightarrow \partial_t q_l(t,T) = (v+b_l)[q_l(t,T) + \zeta w_l q_{\tilde{\lambda}_t}(t,T)] \tag{2.17}$$

We have now derived a system of ODEs that characterizes each of the parametrization functions defining the mgf of the returns. Finally, we substitute $[q_l(t,T) + \zeta w_l q_{\tilde{\lambda}_t}(t,T)]$ in (2.15) using the relation (2.17)

$$\frac{\partial_t q_l(t,T)}{v+b_l},$$

and we integrate this last Equation (2.17). This gives the following system of equations:

$$\begin{cases} q_l(t,T) = (v+b_l)\int_t^T \exp((v+b_l)(u-t))\zeta w_l q_{\tilde{\lambda}_t}(u,T)du \\ \partial_t q_{\tilde{\lambda}}(t,T) = -w\mathbb{E}(J) - \exp(\sum_{l=0}^{n-1} \frac{\partial_t q_l(t,T)}{v+b_l})\mathbb{E}(\exp(-wJ)) + 1 \\ \partial_t q_0(t,T) = w(\mu - \frac{\sigma^2}{2}) - \frac{\sigma^2}{2}w^2 \end{cases}$$

In conclusion, the backward Kolmogorov equation lead to a system of Ricatti equations. However, this is not possible to compute the limit of the partition in the rough Hawkes setting because of the singularity at the origin. In consequence,

$$\lim_{n\to\infty} \sum_{l=0}^{n-1} w_l = \infty.$$

Fortunately, forward equations solve that problem. The singularity does not introduce any problem in term of density.

2.1.4 Step 4

Hainaut [3] has already proved that

1. parametrization functions are stationary : $q_*(t,s) = q_*(s-t)$;
2. $\partial_s q_*(t,s) = -\partial_t q_*(t,s)$.

Especially, the second relationship is essential to invert the Kolmogorov equation into its forward form :

(1) Equation 2.13 becomes:

$$\partial_s q_0(t,s) = -w(\mu - \frac{\sigma^2}{2}) + \frac{\sigma^2}{2}w^2.$$

After integration, we have

$$q_0(t,s) = (s-t)\Big[-w(\mu - \frac{\sigma^2}{2}) + \frac{\sigma^2}{2}w^2\Big].$$

(2) Equation 2.15 becomes:

$$\partial_s q_{\tilde{\lambda}_t}(t,s) = w\mathbb{E}(J) + \exp(-\sum_{l=0}^{n-1} \frac{\partial_s q_l(t,s)}{v+b_l})\mathbb{E}(\exp(-wJ)) - 1$$

(3) Equation 2.17 becomes:

$$\begin{aligned} \partial_s q_l(t,s) &= -\int_t^s (v+b_l)\exp(-(v+b_l)(u-t))\zeta w_l \partial_s q_{\tilde{\lambda}_t}(s-u)du \\ &= -\int_0^{s-t} (v+b_l)\exp(-(v+b_l)(s-r-t))\zeta w_l \partial_r q_{\tilde{\lambda}_t}(r)dr \quad \text{with variable change } r = s-u. \end{aligned}$$

This second equation can be put in the first one :

$$\partial_s q_{\tilde{\lambda}_t}(t,s) = w\mathbb{E}(J)+\exp(\underbrace{\sum_{l=0}^{n-1}\int_0^{s-t}\exp(-(v+b_l)(s-r-t))\zeta w_l \partial_r q_{\tilde{\lambda}_t}(r)dr}_{A})\mathbb{E}(\exp(-wJ))-1 \tag{2.18}$$

All things done, we can recover the mgf of the return in the initial Markovian setting of infinite dimension by taking the limit :

$$\begin{aligned}\lim_{n\to\infty} A &= \int_0^{s-t}\int_{\mathbb{R}_+}\exp(-(v+\xi)(s-r-t))\mu(d\xi)\zeta\frac{dq_\lambda(r)}{dr}dr \\ &= \int_0^{s-t}\exp(-v(s-r-t))\int_0^{\infty}\exp(-\xi(s-r-t))\mu(d\xi)\zeta\frac{dq_\lambda(r)}{dr}dr \\ &= \int_0^{s-t}\frac{\exp(-v(s-r-t))}{(s-r-t)^{1-\alpha}}\zeta\frac{dq_\lambda(r)}{dr}dr = \zeta(K\frac{dq_\lambda}{ds})(s-t).\end{aligned} \tag{2.19}$$

This gives the result 2.1, reminded in the following box:

$$\mathbb{E}(\exp(-wX_s)\mid\mathcal{F}_0) = \exp(-s[w(\mu-\frac{\sigma^2}{2})-\frac{\sigma^2w^2}{2}]+q_\lambda(0,s)\lambda_0 - wX_0)$$

where $q_\lambda(0,s)$ is solution of the following differential equation :

$$\partial_s q_\lambda(0,s) = w\mathbb{E}(J)+\mathbb{E}(\exp(-wJ))\exp(\zeta(K\frac{dq_\lambda}{ds})(s))-1$$

and K is the Sonine operator defined in Chapter 1 :

$$\begin{aligned}\Psi(s) = (K\frac{dq_\lambda}{ds})(s) &= (k*\frac{dq_\lambda}{ds})(s) = \int_0^s k(s-u)\frac{dq_\lambda}{du}(u)du \\ &= \int_0^s \exp(-v(s-u))\times\frac{(s-u)^{\alpha-1}}{\Gamma(\alpha)}\frac{dq_\lambda}{du}(u)du\end{aligned} \tag{2.20}$$

□

There is still some remarks about this main results. Notably, if we define $\Psi(s) = (K\frac{dq_\lambda}{ds})(s)$, we can take advantage of the relation induced by Sonine operators 1.1:

$$\Psi(s) = (K\frac{dq_\lambda}{ds})(s) \Leftrightarrow K^{-1}\Psi(s) = \frac{dq_\lambda}{ds}(s) = \frac{d}{ds}(L\Psi)(s),$$

where

$$\begin{aligned}(L\Psi)(s) = (l*\Psi)(s) &= \int_0^s l(s-u)\Psi(u)du \\ &= \int_0^s\left[\beta^\alpha+\frac{\alpha}{\Gamma(1-\alpha)}\int_{s-u}^{\infty}\frac{\exp(-vr)}{r^{1+\alpha}}dr\right]\Psi(u)du.\end{aligned}$$

Equation 2.20 can then be written in term of Sonine operator (1.22):

$$\begin{aligned}K^{-1}\Psi(s) &= w\mathbb{E}(J)+\mathbb{E}(\exp(-wJ))\exp(\zeta\Psi(s))-1 \\ \Rightarrow \frac{d}{ds}\int_0^s(\beta^\alpha+\frac{\alpha}{\Gamma(1-\alpha)}\int_{s-u}^{\infty}\frac{\exp(-rv)}{r^{1+\alpha}}dr)\psi(u)du &= w\mathbb{E}(J)+\mathbb{E}(\exp(-wJ))\exp(\zeta\Psi(s))-1.\end{aligned}$$

Now we have obtained a closed-form expression for the mgf, the remaining work includes:

1. calibration of the parameters under $\mathbb{P}$;
2. apply a change of measure towards $\mathbb{Q}$;
3. solve numerically $(K\frac{dq_\lambda}{ds})(s)$, and apply DFFT algorithm in order to compute the Fourier transform.

Each of these remaining steps will be considered in the following sections, and calibration will be further developed through results presented in Chapter 3.

2.2 Calibration under real-world measure

We need to make a choice of distribution for the severity of the jumps, J and calibrate the parameters of the system via Maximum Likelihood Estimation (MLE).

- We consider the rough Hawkes intensity

$$\lambda_t = \lambda_0 + \zeta \int_0^{t^-} e^{-v(t-s)} \frac{(t-s)^{\alpha-1}}{\Gamma(\alpha)}\, dN_s,$$

or equivalently,

$$\lambda_t = \lambda_0 + \zeta \sum_{\tau_j < t} e^{-v(t-\tau_j)} \frac{(t-\tau_j)^{\alpha-1}}{\Gamma(\alpha)}.$$

The log-likelihood associated with the observed jump times $\{\tau_1, \dots, \tau_{N_T}\}$ over the interval $[0, T]$ is given by

$$\log f(T) = \sum_{k=1}^{N_T} \log(\lambda_{\tau_k^-}) - \int_0^T \lambda_s ds.$$

The intensity evaluated immediately before the jump times satisfies

$$\lambda_{\tau_k^-} = \lambda_0 + \zeta \sum_{j=1}^{k-1} e^{-v(\tau_k - \tau_j)} \frac{(\tau_k - \tau_j)^{\alpha-1}}{\Gamma(\alpha)}.$$

For the second term, we obtain

$$\begin{aligned}\int_0^T \lambda_u du &= \lambda_0 T + \zeta \int_0^T \int_0^{u^-} e^{-v(u-s)} \frac{(u-s)^{\alpha-1}}{\Gamma(\alpha)}\, dN_s\, du \\ &= \lambda_0 T + \zeta \int_0^T \int_s^T e^{-v(u-s)} \frac{(u-s)^{\alpha-1}}{\Gamma(\alpha)}\, du\, dN_s.\end{aligned}$$

Using the change of variable $x = u - s$, and the Gamma function identity, we finally obtain

$$\begin{aligned}\int_0^T \lambda_u du &= \lambda_0 T + \zeta \sum_{j=1}^{N_T} \int_0^{T-\tau_j} e^{-vx} \frac{x^{\alpha-1}}{\Gamma(\alpha)} dx. \\ &= \lambda_0 T + \zeta v^{-\alpha} \sum_{j=1}^{N_T} \left(1 - \frac{\Gamma(\alpha, v(T-\tau_j))}{\Gamma(\alpha)}\right).\end{aligned} \tag{2.21}$$

where we used the following Gamma relation:

$$\int_0^a e^{-vx} x^{\alpha-1} dx = v^{-\alpha} \left(\Gamma(\alpha) - \Gamma(\alpha, va) \right),$$

Therefore, the ready-to-code log-likelihood is

$$\log f(T \mid \theta_N) = \sum_{k=1}^{N_T} \log \left(\lambda_0 + \zeta \sum_{j=1}^{k-1} e^{-v(\tau_k - \tau_j)} \frac{(\tau_k - \tau_j)^{\alpha-1}}{\Gamma(\alpha)} \right) - \lambda_0 T - \zeta v^{-\alpha} \sum_{j=1}^{N_T} \left(1 - \frac{\Gamma(\alpha, v(T - \tau_j))}{\Gamma(\alpha)} \right).$$

The optimal parameters $\Theta_N = \{\hat{v}, \hat{\alpha}, \hat{\zeta}\}$ are computed via log-likelihood maximization:

$$\Theta_N = \arg\max_{\theta_N} \log(f(\tau \mid \theta_N)).$$

- For the severity processes $J^k \sim J$, several choices are possible. We first consider the exponential distribution:

$$J \sim \text{Exp}(\rho),$$

whose density is given by

$$\nu_1(x) = \rho e^{-\rho x} \mathbf{1}_{x>0}.$$

Hence,

$$\log \nu_1(x) = \log(\rho) - \rho x.$$

- Since the exponential distribution only produces positive jumps, it is not suitable for modelling the negative jumps observed in log-returns. A natural extension is the double exponential distribution:

$$J \sim \text{DoubleExp}(p, \eta_+, \eta_-).$$

This model generates positive jumps with probability p, and negative jumps with probability $1 - p$. Its density is given by

$$\nu_2(x) = p\eta_+ e^{-\eta_+ x} \mathbf{1}_{x>0} + (1-p)\eta_- e^{\eta_- x} \mathbf{1}_{x<0}.$$

Consequently, the log-density satisfies

$$\log \nu_2(x) = (\log p + \log \eta_+ - \eta_+ x) \mathbf{1}_{x>0} + (\log(1-p) + \log \eta_- + \eta_- x) \mathbf{1}_{x<0}.$$

The optimal parameters are obtained through maximum likelihood estimation:

$$\Theta_J = \arg\max_{\{p, \eta_+, \eta_-\}} \sum_i \log \left(\nu_2(x_i \mid \eta_+, \eta_-) \right).$$

- Finally, we consider a double exponential distribution observed with Gaussian noise:

$$J = X + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2), \qquad X \sim \text{DoubleExp}(p, \eta_+, \eta_-).$$

This setting is suitable for future calibration with peak-over-threshold method discussed in the next Chapter 4. In this case, the jump density is obtained through convolution:

$$\nu_3(x) = \int_{\mathbb{R}} \nu_2(y)\phi(x-y)\,dy,$$

where ϕ denotes the Gaussian density function.

After explicit computation, one obtains

$$\begin{aligned}\nu_3(x) = p\eta_+ \exp\left(\frac{1}{2}\eta_+^2\sigma^2 - \eta_+ x\right)\Phi\left(\frac{x-\eta_+\sigma^2}{\sigma}\right) \\ + (1-p)\eta_- \exp\left(\frac{1}{2}\eta_-^2\sigma^2 + \eta_- x\right)\left[1-\Phi\left(\frac{x+\eta_-\sigma^2}{\sigma}\right)\right],\end{aligned}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard Gaussian distribution.

To compare the exponential and double exponential models, we use a likelihood ratio test:

$$\Lambda = 2\left(\log L_{\text{DoubleExp}} - \log L_{\text{Exp}}\right)$$

Under the null hypothesis that the exponential model is sufficient, the statistic follows asymptotically a central chi-squared distribution of dimension corresponding to the difference of parameters between the more complex double exponential model and the exponential one. Hence,

$$\Lambda \sim \chi^2(1), \tag{2.22}$$

since the double exponential model introduces one additional η_-. If Λ exceeds the 5 % quantile of the $\chi^2(1)$ distribution, the exponential model is rejected in favour of the double exponential model. This kind of test for model comparison will be extensively used in the next Chapter 3 for model selection.

2.3 Change of measure

This section has the objective to detect how parameters calibrated under real-measure $\mathbb{P}$ should be modified under the risk-neutral measure. In particular, the concluding result is:

Dynamics under Q

There is a family of risk-neutral measures $\mathbb{Q}^\gamma$ equivalent to $\mathbb{P}$, for $\gamma \in \mathbb{R}$ (in practice, the domain of γ is restricted by values for which the mgf of $J^{\mathbb{Q}}$ exists), such that

$$\mathbb{E}^{\mathbb{Q}}(\exp(-wJ^{\mathbb{Q}})) = \frac{\mathbb{E}(\exp((\gamma-w)J))}{\mathbb{E}(\exp(J\gamma))},\ \lambda_t^{\mathbb{Q}} = \mathbb{E}(\exp(J\gamma))\lambda_t$$

Proof. Again, the whole section provides the proof of this result. It mainly consists of repeating previous computations, considering the generalized Doleans-Dale adjustment.

We fall back to the step after our approximation procedure, after which we obtained the approximated ('tilde' version) dynamic:

$$\tilde{X}_t = (\mu - \frac{\sigma^2}{2})t + \sigma W_t + \tilde{L}_t - \mathbb{E}(J)\int_0^t \tilde{\lambda}_s ds.$$

The generalized Doleans-Dale exponential is

$$\eta_t = \frac{d\mathbb{Q}^\gamma}{d\mathbb{P}}\Big|_t = \exp(-\frac{1}{2}\int_0^t \phi(s)^2 ds - \int_0^t \phi(s) dW_s) \times \exp(\gamma \tilde{L}_t + (1 - \mathbb{E}(\exp(J\gamma))) \int_0^t \tilde{\lambda}_s ds) \tag{2.23}$$

where

$$\begin{cases} \eta_t \text{ is a local martingale} : \mathbb{E}(d\eta_t) = 0 \\ \gamma \in \mathbb{R} \text{ such that } \mathbb{E}(\exp(J\gamma)) < \infty \\ \phi(s) \text{ is a } \mathcal{F}_t\text{-adapted process} : \int_0^t \mid \phi(s) \mid^2 ds < \infty \end{cases}.$$

In consequence, the differential formulation of η_t is

$$\begin{aligned} d\eta_t &= -\eta_t \phi(t) dW_t + \eta_t \gamma \tilde{L}_t + \eta_t (1 - \mathbb{E}(\exp(J\gamma))) \underbrace{\tilde{\lambda}_t dt}_{d\tilde{N}_t} \\ &= d\eta_t^{Brownian} + d\eta_t^{Jump} \end{aligned} \tag{2.24}$$

We observe that previous equations (2.24) and (2.23) are an extension of those obtained for diffusion. In fact, the reader has recognized the diffusion term and the jump part in the DD exponential (2.23). We take advantage of the separability of the change of measure. On one hand, it is well-known that the diffusion part leads to a risk-neutral Brownian motion

$$dW_t^{\mathbb{Q}^\gamma} = dW_t + \phi^B(t) dt.$$

The diffusion part of $\mathrm{d}\tilde{X}_t$ becomes then

$$[\mu - \frac{\sigma^2}{2} - \sigma\phi^B(t)]dt + \sigma dW_t^{\mathbb{Q}^\gamma}. \tag{2.25}$$

Consequently, since all derivatives grows at the risk-free rate under $\mathbb{Q}$:

$$\mu - \sigma\phi^B(t) = r \Rightarrow \phi^B(t) = \frac{\mu - r}{\sigma}.$$

On the other hand, we just consider the jump part in (2.24):

$$d\eta_t^{Jump} = \eta_t^{Jump} \gamma d\tilde{L}_t + \eta_t^{Jump} (1 - \mathbb{E}(\exp(J\gamma))) \tilde{\lambda}_t dt \tag{2.26}$$

We define $Z_t := \log(\eta_t^{Jump})$. Since we have from the jump part in (2.23):

$$\eta_t^{Jump} = \exp(\gamma \tilde{L}_t + (1 - \mathbb{E}(\exp(J\gamma)) \int_0^t \tilde{\lambda}_s ds),$$

then the dynamic of Z_t can be directly deduced from the one of η_t:

$$Z_t = \gamma \tilde{L}_t + (1 - \mathbb{E}(\exp(J\gamma)) \int_0^t \tilde{\lambda}_s ds \Rightarrow dZ_t = \gamma d\tilde{L}_t + (1 - \mathbb{E}(\exp(J\gamma)) \tilde{\lambda}_t dt.$$

Essentially, we process the same way than for classical diffusion case. We consider the change of measure fundamental theorem :

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}^\gamma}(\exp(-w\tilde{X}_s) \mid \mathcal{F}_t) &= \frac{\mathbb{E}(\exp(-w\tilde{X}_s) \frac{d\mathbb{Q}^\gamma}{d\mathbb{P}}\big|_s \mid \mathcal{F}_t)}{\frac{d\mathbb{Q}^\gamma}{d\mathbb{P}}\big|_t} \\ &= \exp(-Z_t)\mathbb{E}(\exp(-w\tilde{X}_s + Z_s) \mid \mathcal{F}_t); \end{aligned}$$

and we will apply the generalized Itô's lemma to

$$f(t, \tilde{X}_t, \tilde{\lambda}_t, \tilde{Y}_t(l), \tilde{L}_t, Z_t) = \mathbb{E}(\exp(-w\tilde{X}_s + Z_s) \mid \mathcal{F}_t).$$

The following computations are the same than applying Steps 3-4 of our previous procedure. By the generalized Itô's lemma, we can compute the infinitesimal generator:

$$\begin{aligned}\mathbb{E}(\frac{df}{dt} \mid \mathcal{F}_t) = \partial_t f + \left(r - \frac{\sigma^2}{2} - \mathbb{E}(J)\tilde{\lambda}_t\right)\partial_{\tilde{X}} f + \frac{\sigma^2}{2}\partial^2_{\tilde{X}\tilde{X}} f \\ + (1 - \mathbb{E}(\exp(J\gamma)))\tilde{\lambda}_t \partial_Z f \\ - \sum_{l=0}^{n-1}(v + b_l)Y_t(l)\,\partial_{Y_t(l)} f \\ + \partial_{\tilde{\lambda}} f \left[\lambda_0 - \sum_{l=0}^{n-1}(v + b_l)w_l\zeta Y_t(l)\right] \\ + \tilde{\lambda}_t \int_{\mathbb{R}} \Big(f\big(t,\, \tilde{\lambda}_t + \sum_{l=0}^{n-1} w_l\zeta\mathbb{E}(\exp(J\gamma)),\, \tilde{X}_t + z,\, Z_t + \gamma z,\, Y_t(l) + 1\big) - f(\cdot)\Big)\tilde{\nu}(dz).\end{aligned} \tag{2.27}$$

We introduce the exponential affine parametrization:

$$f(t, \tilde{X}_t, \tilde{\lambda}_t, \tilde{Y}_t(l), \tilde{L}_t, Z_t) = \exp(q_0(t,s) + q_{\tilde{\lambda}}(t,s)\tilde{\lambda}_t\mathbb{E}(\exp(J\gamma)) + \sum_{l=0}^{n-1} w_l q_l(t,s)Y_t(l) - w\tilde{X}_t + q_Z(t,s)Z_t).$$

In particular, the last integral in (2.27) can be further simplified, using the distribution property $\int_{\mathbb{R}} \tilde{\nu}(dz) = 1$:

$$\begin{aligned}\int_{\mathbb{R}} \Big(f\big(t, \tilde{\lambda}_t + \sum_{l=0}^{n-1} w_l\zeta\mathbb{E}(\exp(J\gamma)),\, \tilde{X}_t + z,\, Z_t + \gamma z,\, Y_t(l) + 1\big) - f(\cdot)\Big)\tilde{\nu}(dz) \\ = \int_{\mathbb{R}} f(.)\Big(\exp(\sum_{l=0}^{n-1} w_l[\zeta\mathbb{E}(\exp(J\gamma))q_{\tilde{\lambda}}(t,s) + q_l(t,s)] - wz + \gamma z q_Z(t,s)) - 1\Big)\tilde{\nu}(dz) \\ = f(.)\left[\int_{\mathbb{R}} \exp(\sum_{l=0}^{n-1} w_l[\zeta\mathbb{E}(\exp(J\gamma))q_{\tilde{\lambda}}(t,s) + q_l(t,s)] - wz + \gamma z q_Z(t,s))\tilde{\nu}(dz) - 1\right].\end{aligned} \tag{2.28}$$

The closed-form affine exponential provides expression for each derivative:

$$\begin{cases}\partial_{\tilde{X}} f = -wf \\ \partial^2_{\tilde{X}\tilde{X}} f = w^2 f \\ \partial_Y f = w_l q_l(t,s) f \\ \partial_{\tilde{\lambda}} f = q_\lambda(t,s)\mathbb{E}(\exp(J\gamma))f \\ \partial_Z f = q_Z(t,s)f \\ \partial_t f = f[\partial_t q_0(t,s) + \partial_t q_{\tilde{\lambda}}(t,s)\tilde{\lambda}_t\mathbb{E}(\exp(J\gamma)) + Z_t\partial_t q_Z(t,s) + \sum_{l=0}^{n-1} w_l\partial_t q_l(t,s)Y_t(l)].\end{cases}.$$

We put all together and we identify the coefficients of each variable to the coefficients of the zero polynomial. This gives the following system of Fokker-Planck equations:

- For the variable Z_t:
$$\partial_t q_Z(t,s) = 0 \Rightarrow q_Z(t,s) = 1. \tag{2.29}$$
- For the variable $\tilde{\lambda}_t$:
$$\begin{aligned}&\partial_t q_{\tilde{\lambda}}(t,s)\mathbb{E}(\exp(\gamma J)) + w\mathbb{E}(J) + (1 - \mathbb{E}(\exp(J\gamma)))q_Z(t,s) + \\ &\exp(\sum_{l=0}^{n-1} w_l\zeta\mathbb{E}(\exp(J\gamma))q_{\tilde{\lambda}}(t,s) + q_l(t,s)) \int_{\mathbb{R}_+} \exp(-wz + \gamma z q_Z(t,s))\tilde{\nu}(dz) - 1 = 0.\end{aligned} \tag{2.30}$$

We substitute $q_Z(t,s)$ by 1 in Equation 2.29:

$$\Rightarrow \partial_t q_{\tilde{\lambda}}(t,s)\mathbb{E}(\exp(\gamma J)) + w\mathbb{E}(J) - \mathbb{E}(\exp(J\gamma)) + \exp(\sum_{l=0}^{n-1} w_l\zeta\mathbb{E}(\exp(J\zeta))q_{\tilde{\lambda}}(t,s) + q_l(t,s))$$
$$\underbrace{\int_{\mathbb{R}_+} \exp(-wz+\gamma z)\tilde{\nu}(dz)}_{\mathbb{E}(\exp(J(\gamma-w))} = 0. \tag{2.31}$$

$$\Rightarrow \partial_t q_{\tilde{\lambda}}(t,s)\mathbb{E}(\exp(\gamma J)) = -w\mathbb{E}(J) + \mathbb{E}(\exp(J\gamma)) - \exp(\sum_{l=0}^{n-1} w_l\zeta\mathbb{E}(\exp(J\gamma))q_{\tilde{\lambda}}(t,s) + q_l(t,s)) \times \mathbb{E}(\exp(J(\gamma-w)) \tag{2.32}$$

$$\Rightarrow \partial_t q_{\tilde{\lambda}}(t,s) = -w\frac{\mathbb{E}(J)}{\mathbb{E}(\exp(\gamma J))} + 1 - \exp(\sum_{l=0}^{n-1} w_l\zeta\mathbb{E}(\exp(J\gamma))q_{\tilde{\lambda}}(t,s) + q_l(t,s))\frac{\mathbb{E}(\exp(J(\gamma-w))}{\mathbb{E}(\exp(\gamma J))} \tag{2.33}$$

- for each variable $Y_t(l)$:

$$\frac{w_l}{\Gamma(\alpha)}\partial_t q_l(t,s) - \frac{w_l(v+b_l)}{\Gamma(\alpha)}q_l(t,s) - q_{\tilde{\lambda}}(t,s)\mathbb{E}(\exp(J\gamma))(v+b_l)\frac{w_l\zeta}{\Gamma(\alpha)} = 0 \tag{2.34}$$

$$\Rightarrow \partial_t q_l(t,s) = (v+b_l)[q_l(t,s) + \zeta q_{\tilde{\lambda}}(t,s)\mathbb{E}(\exp(J\gamma))] \tag{2.35}$$

- Finally, for the constants terms:

$$\partial_t q_0(t,s) = w(r - \frac{\sigma^2}{2}) - \frac{\sigma^2 w^2}{2}. \tag{2.36}$$

Comparing the system of equations (2.36), (2.35) and (2.33) we have under risk-neutral measure $\mathbb{Q}^\gamma$

$$\begin{cases} \partial_t q_0(t,s) = w(r-\frac{\sigma^2}{2}) - \frac{\sigma^2 w^2}{2} \\ \partial_t q_l(t,s) = (v+b_l)[q_l(t,s) + \zeta q_{\tilde{\lambda}}(t,s)\mathbb{E}(\exp(J\gamma))] \\ \partial_t q_{\tilde{\lambda}}(t,s) = -w\frac{\mathbb{E}(J)}{\mathbb{E}(\exp(\gamma J))} + 1 - \exp(\sum_{l=0}^{n-1} w_l\zeta\mathbb{E}(\exp(J\gamma))q_{\tilde{\lambda}}(t,s) + q_l(t,s))\frac{\mathbb{E}(\exp(J(\gamma-w))}{\mathbb{E}(\exp(\gamma J))} \end{cases}$$

with the one we obtained earlier under $\mathbb{P}$, we can conclude the new dynamics of our system :

1. under $\mathbb{Q}^\gamma$, all assets grow at risk-free rate : $\mu \leftrightarrow r$;
2. the distribution of the jump severity is changed :
$$\mathbb{E}^{\mathbb{Q}}(\exp(-wJ^{\mathbb{Q}^\gamma})) = \frac{\mathbb{E}(\exp((\gamma-w)J))}{\mathbb{E}(\exp(J\gamma))},$$
3. Finally, $\lambda_t^{\mathbb{Q}^\gamma} = \mathbb{E}(\exp(J\gamma))\lambda_t$.

□

In consequence, if $J \sim \text{Exp}(\rho)$, we have a closed-form mgf $\mathbb{E}(\exp(Jw)) = \frac{\rho}{\rho-w}$ under $\mathbb{P}$, hence under $\mathbb{Q}^\gamma$, we have

$$\mathbb{E}^{\mathbb{Q}^\gamma}(\exp(J^{\mathbb{Q}^\gamma}w)) = \frac{\mathbb{E}(\exp((w+\gamma)J))}{\mathbb{E}(\exp(J\gamma))} = \frac{\rho}{\rho-w-\gamma}\frac{\rho-\gamma}{\rho} = \frac{\rho-\gamma}{\rho-\gamma-w}.$$

This means that $J^{\mathbb{Q}^\gamma} \sim \text{Exp}(\rho^{\mathbb{Q}^\gamma} = \rho - \gamma)$. To ensure existence of the mgf, the constraint $\gamma \in [-\rho, \infty)$ must be fulfilled, as shown in [1].

Similarly, if $J \sim \text{DoubleExp}(\eta_+, \eta_-, p)$, the moment generating function under $\mathbb{P}$ is given by

$$\mathbb{E}\left(e^{wJ}\right) = \frac{p\eta_+}{\eta_+ - w} + \frac{(1-p)\eta_-}{\eta_- + w}, \quad w \in [-\eta_-, \eta_+].$$

Under the change of measure $\mathbb{Q}$, we obtain

$$\mathbb{E}^{\mathbb{Q}^\gamma}\left(e^{wJ^{\mathbb{Q}^\gamma}}\right) = \frac{\mathbb{E}\left(e^{(w+\gamma)J}\right)}{\mathbb{E}\left(e^{\gamma J}\right)} = \frac{\frac{p\eta_+}{\eta_+ - (w+\gamma)} + \frac{(1-p)\eta_-}{\eta_- + (w+\gamma)}}{\frac{p\eta_+}{\eta_+ - \gamma} + \frac{(1-p)\eta_-}{\eta_- + \gamma}}.$$

Each negative and positive exponential structures are preserved, while the relative weighting between upward and downward jumps is modified by parameter $\gamma \in [-\eta_-, \eta_+]$.

We observe well than, in both cases, the interval of definition of γ includes $\gamma = 0$. In the family of risk-neutral measure $\mathbb{Q}^\gamma$, we could make the choice to use $\mathbb{Q}^0$, so that change of measure only impacts diffusive part of the dynamic.

Finally, it has already been proven by Hainaut in [1] that the family of change of measure is characterized by

$$\phi(t) = \frac{\mu - r}{\sigma} + \frac{1}{\sigma}\Big[\lambda_t \mathbb{E}(\exp(J\gamma))[\mathbb{E}(\exp(J^{\mathbb{Q}^\gamma})) - 1] - \mathbb{E}(J)\int_0^t \lambda_u du\Big].$$

Proof. According to previous considerations, the dynamic of the stock price under $\mathbb{Q}^\gamma$ is

$$d\tilde{X}_t = (\mu - \frac{\sigma^2}{2} - \sigma\phi(t))dt + \sigma dW_t^{\mathbb{Q}^\gamma} + d\tilde{L}_t^{\mathbb{Q}^\gamma} - \mathbb{E}^{\mathbb{Q}^\gamma}(J^{\mathbb{Q}^\gamma})\int_0^t \tilde{\lambda}_s ds.$$

We replace the expression of $\phi(t)$ in the previous equation, and observe that we obtain

$$\mathbb{E}^{\mathbb{Q}^\gamma}(\frac{dS_t}{dt} \mid \mathcal{F}_0) = rdt.$$

□

2.4 DFFT algorithm

In this last section, we present how the Fourier transform is handled to recover the expression of the pdf. We remind the relation between the mgf and the pdf that I have already highlighted in (1.1):

$$\mathcal{M}_X(z) = \int_{\mathbb{R}} \exp(izx) f(x) dx = \mathcal{F}^{-1}(2\pi f) \Leftrightarrow f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \exp(-izx)\mathcal{M}_X(z)dz = \frac{1}{2\pi}\mathcal{F}(\mathcal{M}_X(z)),$$

where $\mathcal{M}_X(u) = \mathbb{E}(\exp(iuX_t) \mid \mathcal{F}_0)$. Actually, the link can also be viewed on the other way,

$$f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \mathcal{M}_X(-z)\exp(izx)dz = \mathcal{F}^{-1}(\phi(iz)) \Leftrightarrow \mathcal{M}_X(-z) = \int_{\mathbb{R}} f(x)\exp(-izx)dx = \mathcal{F}(f).$$

In other words, in order to recover the pdf $f(-)$, we can either:

- apply a Fourier transform to the moment generating function, with $w = iz$, and scale it by a factor $\frac{1}{2\pi}$;
- apply an inverse Fourier transform to the moment generating function, with $w = -iz$.

The following explanations and summary map 2.4 highlight the dual links between the space in which $f(-)$ is defined and the one of $\mathcal{M}_X(-)$. Numerically, we cannot handle a continuous integral, then a meaningful approximation of the Fourier transform, using finite storage in memory, must be used. Intuitively, this is done by sampling the pdf, that is : we apply a train of Dirac impulse to the function, in order to isolate the values at constant interval, corresponding to the sampling period T_e.

In the dual space, the sampling has a periodization effect. This can be observed because :

1. the Fourier transform is a linear operator, then

$$\mathcal{F}(f_e(x)) = \mathcal{F}(\sum_{n\in\mathbb{Z}} f(x)\delta(x - nT_e)) = \sum_{n\in\mathbb{Z}} \mathcal{F}(f(x)\delta(x - nT_e))$$

2. By duality, the multiplication in x-axis gives a convolution in z-space : this property has already been used when considering Sonine operators 1.1.

3. the Dirac delta is actually the neutral element of the convolution. Think about what the convolution consists : we infinitely sum the product between the first function and all translations of the other one. In case we translate a Dirac delta, this is equivalent to continuously sampling the first function, which has no effect but recovering the function itself.

Consequently, sampling of $f(-)$ amounts to a convolution of $\mathcal{M}_X(-)$ with a train of Dirac impulse, which make the signal periodic. After sampling/periodization dual operation, a simple change of domain scale must be applied : from the continuously Fourier transform, we obtain this way the discrete Fourier transform.

However, this one still implies a infinite sum over integer values. Therefore, a truncature of the sum to ,M terms is applied. Actually, this step may seems not consistent but this is the motivation behind the DFT transform. Since $\mathcal{M}_X(z)$ have been periodized by the sampling of f(x), the approximation may still be exact in case of the period corresponds to M. If the reader has some notion in spectral transform, he could understand the DFT as the Discrete Tranform Fourier Serie (DTFS) applied to $f[n]$ truncated to M samples and M-periodized, up to a factor M. Remind that, for a function $f[n]$ periodic of period M,

$$\mathcal{M}_{DTFS}[k] = \frac{1}{M}\sum_{n=0}^{M-1} f[n]\exp(-j\Omega_0 nk), \quad \mathcal{M}_{DFT}[k] = \sum_{n=0}^{M-1} f[n]\exp(-\frac{2k\pi jn}{M}).$$

Therefore, for a non-periodic $f[n]$, applying DFT corresponds to $DTFS$ on its version periodized of period M and truncated to one period.

$$
\begin{array}{cc}
f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \mathcal{M}_X(z)\exp(izx)dz & \mathcal{M}_X(z) = \int_{\mathbb{R}} f(x)\exp(-izx)dx \\
= \mathcal{F}^{-1}(\mathcal{M}_X(z)) & = \mathcal{F}(f(x)) \\
\Downarrow \text{sampling} & \Downarrow \text{periodization} \\
f_e(x) = \sum_{n\in\mathbb{Z}} f(x)\delta(x - nT_e) & \mathcal{M}_{X,e}(z) = \mathcal{M}_X(z) * \frac{1}{T_e}\sum_{k\in\mathbb{Z}} \delta(w - \frac{2\pi k}{T_e}) \\
\Downarrow x = nT_e & \Downarrow \Omega = wT_e \\
f[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathcal{M}_X(\exp(j\Omega))\exp(j\Omega n)d\Omega & \mathcal{M}_X(\exp(j\Omega)) = \sum_{n\in\mathbb{Z}} f[n]\exp(-j\Omega n) \\
\Downarrow \text{truncature} & \Downarrow \text{truncature} \\
f[n] \approx \sum_{k=0}^{M-1} \mathcal{M}_{X,DFT}[k]\exp(jk\Omega_0 n) & \mathcal{M}_{X,DFT}[k] = \sum_{n=0}^{M-1} f[n]\exp(\frac{-2\pi jk}{M}n)
\end{array}
$$

Overall, we have

$$\mathcal{M}_X(\exp(j\Omega)) \approx \mathcal{M}_{X,DFT}[k] = \sum_{n=0}^{M-1} f[n]\exp(\frac{-2\pi jk}{M}n) = \sum_{n=0}^{M-1} f[n]w^n.$$

This is exactly a polynomial approximation problem on $\mathbb{C}$, where $w = \exp(-ix_k)$, and x_k are the equispaced points at which we take the approximation:

$$x_k = \frac{2\pi k}{M}.$$

This gives a linear system $Wf = \Phi$:

$$\underbrace{\begin{pmatrix} 1 & w_0 & w_0^2 & \cdots & w_0^{M-1} \\ \vdots & \ddots & & & \vdots \\ 1 & w_{M-1} & w_{M-1}^2 & \cdots & w_{M-1}^{M-1} \end{pmatrix}}_{\text{Vandermonde matrix W}} \begin{pmatrix} f[0] \\ f[1] \\ \vdots \\ f[M-1] \end{pmatrix} = \begin{pmatrix} \mathcal{M}_X[0] \\ \mathcal{M}_X[1] \\ \vdots \\ \mathcal{M}_X[M-1] \end{pmatrix} \tag{2.37}$$

One can easily prove that $\frac{1}{M}W^*W$ is an unitary matrix, and then the pdf samples f can be computed without matrix inversion, but using the Hermitian matrix W^*:

$$W^*Wf = W^*\Phi \Leftrightarrow Mf = W^*\Phi \Leftrightarrow f = \frac{1}{M}W^*\Phi \Rightarrow f[k] = \frac{1}{M}\sum_{n=0}^{M-1} \mathcal{M}_X[n]\exp(\frac{2\pi kjn}{M}).$$

At this stage, I have shown that the samples of the pdf can be retrieve by inverting a system, at a cost of $\mathcal{O}(n^3)$. Because the Vandermonde matrix defining the system is unitary, no inversion is needed. That makes the cost drop to $\mathcal{O}(n^2)$. However, FFT divide-and-conqueer algorithm reaches same result with a complexity of $\mathcal{O}(n\log n)$, using a wise split between even and odd interpolation points x_{2h} and x_{2h+1}. Actually, this algorithm is the one used in Python `scipy.fft` function. The divide-and-conqueer use

the following recurrence :

$$\begin{aligned} Mf[k] &= \sum_{h=0}^{M/2+1} \mathcal{M}_{X,DFT}[2h](\exp(\frac{2\pi ik}{M}))^{2h} + \sum_{h=0}^{M/2+1} \mathcal{M}_{X,DFT}[2h+1](\exp(\frac{2\pi ik}{M}))^{2h+1} \\ &= \underbrace{\sum_{h=0}^{M/2+1} \mathcal{M}_{X,DFT}[2h](\exp(\frac{2\pi ik}{M/2}))^{h}}_{\text{even part } E[k]} + \exp(\frac{2\pi ik}{M}) \underbrace{\sum_{h=0}^{M/2+1} \mathcal{M}_{X,DFT}[2h+1](\exp(\frac{2\pi ik}{M/2}))^{h}}_{\text{odd part } O[k]}. \end{aligned} \tag{2.38}$$

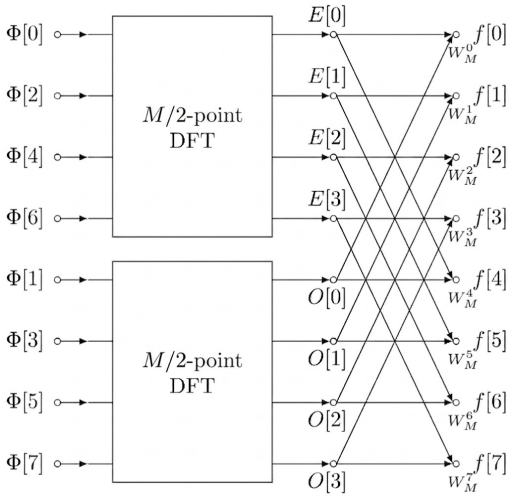

Figure 2.1: Explanation of the recurrence used by the Divide-and-Conqueer algorithm Fast Fourier Transform. Inspired by the graph shown in Wikipedia.

Now, we take a step back: before applying FFT, we need to find a way to solve the Volterra differential equation in (2.1): the last unclosed formula necessary for the computation of the MGF. We use an explicit approximation of $(K\frac{dq_\lambda}{ds})_s$. We subdivide the interval $[0, s]$ into n subintervals $[s_k, s_{k+1}]$ of length Δ:

$$(K\frac{dq_\lambda}{ds})_s = \int_0^s \exp(-h(s-u))\frac{(s-u)^{\alpha-1}}{\Gamma(\alpha)}\frac{dq_\lambda}{du}(u)du \approx \frac{1}{\Gamma(\alpha)}\sum_{u=0}^{k-1}\exp(-h(s_k-s_u))(s_k-s_u)^{\alpha-1}g(u)\Delta \tag{2.39}$$

where $g(u) \approx \frac{dq_\lambda(u)}{du}$ is solved iteratively. The implicit equation (2.3) is computed from the initial condition

$$g(0) = w\mathbb{E}(J) + \mathbb{E}(\exp(-wJ)) - 1; \tag{2.40}$$

and then through the following forward recurrence

$$g(k) = w\mathbb{E}(J) + \mathbb{E}(\exp(-wJ))\exp(\frac{\zeta}{\Gamma(\alpha)}\sum_{u=0}^{k-1}\exp(-h(s_k-s_u))(s_k-s_u)^{\alpha-1}g(u)\Delta) - 1. \tag{2.41}$$

At the end of the loop, $q_\lambda(s)$ is computed using the sum approximation:

$$q_\lambda(s) = \int_0^s \frac{dq_\lambda(u)}{du}du \approx \sum_{k=0}^{s} g(k)\Delta_k.$$

Therefore, we are able to compute numerically the mgf

$$\mathbb{E}(\exp(iwX_s) \mid \mathcal{F}_0) = \exp(-s[-iw(\mu - \frac{\sigma^2}{2}) + \frac{\sigma^2 w^2}{2}] + q_\lambda(s)\lambda_0 + iwX_0).$$

One assumes $X_0 = 0$, and retrieve the pdf either

1. using built-in `fft` function. This applies the previous divide-and-conqueer strategy;
2. applying manually the approximation, as presented by Hainaut in [1].

Approximated Fourier Transform using insight of FFT ([1])

$$f(x) \approx \frac{\Delta w}{\pi}\mathbb{R}\Big[\sum_{j=1}^{M} \delta_j \mathcal{M}_X(w_j)(-1)^{j-1}\exp(-i(k-1)(j-1)\frac{2\pi}{M})\Big]$$

where M is the number of steps and,

$$\begin{cases} \Delta_x = \frac{2x_{max}}{M-1}; \\ \Delta_w = \frac{2\pi}{M\Delta_x}; \\ w_j = (j-1)\Delta w \\ x_k = \frac{-M}{2}\Delta_x + (k-1)\Delta_x \quad k = 1, \dots, M \\ \delta_j = (\frac{1}{2})^{\mathbb{I}(j=1)} + \mathbb{I}(j \neq 1) = \begin{cases} \frac{1}{2} & \text{if } j = 1 \\ 1 & \text{if } j \neq 1 \end{cases} \end{cases}.$$

Proof. This last proof makes the link between both computations, and stress the consistency of a manual approximation with spectral analysis did previously. First-of-all, since $f(x)$ is real, its Fourier transform is Hermitian symmetric $\mathcal{M}_X(-z) = \mathcal{M}_X^*(z)$, and

$$f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \exp(-izx)\mathcal{M}_X(z)dz = \frac{1}{\pi}\mathbb{R}(\int_{\mathbb{R}_+} \underbrace{\exp(-izx)\mathcal{M}_X(z)}_{:=g(z)}\, dz).$$

First of all, we truncate the infinite integral at a specific frequency w_{max}. Secondly, we approximate the integral by a sum , using trapezoidal composite approximation. The interval $[0, w_{\max}]$ is splitted in M parts, at $w_j = (j-1)\Delta w$, with $\Delta w = \frac{w_{max}}{M}$:

$$\begin{aligned}\int_{\mathbb{R}_+} \underbrace{\exp(-izx)\phi(z)}_{:=g(z)}\, dz \approx \int_0^{w_{\max}} \underbrace{\exp(-izx)\phi(z)}_{:=g(z)}\, dz &\approx \Delta w(\frac{1}{2}g(w_1) + \sum_{j=2}^{M-1} g(w_j) + \frac{1}{2}g(w_M)) \\ &= \Delta w \sum_{j=1}^{M} \delta_j g(w_j),\end{aligned}$$

where $\delta_j = \begin{cases} \frac{1}{2} \text{ if } j = 1 \text{ or } j = M \\ 1 \text{ otherwise} \end{cases}$.

Moreover, the pdf must be sampled at $x_k = -x_{\max} + (k-1)\Delta x$. Then,

$$\begin{aligned}\exp(-iw_j x_k) &= \exp(-i(j-1)\Delta w(-x_{\max} + (k-1)\Delta x)) \\ &= \exp(i(j-1)\Delta w x_{\max})\exp(-i(j-1)\Delta w(k-1)\Delta x)\end{aligned}$$

In order to find back exact DFT basis function, we constraint the two exponential factors:

$$\begin{cases} \exp(-i(j-1)\Delta w(k-1)\Delta x) = \exp(-\frac{2\pi i(j-1)(k-1)}{M}) \\ \exp(i(j-1)\Delta w x_{\max}) = \exp(i(j-1)\pi) = (-1)^{j-1} \end{cases} \Rightarrow \begin{cases} \Delta w \Delta x = \frac{2\pi}{M} \\ x_{\max} = \frac{M\Delta x}{2} \end{cases} .$$

$$f(x) \approx \frac{\Delta w}{\pi} \mathbb{R}\Big[\sum_{j=1}^{M} \delta_j \phi(w_j)(-1)^{j-1} \exp(-i(k-1)(j-1)\frac{2\pi}{M}) \Big]$$

All this computations are equivocal to previous spectral analysis. The x-scale must be sampled, and then approximated through the DFT. The DFT is a truncature in the frequency space (as here we truncate at $w_{\max}$), after discrete approximation of the Fourier transform (through the trapezoidal approximation here, and going from continuous Fourier transform to Discrete Fourier Transform DTFT previously). □

Chapter 3

Calibration and illustration: BTC returns

In this chapter, the calibration of the parameters in the Markovian system (2.2) is detailed. This can be summarize in three steps:

- Peak-over-threshold method to distinguish between the diffusive and the jump part of the returns;
- Maximum Likelihood Estimation for parameters of the distribution of J, $\nu(-)$;
- and, finally, Maximum Likelihood Estimation for parameters of the intensity process λ_t.

Comparison between the different memory kernels presented in Chapter 1 is provided, in terms of Log-likelihood, or Chi-squared test (2.22) when the complexity of the models differ. Results for the Bitcoins hourly returns from 2018-02-09 to 2023-02-09 are presented. Finally, using calibrated parameters, I assess the ability of the model to replicate black implied volatility surface, and I compare the obtained results with the mutually exciting rough diffusion dynamic presented by Hainaut in [1].

3.1 Peak-over-threshold method

Let $(S_t)_{t\geq 0}$ denote the process of an asset price observed at discrete times with a fixed time step Δ. For instance, we consider crypto-currency (BTC log-returns 3.1) in this chapter. The rough models have already shown poor results for classical equities, which present no rough jump structure [3]. The method is presented here similarly than in Hainaut [4]. We define the log-returns as:

$$x_k = \log\left(\frac{S_{k\Delta}}{S_{(k-1)\Delta}}\right), \quad k = 1, \dots, n, \tag{3.1}$$

with observation times $T = \{t_1, t_2, \dots, t_n\}$, where $t_k = k\Delta$.

The objective of the Peak-over-threshold method is to split observation times T into three subsets:

- $T^+ \subset T$: times at which positive jumps occur;
- $T^- \subset T$: times at which negative jumps occur;

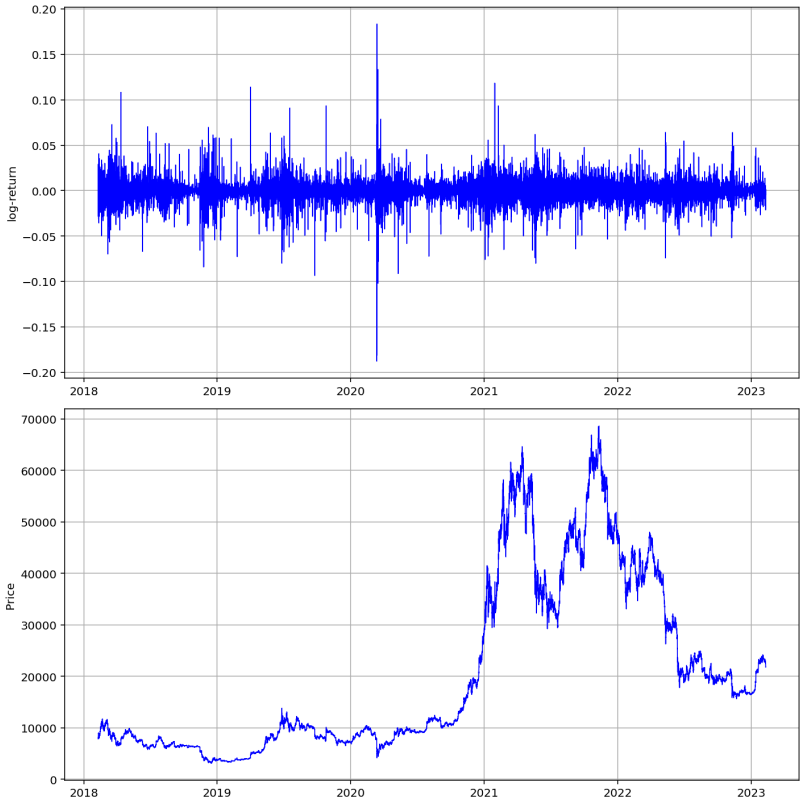

Figure 3.1: BTC log-returns between 2018-02-09 and 2023-02-09

- $T \setminus \{T^+, T^-\}$: times at which no jump occurs. This means that the log-returns occurring in $T \setminus \{T^+, T^-\}$ are strictly distributed in the diffusive part of the log-return dynamic:

$$\{x_k \mid t_k \in T \setminus \{T^+, T^-\}\} \sim \mathcal{N}(\mu_g \Delta, \sigma_g \sqrt{\Delta}).$$

The unbiased estimators of μ_g and σ_g are given by:

$$\mu_g = \frac{1}{n\Delta} \sum_{j=1 | t_j \in T \setminus \{T^+, T^-\}}^{n} x_j, \tag{3.2}$$

$$\sigma_g^2 = \frac{1}{(n-1)\Delta} \sum_{j=1 | t_j \in T \setminus \{T^+, T^-\}}^{n} (x_j - \mu_g)^2. \tag{3.3}$$

Let $\Phi(\cdot)$ denote the cumulative distribution function of a standard normal distribution. For a given confidence level $\alpha' \in (0, 1)$, we define the lower and upper thresholds as:

$$g(\alpha') = \mu_g \Delta + \sigma_g \sqrt{\Delta}\, \Phi^{-1}(\alpha'), \tag{3.4}$$

$$g(1-\alpha') = \mu_g \Delta + \sigma_g \sqrt{\Delta}\, \Phi^{-1}(1-\alpha'). \tag{3.5}$$

These thresholds define a confidence interval:

$$[g(\alpha'),\ g(1-\alpha')]. \tag{3.6}$$

The confidence level α' is selected so that the distribution of returns within the interval $[g(\alpha'), g(1-\alpha')]$ is as close as possible to a normal distribution. Returns lying outside the confidence interval are classified as jumps.

More precisely, we choose the α'^* level so that, inside the confidence interval, data exhibit the smallest skewness and excess kurtosis:

$$\alpha'^* = \arg\min_{\alpha'} \left(\mathbb{S}(\{x_k \mid x_k \in [g(\alpha'), g(1-\alpha')]\})^2 + (\mathbb{K}(\{x_k \mid x_k \in [g(\alpha'), g(1-\alpha')]\}) - 3)^2 \right)^{1/2}. \tag{3.7}$$

For BTC log-returns 3.1, I obtain the results represented in Figure 3.2. Thresholds are identified at -0.00975728 and 0.01001566.

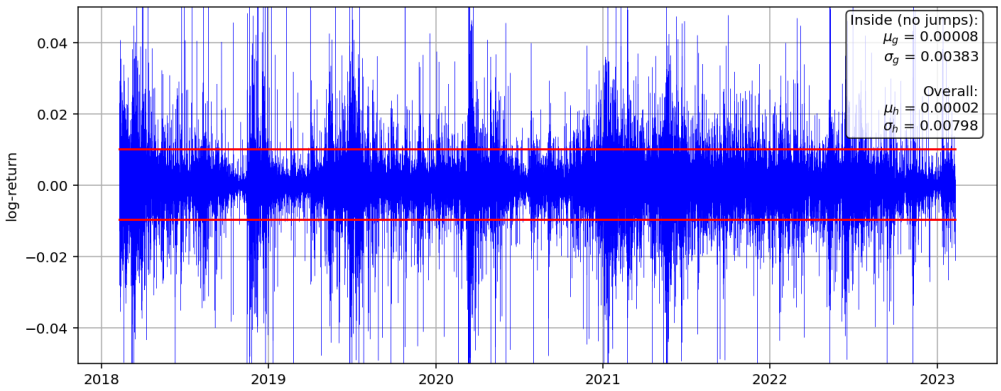

Figure 3.2: Peak-over-threshold method applied on BTC 3.1.

3.2 Maximum log-likelihood estimation of parameters

Firstly, the jump intensity parameters are obtained by log-likelihood maximization, as explained in 2.2. The calibration gives the parameters displayed in Table 3.3.1. The reader can observe that the optimal α value is approximately equal to 1. Actually, the optimal value is 1, but is not reachable because the sigmoid transform is applied to the parameter to ensure range of value between 0 and 1, but 1 is asymptotically never reached exactly. Then, Rough Hawkes kernel do not lead to a significant better fit than classical Hawkes kernel with exponential decreasing memory for the time series considered. This result is underwhelming, as it means that the introduction of a non-Markovian dynamic is not necessary for most returns. Nevertheless, this result confirms the same observation already drawn in Hainaut [3].

Parameter	**Estimated value**
α	0.9999999999919802
v	267.2099
ζ	228.5692
λ_0	147.1476
Log-likelihood	-31653.0345

Table 3.1: Calibrated parameters of the one-dimensional rough Hawkes process.

For these values, the path of the intensity is shown in Figure 3.3 and the stationary intensity given by formula 1.33 gives

$$\frac{\lambda_0 v^\alpha}{v^\alpha - \zeta} = 1017.5616.$$

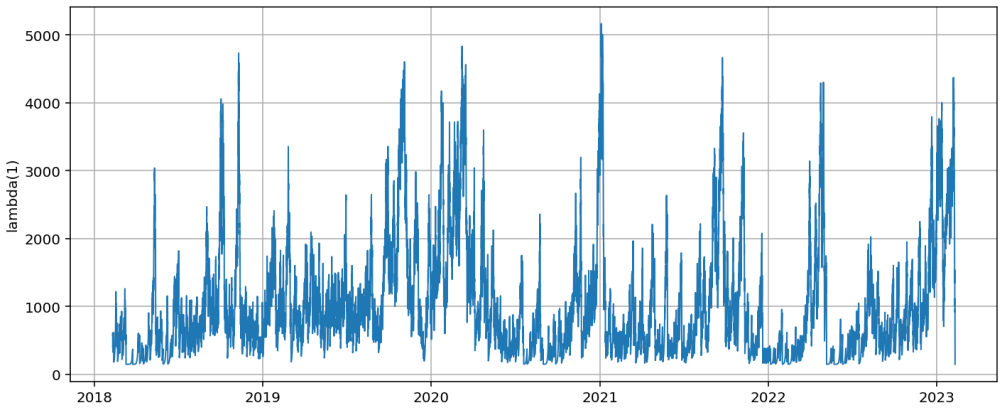

Figure 3.3: Path of Intensity for Rough Hawkes model, with calibrated parameters 3.3.1.

Secondly, one performs the estimation of jump severity by log-likelihood maximization. I refer the reader to the different distributions suggested for J in Section 2.2. The next positive and negative jump times are defined as:

$$t_k^+ = \min\{t_j \in T \mid x_j \geq g(1-\alpha),\, j \geq k\} \in T^+, \tag{3.8}$$

$$t_k^- = \min\{t_j \in T \mid x_j \leq g(\alpha),\, j \geq k\} \in T^-. \tag{3.9}$$

The number of positive and negative jumps up to time t_j are therefore given by:

$$N^+(t_j) = \max\{k \in \mathbb{N} \mid t_k^+ \leq t_j\}, \tag{3.10}$$
$$N^-(t_j) = \max\{k \in \mathbb{N} \mid t_k^- \leq t_j\}, \tag{3.11}$$

and the probability of positive jumps is estimated as:

$$p = \frac{N^+(t_n)}{N^+(t_n) + N^-(t_n)} = 0.4939. \tag{3.12}$$

We see well that log-returns exhibit positive and negative jumps (3.2), then the exponential distribution for magnitude is not suitable.The different distributions that can be considered have already been presented in 2.2:

1. J ∼DoubleExpo(p, η_+, η_-) : we observe than empirical proportion of positive jumps is well recovered during optimization in the calibrated parameters displayed in Table 1. This motivates the fact that for models exhibiting an higher number of parameters to fit, p could be taken empirically. Intuitively, it is not surprising that the model fitting the data the best is the one replicating the same proportion of positive and negative jumps.

Parameter	**Estimated value**
η_+	56.5378
η_-	55.1829
p	0.4939
Log-likelihood	-11838.8304

Table 3.2: Estimated parameters of the jump magnitude distribution (double exponential model)

2. J ∼Noisy DoubleExpo (p, η_+, η_-) with gaussian noise: in this case, the parameter p is not included in the optimization problem. Indeterminacy in the estimation is observed when p is included. However, as observed in 1, the empirical estimation is the minimizer, then it doesn't affect results in 2. The gaussian noise has volatility $\sigma = \sigma_g$ corresponding to the one estimated by peak-over-threshold method for log-returns occuring at $T \setminus \{T^+, T^-\}$.

Parameter	**Estimated value**
η_+	59.4126
η_-	57.8359
Log-likelihood	-11953.0611

Table 3.3: Estimated parameters of the jump magnitude distribution (Gaussian–double exponential model)

3.3 Comparison with other kernels

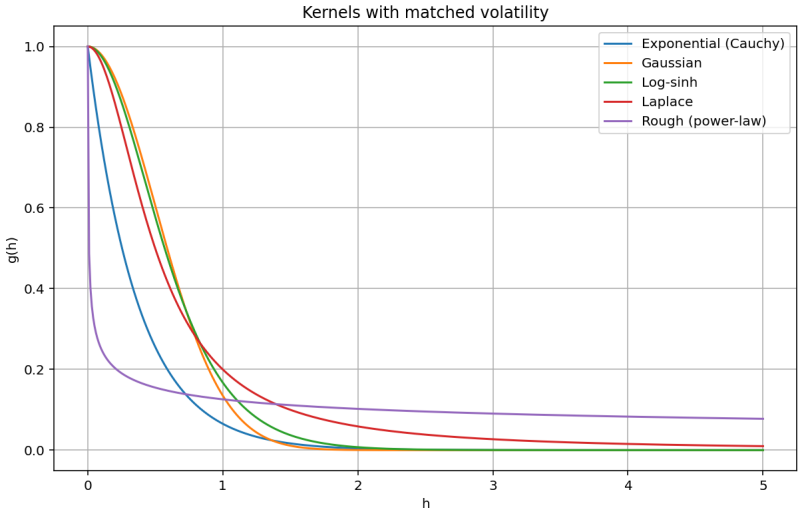

Figure 3.4: Comparison of kernels issued from different distributions, with matching variances (=16). The exponential case issued from Cauchy distribution is shown for parameter $\alpha = 2.73$, and the rough case with $\alpha = 0.7$. Inspired by [4].

The scheme is exactly the same than in previous section. Parameters of the intensity dynamic are estimated with log-likelihood maximization. All kernels have not a closed-form expression for their integral, then numerical approximation methods are used. In each case, we consider a different kernel $k(-)$ and the intensity of the jump process is given by

$$\lambda_t = \lambda_0 + \zeta \sum_{\tau_j < t} k(t - \tau_j). \tag{3.13}$$

3.3.1 Classical Hawkes kernel (exponential kernel from Cauchy distribution)

Without any surprise, this model, whose kernel is given by Equation (1.9), leads to exactly the same results that the ones we obtained for rough Hawkes kernel, with estimated $\alpha = 1$.

Parameter	Estimated value
υ	267.179
ζ	228.545
λ_0	147.1198
Log-likelihood	-31653.0345

Table 3.4: Calibrated parameters of the one-dimensional classical Hawkes process.

3.3.2 Laplacian measure

We consider a process with Laplacian kernel (1.3):

$$k(t) = \frac{\alpha^2}{\alpha^2 + t^2}.$$

Given observed jump times $\{\tau_1, \dots, \tau_{N_T}\}$, the log-likelihood over $[0, T]$ is

$$\log f(T \mid \alpha, \zeta, \lambda_0) = \sum_{k=1}^{N_T} \log(\lambda_{\tau_k^-}) - \int_0^T \lambda_s \, ds. \tag{3.14}$$

The intensity (3.13) evaluated at jump times is

$$\lambda_{\tau_k^-} = \lambda_0 + \zeta \sum_{j=1}^{k-1} k(\tau_k - \tau_j) = \lambda_0 + \zeta \sum_{j=1}^{k-1} \frac{\alpha^2}{\alpha^2 + (\tau_k - \tau_j)^2}.$$

For the second term, we integrate (3.13) and apply the change of measure $u = t - \tau_j$:

$$\int_0^T \lambda_t \, dt = \lambda_0 T + \zeta \sum_{j=1}^{N_T} \int_{\tau_j}^T k(t - \tau_j) \, dt = \lambda_0 T + \zeta \sum_{j=1}^{N_T} \int_0^{T-\tau_j} \frac{\alpha^2}{\alpha^2 + u^2} \, du.$$

The integral has a closed-form

$$\int \frac{\alpha^2}{\alpha^2 + u^2} \, du = \alpha \arctan\left(\frac{u}{\alpha}\right),$$

then we obtain

$$\int_0^T \lambda_t \, dt = \lambda_0 T + \zeta\alpha \sum_{j=1}^{N_T} \arctan\left(\frac{T - \tau_j}{\alpha}\right).$$

Therefore, the log-likelihood (3.14) get the following closed expression:

$$\log f(T \mid \alpha, \zeta, \lambda_0) = \sum_{k=1}^{N_T} \log\left(\lambda_0 + \zeta \sum_{j=1}^{k-1} \frac{\alpha^2}{\alpha^2 + (\tau_k - \tau_j)^2}\right) - \lambda_0 T - \zeta\alpha \sum_{j=1}^{N_T} \arctan\left(\frac{T - \tau_j}{\alpha}\right).$$

This minimization problem is very sensitive to initial guess, but after some trials and error, the algorithm finally converges to parameters in 3.3.2.

Parameter	**Estimated value**
α	0.133
ζ	4.313
λ_0	148.413
Log-likelihood	-30421.31

Table 3.5: Calibrated parameters for Laplace measure.

These parameters satisfy the stability condition

$$\zeta \int_0^\infty k(t) \, dt = \zeta \frac{\pi\alpha}{2} = 0.9042 < 1,$$

and $\lambda_\infty = \frac{1}{1 - \zeta \frac{\pi\alpha}{2}} = 1549.675$.

3.3.3 Gaussian spectral measure

We consider a Gaussian kernel (1.5):

$$k(u) = \exp\left(-\frac{u^2}{4\alpha}\right), \quad \alpha > 0.$$

The log-likelihood of the observed jump times $(\tau_1, \dots, \tau_{N_T})$ over the interval $[0, T]$ has still the same structure than Equation (3.14). At jump times, the intensity (3.13) takes the form

$$\lambda_{\tau_k^-} = \lambda_0 + \zeta \sum_{j<k} \exp\left(-\frac{(\tau_k^- - \tau_j^-)^2}{4\alpha}\right).$$

The integral term can be further computed:

$$\int_0^T \lambda_t dt = \lambda_0 T + \zeta \sum_{j=1}^{N_T} \int_{\tau_j}^T \exp\left(-\frac{(t-\tau_j)^2}{4\alpha}\right) dt, \text{ where } \int_{\tau_j}^T \exp\left(-\frac{(t-\tau_j)^2}{4\alpha}\right) dt = \sqrt{\pi\alpha}\, \mathrm{erf}\left(\frac{T-\tau_j}{2\sqrt{\alpha}}\right).$$

All components considered, the final formulation of the log-likelihood in the gaussian spectral case is

$$\log f(T \mid \alpha, \zeta, \lambda_0) = \sum_{j=1}^{n} \log(\lambda_0 + \zeta \sum_{j<k} \exp\left(-\frac{(\tau_k^- - \tau_j^-)^2}{4\alpha}\right)) - \lambda_0 T - \zeta\sqrt{\pi\alpha} \sum_{j=1}^{N_T} \mathrm{erf}\left(\frac{T-\tau_j}{2\sqrt{\alpha}}\right).$$

The estimated parameters are displayed in Table 3.3.3. The process is stationary because the stability condition is satisfied:

$$\zeta \int_0^\infty k(u)\, du = \zeta\sqrt{\pi\alpha} = 0.8149 < 1.$$

When the stability condition holds, the stationary intensity is given by

$$\lambda_\infty = \frac{\lambda_0}{1 - \zeta\sqrt{\pi\alpha}} = 1016.776.$$

Parameter	**Estimated value**
α	0.000006
ζ	188.210
λ_0	188.179
Log-likelihood	-31617.579

Table 3.6: Calibrated parameters for the Gaussian spectral measure.

3.3.4 Logistic measure

We consider a memory process with logistic-type kernel (1.7):

$$k(t) = \frac{\pi\alpha t}{\sinh(\pi\alpha t)}, \qquad t > 0.$$

The computations are exactly the same than with other kernels. Given observed jump times $\{\tau_1, \dots, \tau_{N_T}\}$ on the interval $[0, T]$, the log-likelihood of the process is given by (3.14). At each jump time τ_k, the intensity (3.13) is

$$\lambda_{\tau_k^-} = \lambda_0 + \zeta \sum_{j=1}^{k-1} k(\tau_k - \tau_j) = \lambda_0 + \zeta \sum_{j=1}^{k-1} \frac{\pi\alpha(\tau_k - \tau_j)}{\sinh\big(\pi\alpha(\tau_k - \tau_j)\big)}.$$

We decompose the second term and we apply the change of variables $u = t - \tau_j$:

$$\int_0^T \lambda_t\, dt = \lambda_0 T + \zeta \sum_{j=1}^{N_T} \int_{\tau_j}^T k(t - \tau_j)\, dt = \lambda_0 T + \zeta \sum_{j=1}^{N_T} \int_0^{T-\tau_j} \frac{\pi\alpha u}{\sinh(\pi\alpha u)}\, du.$$

Unlike previous kernels, this integral does not admit a simple closed-form expression, hence it is evaluated numerically in the implementation (trapezoidal quadrature on a discretized grid). The log-likelihood is therefore given by

$$\log f(T \mid \alpha, \zeta, \lambda_0) = \sum_{k=1}^{N_T} \log\left(\lambda_0 + \eta \sum_{j=1}^{k-1} \frac{\pi\alpha(\tau_k - \tau_j)}{\sinh\big(\pi\alpha(\tau_k - \tau_j)\big)}\right) - \lambda_0 T - \eta \sum_{j=1}^{N_T} \int_0^{T-\tau_j} \frac{\pi\alpha u}{\sinh(\pi\alpha u)}\, du.$$

Calibrated parameters are displayed in Table 3.3.4.

Parameter	Estimated value
α	27.7275
ζ	31.3257
λ_0	101.7520
Log-likelihood	- 31198.3011

Table 3.7: Calibrated parameters for the logistic measure.

Using the identity $\zeta \int_0^\infty \frac{t}{\sinh(t)}\, dt = \zeta \frac{\pi^2}{4}$, the stability condition becomes

$$\zeta \int_0^\infty \frac{\pi\alpha t}{\sinh(\pi\alpha t)}\, dt = \zeta \int_0^\infty \frac{\pi\alpha t}{\sinh(\pi\alpha t)}\, dt = \frac{\zeta\pi}{4\alpha} = 0.887 < 1.$$

We observe a long-run mean $\lambda_\infty = 900.458$.

The empirical results clearly show that the classical exponential kernel, corresponding to the standard Hawkes process, is not the most suitable specification for the considered dataset. Indeed, after recovering the maximized log-likelihood values, the Laplace, Logistic and Gaussian kernels all achieve larger likelihoods than the exponential specification:

$$\ell_{\text{Laplace}} > \ell_{\text{Logistic}} > \ell_{\text{Gaussian}} > \ell_{\text{Exp}} \approx \ell_{\text{Rough}}.$$

The empirical results suggest that kernels exhibiting slower decay structures provide a better description of the observed jump dynamics than the standard exponential Hawkes kernel. In particular, the Laplace and Logistic specifications significantly improve the fit, indicating a more persistent excitation mechanism and a longer memory structure than the one implied by exponential decay. The Gaussian kernel also yields an improvement, although less pronounced. By contrast, the rough kernel produces likelihood values almost identical to those of the exponential Hawkes model, and the associated likelihood-ratio tests fail to detect any significant difference between both specifications. Consequently, introducing roughness does not appear to provide additional explanatory power for the present dataset. Overall, the results indicate that the standard exponential Hawkes process is not sufficiently flexible to capture the temporal dependence structure of the data, whereas more persistent kernels, especially the Laplace kernel, provide a more appropriate representation of jump clustering effects.

3.4 Comparison with two-dimensional rough Hawkes processes

In [1], Hainaut introduces a bidimensional rough-Hawkes process for the intensities. This is an extension to my previously detailed uni-dimensional Hawkes process with double exponential jumps. Instead of considering one unique rough structure of positive

and negative jumps, Hainaut [1] modeled two different yet correlated memory dependence structures for positive jumps at one side, and negative ones at the other. The bidimensional jump frequency intensity takes the following form :

$$\begin{pmatrix} \lambda_t^{(1)} \\ \lambda_t^{(2)} \end{pmatrix} = \begin{pmatrix} \lambda_0^{(1)} \\ \lambda_0^{(2)} \end{pmatrix} + \begin{pmatrix} \zeta_{11} & \zeta_{12} \\ \zeta_{21} & \zeta_{22} \end{pmatrix} \begin{pmatrix} \int_0^t e^{-v(t-s)}(t-s)^{\alpha-1} \dfrac{1}{\Gamma(\alpha)}\, dN_s^{(1)} \\ \int_0^t e^{-v(t-s)}(t-s)^{\alpha-1} \dfrac{1}{\Gamma(\alpha)}\, dN_s^{(2)}. \end{pmatrix} \tag{3.15}$$

In consequence, the log-return process is adapted with two jump components:

$$X_t = \left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W_t + \sum_{j=1}^{2} \left(L_t^{(j)} - \mu_j \int_0^t \lambda_s^{(j)}\, ds \right).$$

The extension of the previous computations to multidimensional Hawkes processes is relatively straightforward. In particular, the reader may compare the results obtained in [1] with those presented in the first two chapters of this master's thesis.

Such a comparison is especially relevant since conclusions similar to those obtained by Hainaut in [1] are recovered. Although the unidimensional rough Hawkes specification does not exhibit a substantial improvement over the classical unidimensional exponential Hawkes model for jump modelling, the introduction of mutually exciting rough jumps in the multidimensional framework leads to significantly better performances. This highlights the importance of cross-excitation effects between positive and negative jumps. The calibrated parameters for dynamic (3.15) are displayed in Table 3.8 : we observe well that the likelihood maximization reaches a parameter $\alpha < 1$ and that the log-likelihood is above all of those obtained with univariate models in previous section. The two jump intensity dynamics are shown in Figure 3.5.

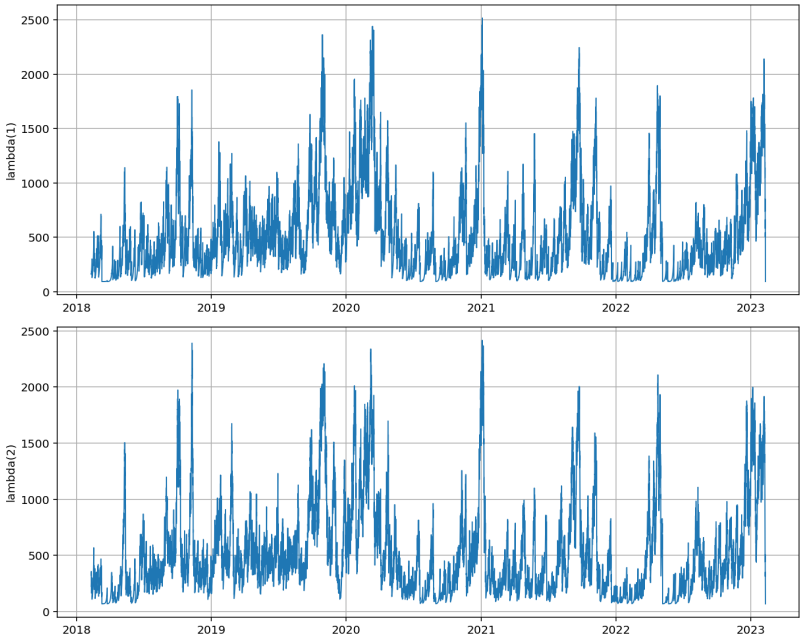

Figure 3.5: Paths of intensity for positive ($\lambda_{t,1}$) and negative ($\lambda_{t,2}$) jumps in the mutually exciting rough Hawkes setting.

Parameter	Value
α	0.909824454026053
v	169.5192568075053
$\lambda_{0,1}$	91.90428071
$\lambda_{0,2}$	68.59596111
η_{11}	82.28624195
η_{12}	4.12289411
η_{21}	25.37357272
η_{22}	67.17047154
$\log L$	-28031.73971409548

Table 3.8: Calibrated parameters for dynamic (3.15).

When considering multidimensional rough Hawkes processes, the step asking for more caution is the computation of stability requirements. In fact, if we take back the extension we made in the first chapter (1.32), mainly instead of a scalar, the long-run mean

$$u(t) = \lambda_0 + \zeta(k * u)_t \Leftrightarrow U(s) = \frac{\lambda_0}{s} + \zeta K(s)U(s), \tag{3.16}$$

behaves like a linear time-invariant (LTI) system with feedback kernel 3.6. This characterization is exploited in the proof of the theorem 3.4.

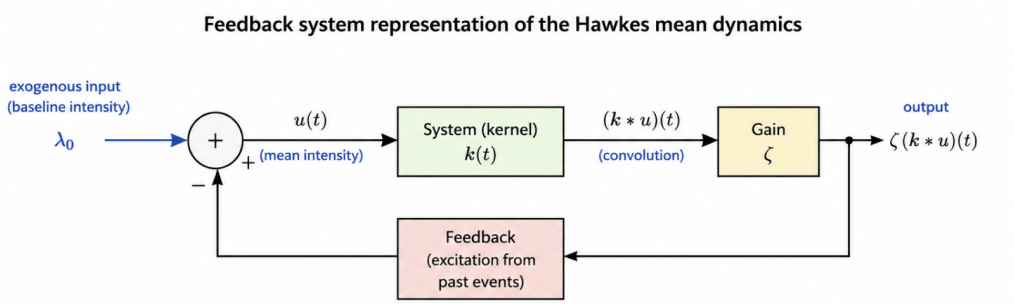

Figure 3.6: system representation of the rough Hawkes long-run mean dynamic

Stability of the mean process 3.6 in the mutually excited jumps dynamic (3.15)

The mutually exciting rough Hawkes system is stable if and only if

$$\begin{cases} \zeta_{11} < v^\alpha, \\ \zeta_{22} < v^\alpha, \\ \zeta_{12}\zeta_{21} < (v^\alpha - \zeta_{11})(v^\alpha - \zeta_{22}) \end{cases}$$

Proof. This system is BIBO (bounded-input bounded-output) stable only if the spectral radius of $\zeta K(0)$ is below unity:

$$\rho(\zeta K(0)) < 1 \Leftrightarrow \rho(\zeta) < K(0)^{-1}.$$

In the unidimensional scalar case, this condition was equivalent to $\zeta < K(0)^{-1} = v^\alpha$. In this two-dimensional case:

On the one hand we have the individual stability condition for each intensity process, taken independently, each being an unidimensional rough process:

$$\zeta_{11} < v^\alpha, \quad \zeta_{22} < v^\alpha.$$

On the other hand, the dependence between both processes induces an additional constraint for stability. We need to compute the maximum eigenvalue of ζ, $\rho(\zeta)$ which zeroes the characteristic polynomial:

$$\begin{aligned} \det(\begin{pmatrix} \lambda - \zeta_{11} & -\zeta_{12} \\ -\zeta_{21} & \lambda - \zeta_{22} \end{pmatrix}) = (\lambda - \zeta_{11})(\lambda - \zeta_{22}) - \zeta_{21}\zeta_{12} = 0 \\ \Leftrightarrow \lambda^2 - (\zeta_{11} + \zeta_{22})\lambda + (\zeta_{11}\zeta_{22} - \zeta_{21}\zeta_{12}) = 0. \end{aligned} \tag{3.17}$$

Among the two roots of this polynomial, the maximum eigenvalue is computed in the next Equation 3.18. Consequently, the condition $\lambda_{\max} = \rho(\zeta) < v^\alpha$ becomes

$$\begin{aligned} \lambda_{\max} &= \frac{1}{2}(\zeta_{11} + \zeta_{22} + \sqrt{(\zeta_{11} + \zeta_{22})^2 - 4(\zeta_{11}\zeta_{22} - \zeta_{12}\zeta_{21})}) < v^\alpha \\ &\Leftrightarrow \frac{1}{2}(\zeta_{11} + \zeta_{22} + \sqrt{(\zeta_{11} - \zeta_{22})^2 + 4\zeta_{12}\zeta_{21}}) < v^\alpha \\ &\Leftrightarrow \sqrt{(\zeta_{11} - \zeta_{22})^2 + 4\zeta_{12}\zeta_{21}} < 2v^\alpha - \zeta_{11} - \zeta_{22} \\ &\Rightarrow (\zeta_{11} - \zeta_{22})^2 + 4\zeta_{12}\zeta_{21} < (2v^\alpha - \zeta_{11} - \zeta_{22})^2 \\ &\Leftrightarrow \zeta_{11}^2 + \zeta_{22}^2 - 2\zeta_{11}\zeta_{22} + 4\zeta_{12}\zeta_{21} < 4v^{2\alpha} + \zeta_{11}^2 + \zeta_{22}^2 + 2\zeta_{11}\zeta_{22} - 4v^\alpha(\zeta_{11} + \zeta_{22}) \\ &\Leftrightarrow \zeta_{12}\zeta_{21} < v^{2\alpha} + \zeta_{11}\zeta_{22} - v^\alpha(\zeta_{11} + \zeta_{22}) = (v^\alpha - \zeta_{11})(v^\alpha - \zeta_{22}) \end{aligned} \tag{3.18}$$

These conditions are the same that the ones obtained in [1], but taking the intuitive point of view allow for lighter computations here. Instead, Hainaut in [1] inverted the full two-dimensional system 3.16 in the Laplace domain, the same way we did in the unidimensional case 1.32. This results is heavy computations. □

The calibration of the rough Hawkes mutually exciting intensities gives parameters fulfilling previous BIBO stability conditions 3.4.Consequently, the system reaches the following long-run intensities:

$$\lambda_{\infty,1} = 485.405, \quad \lambda_{\infty,2} = 496.644. \tag{3.19}$$

Even though we encounter poor results for modeling memory structure in jumps with one-dimensional rough Hawkes process, in two-dimensions, when we separate the memory structure of positive and negative jumps, rough Hawkes kernel is well preferable than exponential kernel. This is an encouraging result about the use of rough kernels, obtained by the chi-squared statistic (with one degree of freedom):

$$2(\text{Loglik}_{rough} - \text{Loglik}_{exp}) = 19.3.$$

The calibrated parameters for the exponential kernel are displayed in Table **??**.

Table 3.9: Calibrated parameters for the two-dimensional exponential Hawkes model.

Parameter	Value
v	230.27
$\lambda_{0,1}$	60.0147
$\lambda_{0,2}$	113.658997
η_{11}	98.922
η_{12}	101.0919
η_{21}	2.1854
η_{22}	177
$\lambda_{\infty,1}$	499.0734
$\lambda_{\infty,2}$	511.7533

3.5 Call pricing

In this last section, the European call surfaces induced by each of the previous uni-dimensional and bidimensional rough models are compared (without drift and Brownian diffusion : $\sigma = r = 0$). S_0 is assumed equal to 100, and $K \in [45, 145]$, $T \in [1, 6]$ months, all other variables are equal to their estimation performed in previous sections. Only the rough parameter α for the uni-dimensional model is set to 0.9098, the optimal value obtained for the bidimensional model. I remind the reader that the a call option is said to be

- *in-the-money (ITM)* if $S_0 > K$, meaning it has positive intrinsic value.
- *out-of-the-money (OTM)* if $S_0 < K$, meaning it has no intrinsic value.
- *at-the-money (ATM)* if $S_0 = K$, meaning the spot price equals the strike price.

In consequence, in the surface, low relative strikes are in-the-money and large ones out-the-money.

The implied volatility surface for European call options is obtained by inverting Black-Scholes call formula. We process via the following steps:

1. First of all, we compute the pdf $f(x)$ of log-returns by DFFT, using technique previously described in Chapter 2;
2. Then, we price the call option in the rough model :

$$\text{Call}_{rough}(K,T) = \exp(-rT)\mathbb{E}^{\mathbb{Q}}((S_T - K)_+ \mid \mathcal{F}_0) = \int_0^\infty (S_0 \exp(x) - K)_+ f(x)dx;$$

3. Finally, the implied volatility is recover by matching the previously obtained value of the call option $\text{Call}_{rough}(K,T)$ with the Black-Scholes'one:

$$\text{Call}_{rough}(K,T) = S_0\,\Phi(d_1) - Ke^{-rT}\Phi(d_2), \text{ where}$$

$$\begin{cases} d_1 = \frac{\ln\left(\frac{S_0}{K}\right) + \left(r + \frac{1}{2}\hat{\sigma}^2\right)T}{\hat{\sigma}\sqrt{T}}. \\ d_2 = d_1 - \hat{\sigma}\sqrt{T} \end{cases}$$

By inverting the Black-Scholes formula, we obtain the implied volatility of the rough call option $\hat{\sigma}$ for a range of strikes K and maturities T observed.

Days	Expectation	Std. deviation	5%	95%
30	0.005	0.191	-0.318	0.318
60	0.005	0.272	-0.455	0.445
90	0.005	0.322	-0.533	0.533
120	0.005	0.357	-0.592	0.592
150	0.005	0.384	-0.641	0.631
180	0.005	0.406	-0.670	0.670

Table 3.10: pdf statistics after DFFT for the 1D-rough model, with truncature $x_{\max} = 2.5$, $M = 2^9$ components, and $\sigma = r = 0$.

Days	Expectation	Std. deviation	5%	95%
30	0.005	0.169	-0.308	0.250
60	0.005	0.244	-0.435	0.367
90	0.005	0.294	-0.514	0.455
120	0.005	0.331	-0.582	0.514
150	0.005	0.361	-0.621	0.563
180	0.005	0.386	-0.670	0.612

Table 3.11: pdf statistics after DFFT for the 2D-rough model, with truncature $x_{\max} = 2.5$, $M = 2^9$ components, and $\sigma = r = 0$.

This inverse-engineering algorithm gives the European call surfaces in Figure 3.7.

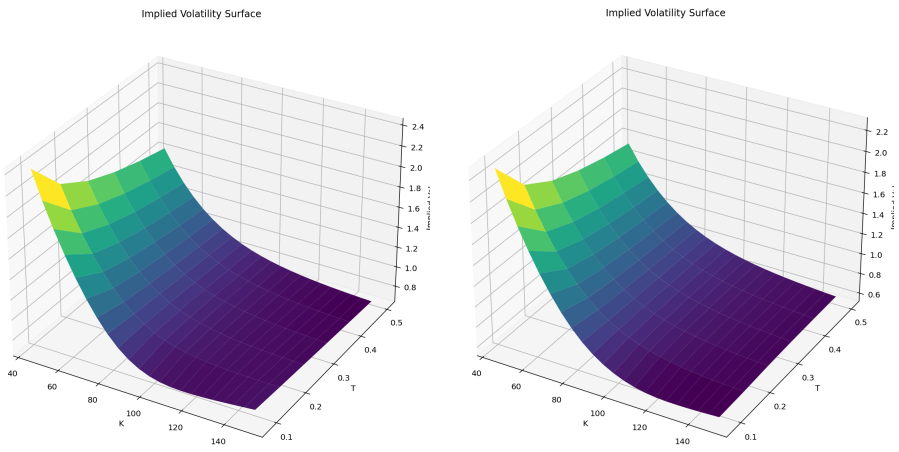

Figure 3.7: On the left : implied volatility surface of 1D rough Hawkes model (2.2), and on the right : implied volatility surface of 2D mutually exciting rough Hawkes model (3.16).

We check the properties of the pdf in Tables 3.5 and 3.5. The 90% interval grows up to $[-67, 67]\%$ for the 1D rough Hawkes model and up to $[-67, 61.2]\%$ for the mutual 2D rough Hawkes model. This highlights the non-symmetric structure introduced by the mutually exciting model compared to the one-dimensional case.

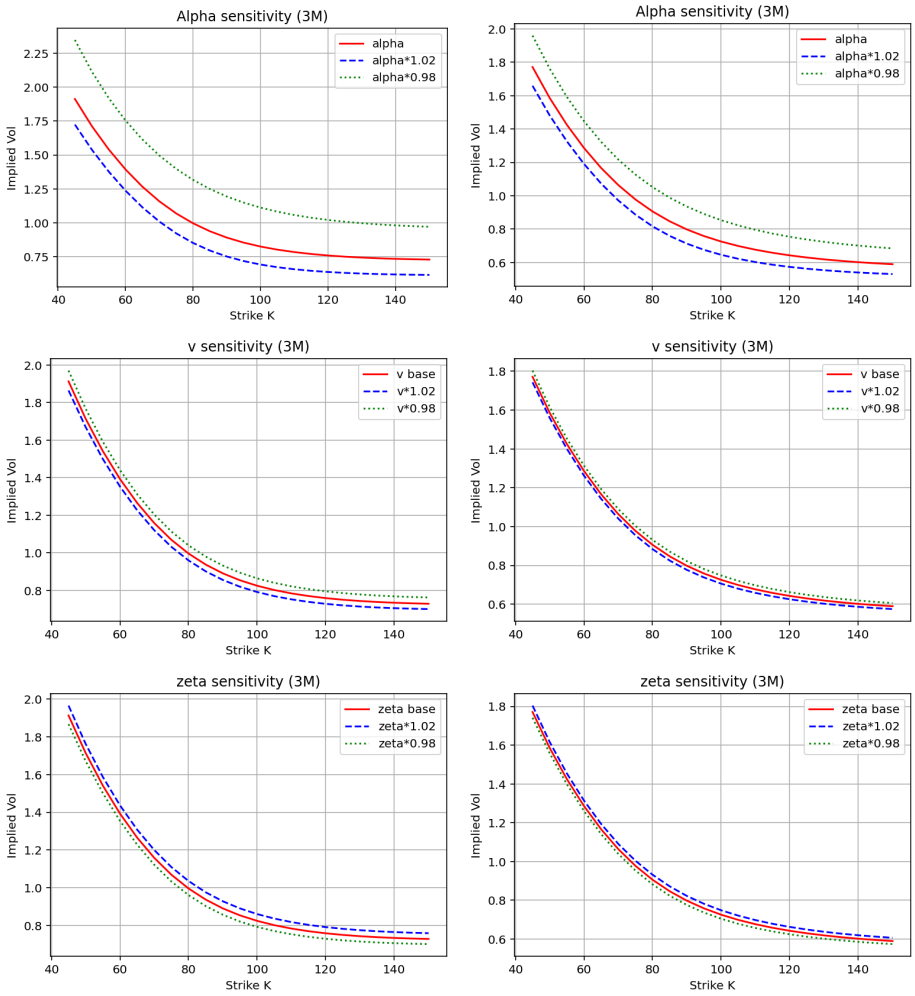

Figure 3.8: Sensitive analysis of the European call curve for $T = 3$ months. On the left, sensitiveness for the 1D rough Hawkes model, and on the right for the 2D mutually exciting rough jumps (in order to match with the definitions in [1], here we have Alpha $= \alpha$, v $= v$ and zeta $= \zeta$.)

In Figure 3.8, after stressing each parameter of the intensity with shocks of $\pm 2\%$, we observe that

- decreasing α promotes the roughness of the model, and increase the implied volatility. This shock is stronger in the univariate model.
- increasing v makes the damping of the rough Hawkes kernel stronger, which makes memory be shorter in the process, and decrease the implied volatilities. Again, slightly stronger effect are observed in the unidimensional case.
- Finally, increasing/decreasing the parameter(s) ζ has a direct proportional impact on the implied volatility, as this factor directly drives the proportion of the memory part in the intensity.

Chapter 4

Rough Hawkes-Heston models

In this chapter, the rough Hawkes-Heston model is introduced in two different but complementary settings. The first one introduces jumps in the volatility dynamic. The second one arises from market observation: a single jump is modeled for impacting both the volatility and the underlying.

Firstly, I present a small state-of-the art of literature using rough processes in Heston models. Secondly, the moment generating functions of the log-returns X_t are computing in both settings.

In [5], Hainaut introduces the rough-diffusive Heston model. Instead of rough jumps, rough diffusion is studied. I recall the definition of a classical Brownian motion [4] to the reader.

Brownian Motion

A Brownian motion $(B_t)_{t\geq 0}$ is a process such that

- $B_0 = 0$;
- $B_t - B_s \perp \mathcal{F}_s$ (independent increments)
- $B_t - B_s \sim \mathcal{N}(0, \sqrt{t-s})$ (stationary increments)
- its trajectory is continuous with $\mathbb{V}(B_t) = t$.

Generalization of this definition naturally arises in literature. For instance, Levy processes can be introduced as an extension of the Brownian motion for which the continuity of its trajectory is only ensured continuously in probability [2].

A fractional Brownian motion generalizes the $(B_t)_{t\geq 0}$ by introducing dependency in increments through the Hurst index $H \in (0, 1)$.

Fractional Brownian Motion

A fractional Brownian motion $(B_t^H)_{t\geq 0}$ is a process such that

- increments are dependent with
  - $\mathbb{C}(B_{t+h}^H - B_t^H, B_{s+h}^H - B_s^H) \propto H(2H-1) \mid t-s \mid^{2H-2}$
  - $\mathbb{E}(B_t^H B_s^H) = \frac{1}{2}(t^{2H} + s^{2H} - \mid t-s \mid^{2H})$
- $\mathbb{V}(B_t^H) = t^{2H}$ and $\mathbb{E}(B_t^H) = 0$

Similarly to our rough jump setting, the fractional diffusion can be expressed using a Sonine operator :

$$B_t^H = (k * dW)(t) = \int_0^t k(t-s)dW_s,$$

where $k(-)$ is a chosen kernel. The parameter H measures the dependence degree, and relates to our previous α:

$$\alpha = H + \frac{1}{2}.$$

For a choice $H = \frac{1}{2}$ (i.e. $\alpha = 1$), we recover the classical Brownian Motion[1]. For $H < \frac{1}{2}$ (i.e. $\alpha \in (\frac{1}{2}, 1)$), we observe a negative covariance between increments. Hainaut[5] has introduced the rough Hawkes-Heston model using the fractional Brownian motion in the volatility process:

$$V_t = V_0 + \int_0^t g(t-s)\kappa(\theta - V_s)ds + \int_0^t g(t-s)\sigma\sqrt{V_s}dW_s. \tag{4.1}$$

Observing the large jumps in the VIX dynamic 4.1, I got inspired by this model for introducing rough jumps in volatility.

I present in this chapter two different rough jumps Hawkes-Heston models. Indeed, the literature and our previous results have already shown that the usefulness of rough jumps in stock price dynamics is limited. They are generally better suited to cryptocurrencies or to stock volatility, where more jumps and irregularities tend to occur. Notice that the model 4.2 cannot be considered because it is not an affine model (because of the factor $\sqrt{V_t}\lambda_t$):

$$\begin{cases} dX_t = (\mu - \frac{V_t}{2})dt + \sqrt{V_t}dW_t^X \\ dV_t = \kappa(\theta - V_t)dt + \sigma\sqrt{V_t}(dW_t^V + dL_t - \lambda_t\mathbb{E}(J)dt) \end{cases} \tag{4.2}$$

The framework we introduced in the first chapter is only applicable to the affine Volterra processes. Without this property, we cannot postulate an affine exponential parametrization of the mgf. Following the expression of Hainaut 4.1, adding rough jumps to the rough diffusion would introduce the following dynamic for volatility :

$$\begin{aligned} V_t &= V_0 + \int_0^t g(t-s)\kappa(\theta - V_s)ds + \int_0^t g(t-s)\sigma\sqrt{V_s}dW_s + \int_0^t g(t-s)dL_s \\ &= V_0 + \kappa\theta\int_0^t g(t-s)ds - \kappa\int_0^t g(t-s)\sigma V_s ds + \sigma\int_0^t g(t-s)\sqrt{V_s}dW_s + \int_0^t g(t-s)dL_s \\ &= f(t) + \int_0^t g(t-s)\underbrace{[-\kappa V_s ds + \sigma\sqrt{V_s}dW_s + dL_s]}_{dY_s} \\ &= f(t) + (g * dY)_t. \end{aligned} \tag{4.3}$$

The rough diffusion won't be treated in this thesis, but coupling rough diffusion and jump dynamic could definitely be an interesting framework for further research.

In particular, I will consider two distinct dynamics in this chapter.The rough Hawkes-Heston model:

$$\begin{cases} dX_t = (\mu - \frac{V_t}{2})dt + \sqrt{V_t}dW_t^X \\ dV_t = \kappa(\theta - V_t)dt + \sigma\sqrt{V_t}dW_t^V + dL_t - \lambda_t\mathbb{E}(J)dt \\ d\lambda_t = \zeta\int_{\mathbb{R}_+} dY_t(z)\mu(dz) \\ dY_t(z) = -(v + b_l)Y_t(z)dt + dN_t \end{cases}, \tag{4.4}$$

[1] Observe that $H = \frac{1}{2}$ is the root of the covariance term in 4.

and its extension, namely the mutually roughly excited stock-volatility model:

$$\begin{cases} dX_t = (\mu - \frac{V_t}{2})dt + \sqrt{V_t}dW_t^X - \delta(dL_t - \lambda_t \mathbb{E}(J)dt) \\ dV_t = \kappa(\theta - V_t)dt + \sigma\sqrt{V_t}dW_t^V + dL_t - \lambda_t \mathbb{E}(J)dt \\ d\lambda_t = \zeta \int_{\mathbb{R}_+} dY_t(z)\mu(dz) \\ dY_t(z) = -(v + b_l)Y_t(z)dt + dN_t \end{cases} \tag{4.5}$$

where W_t^X and W_t^V are two correlated brownian motions with correlation ρ. In other words, we can simulate them using two independent and normally distributed variables ϵ_t^1 and ϵ_t^2 so that :

$$\begin{cases} dW_t^X = \epsilon_t^1 \\ dW_t^V = \rho\epsilon_t^1 + \sqrt{1-\rho^2}\epsilon_t^2 \end{cases}.$$

The first model directly extends the Hawkes-Heston model with volatility in a fractional diffusive market of Hainaut [5]. The objective of the second one is to introduce a parameter in the model to encounter from an empirical observation : stock price tends to present shocks following jumps in the volatility structure, in opposite sign. I illustrate this in 4.1

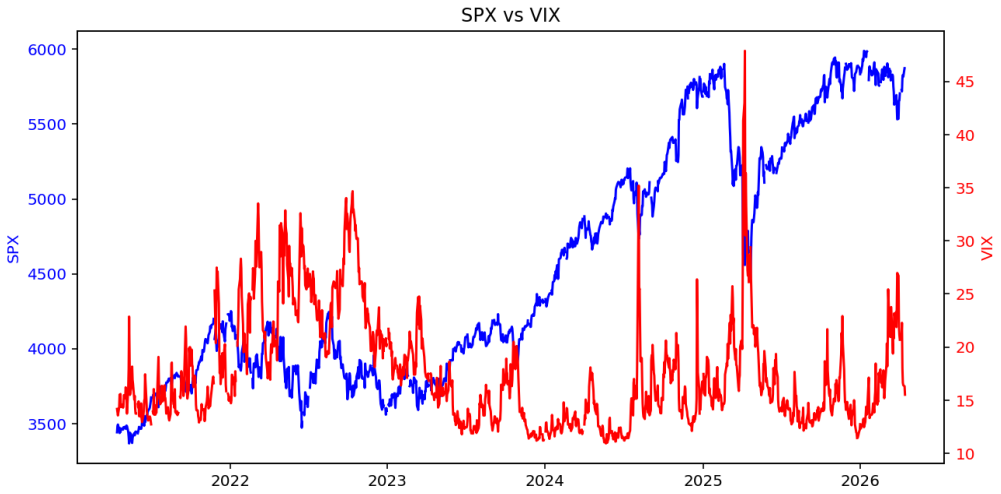

Figure 4.1: Historical S&P and VIX from June 2021 up to end March 2026. One observes that S&P tends to decrease when VIX jumps. This observation is the incentive to introduce a contribution of the VIX jumps into the stock dynamic, through a parameter δ. This parameter should be calibrated to capture the impact of shocks in volatility on the stock dynamic. In fact, increase (resp. decrease) in VIX is generally followed by decrease (resp. increase) in stock, but shocks in volatility are higher than in stocks. Parameter δ can be view as the scaling parameter which translates shocks in volatility into shocks in stock. Therefore, we expect $\delta < 1$.

4.1 Rough Hawkes-Heston volatility model

In this section, we follow the steps of the procedure for obtaining the mgf of the affine Volterra system 4.4 (rough Hawkes-Heston), as described in Chapter 1. The final result is presented in 4.1.

Log-returns mgf of the rough Hawkes-Heston model 4.4

$$\mathbb{E}(\exp(-wX_s) \mid \mathcal{F}_0) = \exp([-w\mu T + \kappa\theta \int_0^s q_V(u)du] + q_\lambda(s)\lambda_0 + q_V(s)V_0 - wX_0)$$

where

$$\partial_s q_\lambda(s) = -\mathbb{E}(J)q_V(s) + \exp(\zeta(K\frac{dq_\lambda}{ds})_s)\mathbb{E}(\exp(q_V(s)J)) - 1;$$

$$\partial_s q_V(s) = -q_V(s)[\rho\sigma w + \kappa - \frac{q_V(s)\sigma^2}{2}] + \frac{w}{2}(1+w)$$

Proof. **4.1.1 Step 1**

The dampened version of $k(-)$ has a spectral representation via Fourier transform, then we can embed all the memory of the system in an infinity of new state variables. This is represented by the continuous process $\{Y_t(u)\}_{u\in\mathbb{R}_+}$. However, because the kernel has a pole at $h = 0$, we consider the unilateral transform. This is why $Y_t(u)$ is defined on $\mathbb{R}_+$.

4.1.2 Step 2

After the unilateral finite-dimensional approximation procedure (See the procedure in first chapter 1.2), we obtain the following system :

$$\begin{cases} d\tilde{X}_t = (\mu - \frac{\tilde{V}_t^2}{2})dt + \sqrt{\tilde{V}_t}dW_t^X \\ d\tilde{V}_t = \kappa(\theta - \tilde{V}_t)dt + \sigma\sqrt{\tilde{V}_t}dW_t^V + d\tilde{L}_t - \tilde{\lambda}_t\mathbb{E}(J)dt \\ d\tilde{\lambda}_t = \zeta\sum_{l=0}^{n-1} w_l d\tilde{Y}_t(l) \\ d\tilde{Y}_t(l) = -(v + b_l)\tilde{Y}_t(l)dt + d\tilde{N}_t \end{cases} \tag{4.6}$$

4.1.3 Step 3

As before, we aim computing the backward Kolmogorov equation :

$$\mathcal{A}f + \partial_t f = 0 \tag{4.7}$$

where

$$f := f(t, \tilde{X}_t, \tilde{V}_t, \tilde{\lambda}_t, \tilde{Y}_t(l), \tilde{N}_t).$$

The computation of the infinitesimal generator is done through the generalized Itô's lemma applied to f :

$$\begin{aligned} df = {} & \partial_t f dt + \partial_{\tilde{X}} f d\tilde{X}_t + \partial_{\tilde{V}} f d\tilde{V}_t + \partial_{\tilde{\lambda}} f d\tilde{\lambda}_t + \sum_{l=0}^{n-1} \partial_{Y(l)} f dY_t(l) \\ & + \frac{1}{2}\partial^2_{\tilde{X}\tilde{X}} f \underbrace{d\langle\tilde{X}\rangle_t}_{=\tilde{V}_t dt} + \frac{1}{2}\partial^2_{\tilde{V}\tilde{V}} f \underbrace{d\langle\tilde{V}\rangle_t}_{=\sigma^2\tilde{V}_t dt} + \frac{1}{2}\partial^2_{\tilde{\lambda}\tilde{\lambda}} f \underbrace{d\langle\tilde{\lambda}\rangle_t}_{=0} + \partial^2_{\tilde{X}\tilde{V}} f \underbrace{d\langle\tilde{X},\tilde{V}\rangle_t}_{\rho\sigma\tilde{V}_t dt} \\ & + (f(\tilde{X}_t, \tilde{V}_{t^-} + z, \tilde{\lambda}_t + \zeta, Y_t + 1, \tilde{N}_t + 1) - f)d\tilde{N}_t. \end{aligned} \tag{4.8}$$

The infinitesimal generator directly follows via

$$\begin{aligned}\mathbb{E}(\frac{df}{dt} \mid \mathcal{F}_t) = \partial_t f + \partial_{\tilde{X}} f \mathbb{E}(\frac{d\tilde{X}_t}{dt} \mid \mathcal{F}_t) + \partial_{\tilde{V}} f \mathbb{E}(\frac{d\tilde{V}_t}{dt} \mid \mathcal{F}_t) + \partial_{\tilde{\lambda}} f \mathbb{E}(\frac{d\tilde{\lambda}_t}{dt} \mid \mathcal{F}_t) \\ + \sum_{l=0}^{n-1} \partial_{Y(l)} f \mathbb{E}(\frac{dY_t(l)}{dt} \mid \mathcal{F}_t) + \frac{\tilde{V}_t}{2} \partial_{\tilde{X}\tilde{X}} f + \frac{\sigma^2 \tilde{V}_t}{2} \partial_{\tilde{V}\tilde{V}} f + \rho\sigma \tilde{V}_t \partial_{\tilde{X}\tilde{V}} f \\ + \tilde{\lambda}_t \int_0^t \Big[f(\tilde{X}_t, \tilde{V}_{t^-} + z, \tilde{\lambda}_t + \zeta, Y_t(l) + 1, \tilde{N}_t + 1) - f \Big] \tilde{\nu}(dz).\end{aligned} \tag{4.9}$$

The drift terms are obtained from the dynamic of our Heston model 4.4:

$$\begin{cases} \mathbb{E}(\frac{d\tilde{X}_t}{dt} \mid \mathcal{F}_t) = \mu - \frac{\tilde{V}_t}{2} \\ \mathbb{E}(\frac{d\tilde{V}_t}{dt} \mid \mathcal{F}_t) = \kappa(\theta - \tilde{V}_t) - \tilde{\lambda}_t \mathbb{E}(J) \\ \mathbb{E}(\frac{d\tilde{\lambda}_t}{dt} \mid \mathcal{F}_t) = -\zeta \sum_{l=0}^{n-1} w_l (v + b_l) Y_t(l) \\ \mathbb{E}(\frac{dY_t(l)}{dt} \mid \mathcal{F}_t) = -(v + b_l) Y_t(l) \end{cases}.$$

Then,

$$\begin{aligned}\mathbb{E}(\frac{df}{dt} \mid \mathcal{F}_t) = \partial_t f + \partial_{\tilde{X}} f \Big[\mu - \frac{\tilde{V}_t}{2} \Big] + \partial_{\tilde{V}} f \Big[\kappa(\theta - \tilde{V}_t) - \tilde{\lambda}_t \mathbb{E}(J) \Big] - \partial_{\tilde{\lambda}} f \Big[\zeta \sum_{l=0}^{n-1} w_l (v + b_l) Y_t(l) \Big] \\ - \sum_{l=0}^{n-1} \partial_{Y(l)} f (v + b_l) Y_t(l) + \frac{\tilde{V}_t}{2} \partial_{\tilde{X}\tilde{X}} f + \frac{\sigma^2 \tilde{V}_t}{2} \partial_{\tilde{V}\tilde{V}} f + \rho\sigma \tilde{V}_t \partial_{\tilde{X}\tilde{V}} f \\ \underbrace{+ \tilde{\lambda}_t \int_0^t \Big[f(\tilde{X}_t, \tilde{V}_{t^-} + z, \tilde{\lambda}_t + \zeta, Y_t(l) + 1, \tilde{N}_t + 1) - f \Big] \tilde{\nu}(dz)}_{\mathcal{A}f(-)} = 0.\end{aligned} \tag{4.10}$$

The solution of the backward Kolmogorov equation

$$\mathcal{A}f(-) + \partial_t f(-) = 0$$

is exactly the one that zeroes the infinitesimal generator

$$\mathbb{E}(\frac{df}{dt} \mid \mathcal{F}_t) = 0.$$

As before, we observe that the mgf of affine Volterra processes can take an affine exponential parametrization such that

$$u(t,x) = \mathbb{E}(\exp(-wX_T) \mid \mathcal{F}_t) = \exp(q_0(t,T) + q_{\tilde{\lambda}}(t,T)\tilde{\lambda}_t + q_{\tilde{V}}(t,T)\tilde{V}_t + \sum_{l=0}^{n-1} q_l(t,T) Y_t(l) - w\tilde{X}_t).$$

4.1.4 Step 4

For this last step, we take the formula of the infinitesimal generator developed in previous step, and we apply it to $u(t,x)$:

$$\mathbb{E}(\frac{du}{dt}(t,x) \mid \mathcal{F}_t) = 0.$$

We can compute the derivatives of $u(t,x)$ via its parametrization :

$$\begin{cases} \partial_t u(t,x) = \left[\partial_t q_0(t,T) + \partial_t q_{\tilde{\lambda}}(t,T)\tilde{\lambda}_t + \partial q_{\tilde{V}}(t,T)\tilde{V}_t + \sum_{l=0}^{n-1} \partial q_l(t,T) Y_t(l)\right] u(t,x); \\ \partial_{\tilde{X}} u(t,x) = -w u(t,x); \\ \partial_{\tilde{V}} u(t,x) = q_{\tilde{V}}(t,T) u(t,x); \\ \partial_{\tilde{\lambda}} u(t,x) = q_{\tilde{\lambda}}(t,T) u(t,x); \\ \partial_{Y_t(l)} u(t,x) = q_l(t,T) u(t,x); \\ \partial^2_{\tilde{X}\tilde{X}} u(t,x) = w^2 u(t,x); \\ \partial^2_{\tilde{V}\tilde{V}} u(t,x) = q^2_{\tilde{V}}(t,T) u(t,x); \\ \partial^2_{\tilde{X}\tilde{V}} u(t,x) = -w q_{\tilde{V}}(t,T) u(t,x). \end{cases}$$

Moreover, we can simplify the integral term as in Chapter 2 :

$$\begin{aligned} &\int_0^t \left[f(\tilde{X}_t, \tilde{V}_{t^-} + z, \tilde{\lambda}_t + \zeta \sum_{l=0}^{n-1} w_l, Y_t(l) + 1, \tilde{N}_t + 1) - f \right] \tilde{\nu}(dz) \\ &= u(t,x) \left[\int_0^t \exp(q_{\tilde{V}}(t,T) z + \zeta \sum_{l=0}^{n-1} w_l q_{\tilde{\lambda}}(t,T) + q_l(t,T)) \tilde{\nu}(dz) - 1 \right] \\ &= u(t,x) \left[\exp(\zeta \sum_{l=0}^{n-1} w_l q_{\tilde{\lambda}}(t,T) + q_l(t,T)) \int_0^t \exp(q_{\tilde{V}}(t,T) z) \tilde{\nu}(dz) - 1 \right] \\ &= u(t,x) \left[\exp(\zeta \sum_{l=0}^{n-1} w_l q_{\tilde{\lambda}}(t,T) + q_l(t,T)) \mathbb{E}(\exp(q_{\tilde{V}}(t,T) J) - 1 \right] \end{aligned} \tag{4.11}$$

Consequently, all things gathered in Equation (4.10), we have :

$$\begin{aligned} \mathbb{E}(\frac{du}{dt} \mid \mathcal{F}_t) = u(t,x) \Bigg(&\partial_t q_0(t,T) + \partial_t q_{\tilde{\lambda}}(t,T)\tilde{\lambda}_t + \partial q_{\tilde{V}}(t,T)\tilde{V}_t + \sum_{l=0}^{n-1} \partial q_l(t,T) Y_t(l) \\ &- w \left[\mu - \frac{\tilde{V}_t}{2}\right] + q_{\tilde{V}}(t,T) \left[\kappa(\theta - \tilde{V}_t) - \tilde{\lambda}_t \mathbb{E}(J)\right] - q_{\tilde{\lambda}}(t,T) \left[\zeta \sum_{l=0}^{n-1} w_l (v + b_l) Y_t(l)\right] \\ &- \sum_{l=0}^{n-1} q_l(t,T)(v + b_l) Y_t(l) + \frac{\tilde{V}_t}{2} w^2 + \frac{\sigma^2 \tilde{V}_t}{2} q^2_{\tilde{V}}(t,T) - w q_{\tilde{V}}(t,T) \rho \sigma \tilde{V}_t \\ &+ \tilde{\lambda}_t \left[\exp(\zeta \sum_{l=0}^{n-1} w_l q_{\tilde{\lambda}}(t,T) + q_l(t,T)) \mathbb{E}(\exp(q_{\tilde{V}}(t,T) J) - 1 \right] \Bigg) = 0. \end{aligned} \tag{4.12}$$

We identify the coefficients of the different variables and equal them to zero :

- for the constant terms :

$$\partial_t q_0(t,T) - w\mu + q_{\tilde{V}}(t,T)\kappa\theta = 0$$

$$\Rightarrow \partial_t q_0(t,T) = w\mu - q_{\tilde{V}}(t,T)\kappa\theta$$

- for coefficients of $\tilde{\lambda}_t$:

$$\partial_t q_{\tilde{\lambda}}(t,T) - \mathbb{E}(J) q_{\tilde{V}}(t,T) + \exp(\sum_{l=0}^{n-1} q_{\tilde{\lambda}}(t,T)\zeta w_l + q_l(t,T)) \mathbb{E}(\exp(q_{\tilde{V}}(t,T) J)) - 1 = 0$$

$$\Rightarrow \partial_t q_\lambda(t,T) = \mathbb{E}(J) q_{\tilde{V}}(t,T) - \exp(\sum_{l=0}^{n-1} q_{\tilde{\lambda}}(t,T)\zeta w_l + q_l(t,T))\mathbb{E}(\exp(q_{\tilde{V}}(t,T)J)) + 1$$

- for coefficients of $\tilde{V}_t$:

$$\frac{w}{2} - \kappa q_{\tilde{V}}(t,T) + \frac{w^2}{2} + \frac{q_{\tilde{V}}^2(t,T)\sigma^2}{2} - \rho\sigma q_{\tilde{V}}(t,T)w + \partial_t q_{\tilde{V}}(t,T) = 0$$

$$\Rightarrow \partial_t q_{\tilde{V}}(t,T) = q_{\tilde{V}}(t,T)[\rho\sigma w + \kappa - \frac{q_{\tilde{V}}(t,T)\sigma^2}{2}] - \frac{w}{2}(1+w)$$

- for coefficients of $Y_t(l)$:

$$\partial_t q_l(t,T) - q_{\tilde{\lambda}}(t,T)\zeta w_l(v+b_l) - (v+b_l)q_l(t,T) = 0$$

$$\begin{aligned}\Rightarrow \partial_t q_l(t,T) &= (v+b_l)(q_\lambda(t,T)\zeta w_l + q_l(t,T)) \\ &= -\int_t^T (v+b_l)\exp(-(v+b_l)(u-t))\zeta w_l \partial_t q_{\tilde{\lambda}}(t,u)du\end{aligned}$$

When considering the rough Hawkes kernel, Galerkin approximation makes the solution of Backward equation diverge. Then, for the same considerations than in Chapter 1, we consider the forward Kolmogorov equations. We invert the equations using the property proved in Hainaut [3] :

$$\partial_s q_*(t,s) = -\partial_t q_*(t,s),\ q_*(t,s) = q_*(s-t).$$

- For the constant terms :

$$\partial_s q_0(t,s) = -w\mu + q_{\tilde{V}}(t,s)\kappa\theta;$$

- For the coefficients $\tilde{V}_t$:

$$\partial_s q_{\tilde{V}}(t,s) = -q_{\tilde{V}}(t,s)[\rho\sigma w + \kappa - \frac{q_{\tilde{V}}(t,s)\sigma^2}{2}] + \frac{w}{2}(1+w);$$

- For the coefficients $Y_t(l)$:

$$\partial_s q_l(t,s) = -\int_t^s (v+b_l)\exp(\alpha(u-t))\zeta w_l \partial_u q_{\tilde{\lambda}}(u,s)du$$

After a change of variable $r = s-u$,

$$\partial_s q_l(t,s) = -\int_0^{s-t} (v+b_l)\exp(-(v+b_l)(s-r-t))\zeta w_l \partial_r q_{\tilde{\lambda}}(r)dr;$$

- For the coefficients $\tilde{\lambda}$:

$$\partial_s q_\lambda(s-t) = -\mathbb{E}(J) q_{\tilde{V}}(t,s) + \exp(\sum_{l=0}^{n-1} q_{\tilde{\lambda}}(t,s)\zeta w_l + q_l(t,s))\mathbb{E}(\exp(q_{\tilde{V}}(t,s)J)) - 1;$$

We notice that the argument of the exponential, $\sum_{l=0}^{n-1} q_{\tilde{\lambda}}(t,s)\zeta w_l + q_l(t,s)$, is equal to $\sum_{l=0}^{n-1} \frac{\partial_t q_l(t,s)}{v+b_l} = \sum_{l=0}^{n-1} \frac{-\partial_s q_l(t,s)}{v+b_l}$. When n tends to ∞, we recover the same convergence result proved in Chapter 2 (2.1):

$$\lim_{n\to\infty} -\sum_{l=0}^{n-1} \frac{\partial_s q_l(t,s)}{v+b_l} = \zeta(K\frac{dq_\lambda}{ds})_s.$$

Finally, the mgf can be computed as

$$\mathbb{E}(\exp(-wX_s) \mid \mathcal{F}_0) = \exp([-w\mu s + \kappa\theta \int_0^s q_V(u)du] + q_\lambda(s)\lambda_0 + q_V(s)V_0)$$

where

$$\partial_s q_\lambda(s) = -\mathbb{E}(J)q_V(s) + \exp(\zeta(K\frac{dq_\lambda}{ds})_s)\mathbb{E}(\exp(q_V(s)J)) - 1;$$

$$\partial_s q_V(s) = -q_V(s)[\rho\sigma w + \kappa - \frac{q_V(s)\sigma^2}{2}] + \frac{w}{2}(1+w)$$

□

4.2 Mutual rough Hawkes jump in Heston volatility model

When applying the framework, the computations in this section are exactly the same than those in previous section for the rough Hawkes-Heston volatility model 4.4. Only the dynamic of the log-returns has been changed with one additional jump term 4.5. In consequence, I only highlight the changes in the formula instead of repeating all equations in the proof. The final result is presented in 4.2.

Log-return mgf with the mutual rough Hawkes jump in Heston volatility model

$$\mathbb{E}(\exp(-wX_s) \mid \mathcal{F}_0) = \exp([-w\mu s + \kappa\theta \int_0^s q_V(u)du] + q_\lambda(s)\lambda_0 + q_V(s)V_0 - wX_0)$$

where

$$\partial_s q_\lambda(s) = -(w\delta + q_V(s))\mathbb{E}(J) + \exp(\zeta(K\frac{dq_\lambda}{ds})_s)\mathbb{E}(\exp(J(q_V(s) - w\delta))) - 1;$$

$$\partial_s q_V(s) = -q_V(s)[\rho\sigma w + \kappa - \frac{q_V(s)\sigma^2}{2}] + \frac{w}{2}(1+w)$$

Proof. **4.2.1 Step 1**

The dampened version of $k(-)$ has a spectral representation via Fourier transform, then we can embed all the memory of the system in an infinity of new state variables. This is represented by the continuous process $\{Y_t(u)\}_{u\in\mathbb{R}_+}$. However, because the kernel has a pole at $h = 0$, we consider the unilateral transform. This is why $Y_t(u)$ is defined on $\mathbb{R}_+$.

4.2.2 Step 2

After the unilateral finite-dimensional approximation procedure (See the procedure in previous chapter 1.2), we obtain the following system :

$$\begin{cases} d\tilde{X}_t = (\mu - \frac{\tilde{V}_t^2}{2})dt + \sqrt{\tilde{V}_t}dW_t^X - \delta(d\tilde{L}_t - \tilde{\lambda}_t\mathbb{E}(J)dt) \\ d\tilde{V}_t = \kappa(\theta - \tilde{V}_t)dt + \sigma\sqrt{\tilde{V}_t}dW_t^V + d\tilde{L}_t - \tilde{\lambda}_t\mathbb{E}(J)dt \\ d\tilde{\lambda}_t = \zeta\sum_{l=0}^{n-1} w_l d\tilde{Y}_t(l) \\ d\tilde{Y}_t(l) = -(v + b_l)\tilde{Y}_t(l)dt + d\tilde{N}_t \end{cases} \tag{4.13}$$

4.2.3 Step 3

We apply the same Itô's calculus as in the previous case, except that the drift of $\tilde{X}_t$ is modified by the jump compensator:

$$\begin{cases} \mathbb{E}(\frac{d\tilde{X}_t}{dt} \mid \mathcal{F}_t) = \mu - \frac{\tilde{V}_t}{2} + \delta\tilde{\lambda}_t\mathbb{E}(J) \\ \mathbb{E}(\frac{d\tilde{V}_t}{dt} \mid \mathcal{F}_t) = \kappa(\theta - \tilde{V}_t) - \tilde{\lambda}_t\mathbb{E}(J) \\ \mathbb{E}(\frac{d\tilde{\lambda}_t}{dt} \mid \mathcal{F}_t) = \zeta\sum_{l=0}^{n-1} w_l\alpha Y_t(l) \\ \mathbb{E}(\frac{dY_t(l)}{dt} \mid \mathcal{F}_t) = -(v + b_l)Y_t(l) \end{cases}$$

In consequence, the formula for the infinitesimal generator 4.10 becomes:

$$\begin{aligned} \mathbb{E}(\frac{df}{dt} \mid \mathcal{F}_t) = \partial_t f + \partial_{\tilde{X}} f\left[\mu - \frac{\tilde{V}_t}{2} + \delta\tilde{\lambda}_t\mathbb{E}(J)\right] + \partial_{\tilde{V}} f\left[\kappa(\theta - \tilde{V}_t) - \tilde{\lambda}_t\mathbb{E}(J)\right] - \partial_{\tilde{\lambda}} f\left[\zeta\sum_{l=0}^{n-1} w_l(v + b_l)Y_t(l)\right] \\ - \sum_{l=0}^{n-1}\partial_{Y(l)} f(v + b_l)Y_t(l) + \frac{\tilde{V}_t}{2}\partial_{\tilde{X}\tilde{X}} f + \frac{\sigma^2\tilde{V}_t}{2}\partial_{\tilde{V}\tilde{V}} f + \rho\sigma\tilde{V}_t\partial_{\tilde{X}\tilde{V}} f \\ \underbrace{+\tilde{\lambda}_t\int_0^t \left[f(\tilde{X}_t - \delta z, \tilde{V}_{t^-} + z, \tilde{\lambda}_t + \zeta\sum_{l=0}^{n-1} w_l, Y_t(l) + 1, \tilde{N}_t + 1) - f\right]\tilde{\nu}(dz)}_{\mathcal{A}f(-)} = 0. \end{aligned} \tag{4.14}$$

We observe that both the drift of $d\tilde{X}_t$ and the jump integral have been influenced by the mutual jump dynamic. Therefore, it also has consequence on the simplification of the integral term. In consequence,

$$\begin{aligned} &\int_0^t \left[f(\tilde{X}_t - \delta z, \tilde{V}_{t^-} + z, \tilde{\lambda}_t + \zeta\sum_{l=0}^{n-1} w_l, Y_t(l) + 1, \tilde{N}_t + 1) - f\right]\tilde{\nu}(dz) \\ &= u(t,x)\left[\int_0^t \exp((q_{\tilde{V}}(t,T) - w\delta)z + \zeta\sum_{l=0}^{n-1} w_l q_{\tilde{\lambda}}(t,T) + q_l(t,T))\tilde{\nu}(dz) - 1\right] \\ &= f\left[\exp(\zeta\sum_{l=0}^{n-1} w_l q_{\tilde{\lambda}}(t,T) + q_l(t,T))\int_0^t \exp((q_{\tilde{V}}(t,T) - w\delta)z)\tilde{\nu}(dz) - 1\right] \\ &= f\left[\exp(\zeta\sum_{l=0}^{n-1} w_l q_{\tilde{\lambda}}(t,T) + q_l(t,T))\mathbb{E}(\exp((q_{\tilde{V}}(t,T) - w\delta)J) - 1\right] \end{aligned} \tag{4.15}$$

4.2.4 Step 4

When identifying coefficients of each variable, we observe that only the jump parameter $q_{\tilde{\lambda}}$ has been changed, without any surprise since the additional jump term in $\tilde{X}_t$ dynamic depends on $\tilde{\lambda}_t$:

$$\partial_t q_{\tilde{\lambda}}(t,T) = (w\delta + q_V(t,T))\mathbb{E}(J) + 1 - \exp(\sum_{l=0}^{n-1} q_l(t,T) + \zeta w_l q_{\tilde{\lambda}}(t,T))\mathbb{E}(\exp(J(q_V(t,T) - w\delta)))$$

For the same convergence constraint than before when considering the rough Hawkes kernel, we invert the system into its forward form :

$$\partial_s q_{\tilde{\lambda}}(t,s) = -(w\delta + q_V(t,T))\mathbb{E}(J) - 1 + \exp(\sum_{l=0}^{n-1} q_l(t,s) + \zeta w_l q_{\tilde{\lambda}}(t,s))\mathbb{E}(\exp(J(q_V(t,s) - w\delta)))$$

Finally, all things kept identical to the previous case except the differential equation for $\tilde{\lambda}_t$ (4.2.4), mgf can be directly computed:

$$\mathbb{E}(\exp(-wX_s) \mid \mathcal{F}_0) = \exp([-w\mu s + \kappa\theta \int_0^s q_V(u)du] + q_\lambda(s)\lambda_0 + q_V(s)V_0)$$

where

$$\partial_s q_\lambda(s) = -(w\delta + q_V(s))\mathbb{E}(J) + \exp(\zeta(K\frac{dq_\lambda}{ds})_s)\mathbb{E}(\exp(J(q_V(s) - w\delta))) - 1;$$

$$\partial_s q_V(s) = -q_V(s)[\rho\sigma w + \kappa - \frac{q_V(s)\sigma^2}{2}] + \frac{w}{2}(1+w)$$

□

We will only consider this last mutual rough Hawkes-Heston volatility model in the following, as the previously presented rough Hawkes-Heston volatility model is a specific case corresponding to $\delta = 0$.

4.3 Change of measure

In this section, change of measure from market real measure $\mathbb{P}$ to a chosen equivalent measure from the family of risk-neutral measure $\mathbb{Q}$ is performed the same way we did in chapter 2. Only Heston model with leverage effect on stock price δ is considered, hence the first model is a specific case for $\delta = 0$. The main result 4.3 remains the same than in the uni-dimensional space.

In addition to the parameter γ that determines the family of risk-neutral measures for the jumps, a second parameter θ^{VV} appears in 4.3, representing the virtual risk premium for the volatility process. This parameter introduces a second degree of liberty, but is generally set to zero as volatility is mostly not directly tradable. Consequently, the choice $\theta^{VV} = 0$ is reasonable for selecting a single risk measure $\mathbb{Q}^{\gamma,\theta^{VV}}$.

For exponential or double-exponential distribution of jump magnitude J under $\mathbb{P}$, the choice $\gamma = 0$ remains valid.

Hence, the simplest change of measure from $\mathbb{P}$ is towards $\mathbb{Q}^{0,0}$, denoted $\mathbb{Q}$ for convenience. The change of measure only impacts the drift of the asset dynamic, which grows at risk-free rate under any measure in the family $\mathbb{Q}^{\gamma,\theta^{VV}}$.

rough Hawkes-Heston model under risk-neutral measure

For rough Hawkes jumps incorporated in Heston volatility model 4.5, Brownian and jump parts can be analyze separately:

- Brownian motions are shifted according to Girsanov's theorem, the shifts representing the risk premium according to the measure $\mathbb{P}$:

$$\theta_t^X = \frac{\mu - r}{\sqrt{V_t}}, \quad \kappa^{\mathbb{Q}^{\gamma,\theta^{VV}}} = \kappa + \sigma\theta^{VV}, \quad \theta^{\mathbb{Q}^{\gamma,\theta^{VV}}} = \frac{\kappa\theta}{\kappa + \sigma\theta^{VV}},$$

  where r is the risk-free rate and θ^{VV} is a risk premium parameter in volatility.

- jump component undergoes the following transform (Esscher transform, more information in [6]):

$$\mathbb{E}^{\mathbb{Q}^{\gamma,\theta^{VV}}}(\exp(-wJ^{\mathbb{Q}^{\gamma,\theta^{VV}}})) = \frac{\mathbb{E}(\exp((\gamma - w)J))}{\mathbb{E}(\exp(J\gamma))}, \ \lambda_t^{\mathbb{Q}^{\gamma,\theta^{VV}}} = \mathbb{E}(\exp(J\gamma))\lambda_t.$$

Proof. We split the analysis between the diffusive and the jump parts.

Brownian part In this part, the application of the Girsanov's theorem is less trivial than in the unidimensaional case because the model is driven by two correlated brownian motions. Introducing the vector of market prices of risk:

$$\theta_t = \begin{pmatrix} \theta_t^S \\ \theta_t^V \end{pmatrix},$$

the Radon–Nikodym derivative is defined by the Doléans–Dade exponential:

$$\eta_t^{\text{Brownian}} = \frac{d\mathbb{Q}^{\gamma,\theta^{VV}}}{d\mathbb{P}}\Big|_{\mathcal{F}_t} = \exp\left(-\int_0^t \lambda_s^\top dW_s^{\mathbb{P}} - \frac{1}{2}\int_0^t \|\lambda_s\|^2 ds\right).$$

By Girsanov's theorem, the processes

$$\begin{cases} dW_t^{X,\mathbb{Q}^{\gamma,\theta^{VV}}} = dW_t^X + \theta_t^X dt \Rightarrow dW_t^X = dW_t^{X,\mathbb{Q}^{\gamma,\theta^{VV}}} - \theta_t^X dt, \\ dW_t^{V,\mathbb{Q}^{\gamma,\theta^{VV}}} = dW_t^V + \theta_t^V dt \Rightarrow dW_t^V = dW_t^{V,\mathbb{Q}^{\gamma,\theta^{VV}}} - \theta_t^V dt \end{cases}$$

are Brownian motions under $\mathbb{Q}^{\gamma,\theta^{VV}}$. We incorporate the first Brownian motion is the stock model:

$$dS_t = \underbrace{(\mu - \sqrt{V_t}\theta_t^X)}_{r} S_t dt + S_t\sqrt{V_t} dW_t^{X,\mathbb{Q}^{\gamma,\theta^{VV}}} \tag{4.16}$$

Under the risk-neutral measure, discounted asset prices must be martingales. Therefore, the drift in (4.16) must equal the risk-free rate r:

$$\mu - \sqrt{V_t}\theta_t^X = r \Rightarrow \theta_t^X = \frac{\mu - r}{\sqrt{V_t}}.$$

We process similarly for the volatility process:

$$dV_t = \left[\kappa(\theta - V_t) - \sigma\sqrt{V_t}\theta_t^V\right] dt + \sigma\sqrt{V_t} dW_t^{V,\mathbb{Q}^{\gamma,\theta^{VV}}}.$$

Parameters under $\mathbb{Q}^{\gamma,\theta^{VV}}$ for the CIR process are obtained by observing, given a choice $\theta_t^V = \sqrt{V_t}\theta^{VV}$, how a similar mean-reverting structure can be recovered:

$$\kappa(\theta - V_t) - \sigma\sqrt{V_t}\theta_t^V = \underbrace{(\kappa + \sigma\theta^{VV})}_{\kappa^{\mathbb{Q}^{\gamma,\theta^{VV}}}}[\underbrace{\frac{\kappa\theta}{\kappa + \sigma\theta^{VV}}}_{\theta^{\mathbb{Q}^{\gamma,\theta^{VV}}}} - V_t].$$

Jump part For the jump part, the counterpart of the Doleans-Dale exponential is

$$\eta_t^{Jump} = \exp(\gamma\tilde{L}_t + (1 - \mathbb{E}(\exp(J\gamma))\int_0^t \tilde{\lambda}_s ds).$$

Applying step 3 and 4 of the affine Volterra procedure on

$$\mathbb{E}(\exp(-w\tilde{X}_s + \log(\eta_s^{Jump})) \mid \mathcal{F}_t),$$

we recover the same dynamics up to an Esscher transform of the jump magnitude distribution J (following exactly the same procedure than in 2.3). □

Chapter 5

Illustrations : VIX

In this chapter, I present the classical mean to calibrate the variables of System 4.5 under risk-neutral measure, using European call volatility surface. This method has not been implemented, but I explored an approximated calibration under $\mathbb{P}$.

Based on these calibrated parameters, I demonstrate through the VIX example that the rough Hawkes-Heston has the potential to replicate a volatility smile in the European call surface.

5.1 Calibration under risk-neutral measure

The calibration of the rough Hawkes-Heston models should intuitively be performed the same way than for the classical Heston model. As a reminder, the Heston model takes the following Markovian form under the risk-neutral measure:

$$\begin{cases} dX_t = \left(r - \frac{V_t}{2}\right) dt + \sqrt{V_t}\, dW_t^X, \\ dV_t = \kappa(\theta - V_t)\, dt + \sigma\sqrt{V_t}\, dW_t^V, \\ \mathrm{Corr}(dW_t^X, dW_t^V) = \rho. \end{cases}$$

Since this model is affine and Markovian, it directly admits a closed-form expression for its moment generating function (mgf), characterized by classical Riccati equations.

Similarly to what was presented in the first chapter, we obtain the mgf by imposing an exponentially affine structure:

$$\phi(w,t,T) = \mathbb{E}\left(\exp(wX_T) \mid \mathcal{F}_t\right) = \exp\left(A(w,t,T) + B(w,t,T)V_t + wX_t\right).$$

In this setting, the function $\phi(w,t,T)$ satisfies the forward and backward Kolmogorov equations (also known as the Feynman-Kac PDE):

- Backward equation:

$$\mathbb{E}(\frac{d\phi}{dt} \mid \mathcal{F}_t) = \mathcal{A}\phi(w,t,T) + \partial_t\phi(w,t,T) = 0$$

- Forward equation:

$$\mathcal{A}\phi(w,t,T) - \partial_t\phi(w,t,T) = 0$$

Using the same Markovian procedure we used massively in this master's thesis, we obtain the following Ricatti backward equations, and their closed-form expression :

$$\begin{cases} \frac{\partial A(w,t,T)}{\partial t} = -wr - \kappa\theta B(w,t,T) \\ \frac{\partial B(w,t,T)}{\partial t} = \frac{w}{2} + B(w,t,T)[\kappa - \rho\sigma w] - \frac{\sigma^2}{2}B^2(w,t,T) \\ A(w,T,T) = B(w,T,T) = 0 \end{cases} \tag{5.1}$$

$$\Leftrightarrow \begin{cases} A(\omega,t,s) = r\omega(s-t) + \frac{\kappa\gamma}{\sigma^2}\left[(\kappa-\rho\sigma\omega+d)(s-t) - 2\ln\left(\frac{1-ge^{d(s-t)}}{1-g}\right)\right] \\ B(\omega,t,s) = \frac{\kappa-\rho\sigma\omega+d}{\sigma^2}\,\frac{1-e^{d(s-t)}}{1-ge^{d(s-t)}} \end{cases} \tag{5.2}$$

where

$$d = \sqrt{(\rho\sigma\omega-\kappa)^2 + \sigma^2(\omega-\omega^2)}, \quad g = \frac{\kappa-\rho\sigma\omega+d}{\kappa-\rho\sigma\omega-d}.$$

In previously developed calibration tool for the Heston model, as in [7], parameters $\Theta = \{\kappa, \theta, \sigma, V_0\}$ are generally obtained via Heston raw prices RMSE, that is

$$\hat{\Theta} = \arg\min_{\Theta} \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\text{Call}_i^{model}(K_i,T_i;\Theta) - \text{Call}_i(K_i,T_i)\right]^2}. \tag{5.3}$$

However, in practice, this criterion is often reformulated in terms of implied volatility errors. Given that option prices are nonlinear functions of volatility through the Black–Scholes mapping, calibration in price space may overweight deep in-the-money or high-premium options and underrepresent the shape of the volatility surface.

As a result, a more stable approach consists in minimizing the discrepancy between model-implied and market-implied volatilities is generally:

$$\hat{\Theta} = \arg\min_{\Theta} \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\sigma_{imp}^{model}(K_i,T_i;\Theta) - \sigma_{imp}^{market}(K_i,T_i)\right]^2}, \tag{5.4}$$

where σ_{imp}^{market} is obtained from implied volatility surface, and σ_{imp}^{model} is computed by Black-Scholes Call price inversion. Pragmatically, the procedure has already been described in Chapter 3 3.5. After DFFT on $\phi(.,.,.)$, we compute the value of a call in the Heston setting for different strikes K_i and maturities T_i. The couples (K_i, T_i) should correspond to the ones available for the volatility surface, but additional points may be introduced by interpolation to stabilize the calibration. Then, the value of the Heston call options are equalized to the Black-Scholes theoretical call formula, with Black implied volatility as unique parameter. Inverting these formula for each couple gives the model induced Black volatility surface.

Nevertheless, the extension to the rough Hawkes-Heston model 4.5 is not straightforward. In comparison, we have a larger set of parameters to calibrate:

$$\Theta = \{ \underbrace{\kappa,\theta,\sigma,V_0}_{\text{CIR vol dynamic}} , \underbrace{\zeta,\lambda_0,\alpha}_{\text{rough parameters}} , \delta\} \cup \text{ set of parameters for jump } J \text{ distribution}$$

Calibration of such a large number of parameters may become very unstable.

Further research should be dedicated to this calibration. The extension of known algorithms for Heston calibration, such as Neural Networks [8] and Heston raw prices RMSE [7] could be studied following this master's thesis. In the specific case of the SPX analysis, since VIX options prices are available, and

$$VIX \approx 100\sqrt{\frac{1}{T}\mathbb{E}^{\mathbb{Q}}\left(\int_0^T V_s ds\right)}, \tag{5.5}$$

one could imagine a double objective function, minimizing the gap with both stock implied volatilities and VIX implied volatilities. The loss function becomes then:

$$\sum_{i=1}^{N}(\sigma_{imp}^{model,SPX}(K_i,T_i;\Theta) - \sigma_{imp}^{market,SPX}(K_i,T_i))^2 + \sum_{j=1}^{M}(\sigma_{imp}^{model,VIX}(K_i,T_i;\Theta) - \sigma_{imp}^{market,VIX}(K_i,T_i))^2.$$

Ideally, all these optimization statement should be led under the constraint that $2\kappa\theta > \sigma^2$ for ensuring positivity of volatility. I would be pleased to be informed if such research on that subject is once led.

5.2 Alternative calibration under $\mathbb{P}$

I suggest a pragmatic calibration under $\mathbb{P}$ in this section. The main inconsistency in this modeling is that the VIX is constructed on risk-neutral expectation (cfr Equation (5.5)), and I include it in my real-world measure $\mathbb{P}$. Nevertheless, this remark could be outweigh when the risk-premium for volatility θ^{VV} is zero, as discussed in application of Theorem 4.3. Mainly, the technique splits the calibration, in order to ensure stability:

1. First of all, we estimate the latent variance V_t. This can be performed in different settings:

   - We can consider the realized variance $\hat{V}_t \approx (\Delta X_t)^2$, where
   $$\Delta X_t = \log(\frac{S_t}{S_{t-1}}).$$
   - Estimation through Kalman or particle filter based on the observed ΔX_t could also be led;
   - In the specific case of SPX, I consider $\hat{V}_t \approx (\frac{VIX_t}{100})^2$ in view of Relation 5.5.

   I opt for the last case by simplicity, since the VIX is directly observed on Bloomberg.

2. Then, we estimate the CIR parameters via MLE. The idea of this second step arises during search for applying a similar method than peak-over-threshold on $\hat{V}_t$. However, CIR processes are not normally distributed like GBM. CIR processes are conditionally distributed as a non-centered chi-squared [9]. Only conditional expectation and variance can be computed, then the peak-over-threshold is killed for this application. More detailed about this distribution are shown in the Appendix 7.1. Nevertheless, CIR parameters can conditionally be calibrated through the non-centered Chi-squared log-likelihood maximization on the estimated volatility time series $\hat{V}_t$:
   $$(\hat{\kappa}, \hat{\theta}, \hat{\sigma}) = \arg\max_{\kappa,\theta,\sigma} \sum_t \log(f(\hat{V}_{t+1} \mid \hat{V}_t)). \tag{5.6}$$
   Therefore, in this calibration procedure, we initially assume that there are no jumps in the volatility dynamics, since the residuals of the mean-reverting model will subsequently be explained by jumps in a second stage, during the third step. This first calibration step gives the CIR parameters, Table 5.1.

Parameter	Estimated value
κ	0.7012
θ	0.3784
σ	0.5142
Log-likelihood	-4864.0587

Table 5.1: Estimated Heston parameters obtained by maximum likelihood calibration.

3. Thirdly, we estimate the jumps in the volatility, so that they explain large residuals between time series $\hat{V}_t$ and CIR model predictions. We make the distinction between the process $\hat{V}_t$ estimated in first step, and the simulated paths $\hat{V}_t^{cir}$ from the model calibrated in the second step. We cannot apply a similar technique than the peak over threshold here because CIR processes are not stationary : they only have a closed-form for their conditional expectation and variance (cfr the Appendix 7.1 for their expressions).

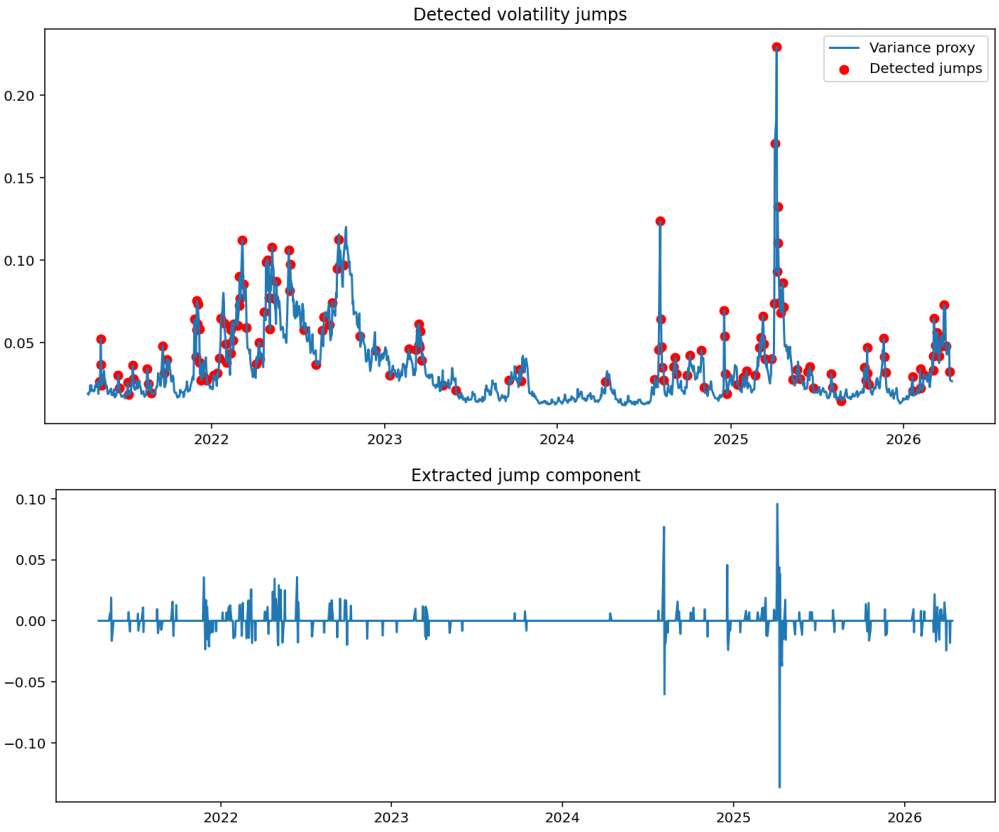

Figure 5.1: Detected jump by the calibration.

We define :

$$\epsilon_t = \frac{\hat{V}_{t+1} - \mathbb{E}(\hat{V}_{cir}^{mod} \mid \hat{V}_t^{cir})}{\sqrt{\mathbb{V}(\hat{V}_{t+1}^{cir} \mid \hat{V}_t^{cir})}},$$

so that jumps explain large deviance from the model :

$$\Delta L_t = \begin{cases} \hat{V}_{t+1} - \mathbb{E}(\hat{V}_{t+1}^{cir} \mid \hat{V}_t^{cir}) \text{ if } |\epsilon_t| \geq c \\ 0 \text{ otherwise} \end{cases}.$$

The best value for c is chosen by cross-validation, such that the sample of residuals within the threshold interval $\mathcal{E} = \{\epsilon_t \mid |\epsilon_t| < c\}$ is normally distributed. We obtain $\hat{c} = 1.2857$, for which residuals presents a reduced skewness and excess kurtosis:

$$\mathbb{K}(\mathcal{E}) = 2.9864, \mathbb{S}(\mathcal{E}) = 0.4845.$$

This already provides an improvement towards normality of residuals, since the initial skewness and kurtosis of residuals ϵ_t are :

$$\mathbb{K}(\epsilon_t) = 31.8685, \mathbb{S}(\epsilon_t) = 2.5965.$$

The detected jumps in volatility are shown in Figure 5.1.

4. Finally, the correlation ρ between the residuals is computed empirically:

$$\hat{\epsilon}_t^S = \frac{\Delta X_t}{\sqrt{\hat{V}_t}}, \quad \hat{\epsilon}_t^V = \frac{\Delta V_t}{\sqrt{\hat{V}_t}} \Rightarrow \hat{\rho} = \frac{\sum_t \hat{\epsilon}_t^S \hat{\epsilon}_t^V}{\sqrt{\sum_t (\hat{\epsilon}_t^S)^2}\sqrt{\sum_t (\hat{\epsilon}_t^V)^2}} = -0.62142.$$

The negative value corresponds well to the intuition.

5. In case we consider the mutually exciting rough Hawkes-Heston statement (4.5), the damping parameter δ can be calibrated via the least square minimization between observed log-returns and model:

$$\hat{\delta} = \arg\min_{\delta} \sum_{t} (\log(S_t) - X_t(\delta))^2, \text{ where} \tag{5.7}$$

$$X_t(\delta) = X_{t-1}(\delta) + \underbrace{\left(\mu - \frac{\hat{V}_{t-1}}{2}\right)\Delta t}_{\text{drift}} + \underbrace{\sqrt{\hat{V}_{t-1}}\, Z_t \sqrt{\Delta t}}_{\text{diffusion}} - \underbrace{\delta\, \Delta L_t}_{\text{jump effect}},$$

where Z is distributed as a standard normal. This first minimization problem is a simple OLS, for which we have a closed-form solution:

$$\hat{\mu} = \frac{1}{\Delta t} \cdot \frac{1}{N} \sum_{t=1}^{N} \Delta X_t, \quad \hat{\delta} = -\frac{\sum_t \left(\Delta X_t - \left(\hat{\mu} - \frac{1}{2} V_t\right) \Delta t\right) \Delta L_t}{\sum_t (\Delta L_t)^2} = 0.785$$

Otherwise, similarly to the peak-over-threshold method, we could consider the $\hat{\delta}$ such that $\log(S_t) + \delta \Delta L_t$ is the most normally distributed, that is:

$$\hat{\delta} = \arg\min_{\delta} (\mathbb{K}(\{\log(S_t) + \delta \Delta L_t\}) - 3)^2 + \mathbb{S}^2(\{\log(S_t) + \delta \Delta L_t\}). \tag{5.8}$$

This second method gives $\hat{\delta} = 0.781$.

6. Now that jumps have been isolated, one can perform calibration of jump severity and estimate the parameters of J exactly the same way we did in Chapter 3 and presented in 2.2, using Log-Likelihood Maximization. The calibrated parameters are presented in Table 5.2.

Parameter	Value
α	0.99745
ζ	0.05787
v	0.14498
λ_0	0.06929
λ_∞	0.11495
η_+	60.8861
η_-	62.9143
p	0.5223
Log-likelihood	-381.9036

Table 5.2: Calibration results for the double exponential jump model.

5.3 Call pricing

We apply the same methodology as in Chapter 3 (3.5). In Figures 5.2 and 5.3, we display the implied volatility surfaces for both the rough Hawkes–Heston model ($\delta = 0$) (4.4) and the mutually exciting rough stock–volatility model within the rough Hawkes–Heston framework (4.5).

We observe that both models are able to reproduce the volatility smile, with the minimum implied volatility attained approximately at-the-money for the call option.

The implied volatility surface ranges between 10% and 50%. The smile effect appears to be more pronounced in the second setting (4.5).

We also verify the properties of the probability density function recovered through the DFFT algorithm; see Table 5.3. The 90% confidence interval widens up to $[-29.8, 25]\%$.

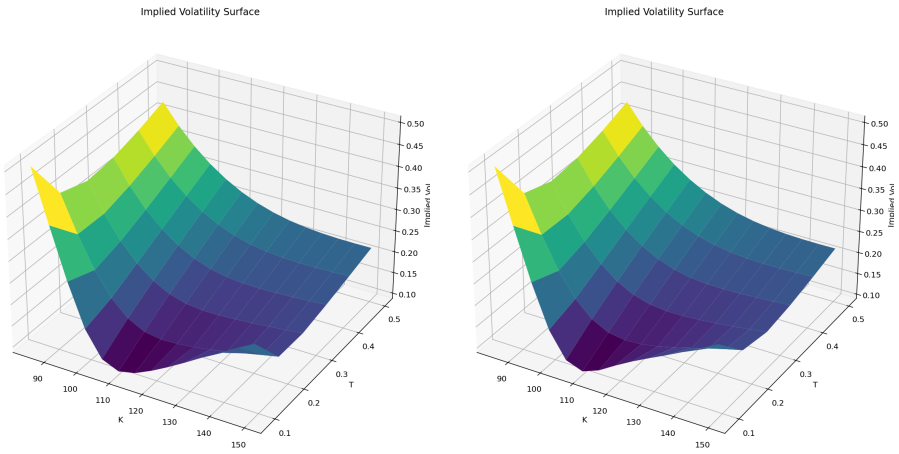

Figure 5.2: European Call Volatility surfaces obtained from Heston 4.4 (on the left) and Heston 4.5 (on the right), for $M = 2^9$, $x\_\max = 2.5$, $V_0 = r = \sigma = 0$, $S_0 = 100$ and other parameters values taken from the calibration.

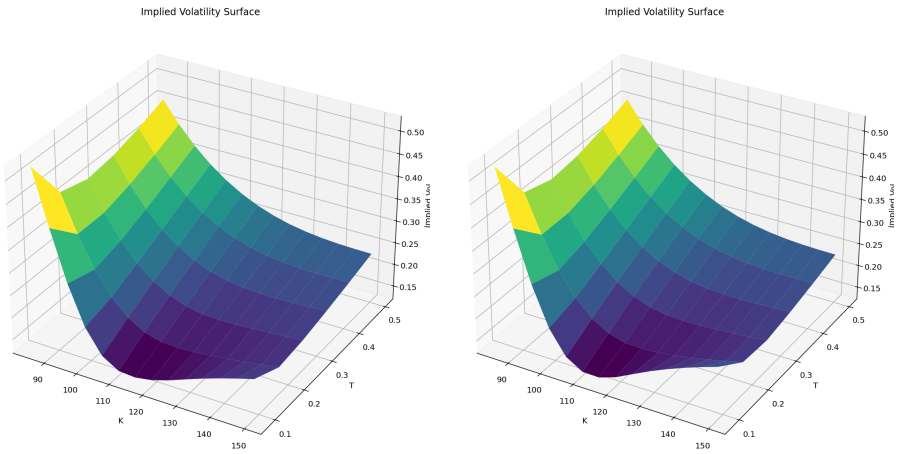

Figure 5.3: European Call Volatility surfaces obtained from Heston 4.4 (on the left) and Heston 4.5 (on the right), for $M = 2^9$, $x\_\max = 2.5$, $V_0 = 0.01, r = \sigma = 0$, $S_0 = 100$ and other parameters values taken from the calibration.

Days	Expectation	Std. deviation	5%	95%
30	0.005	0.030	-0.064	0.044
60	0.007	0.059	-0.113	0.093
90	0.009	0.087	-0.161	0.132
120	0.012	0.115	-0.210	0.171
150	0.015	0.142	-0.250	0.210
180	0.020	0.168	-0.298	0.250

Table 5.3: PDF statistics after DFFT for the mutual rough Hawkes Heston model 4.5, with truncature $x_{\max=2.5}$, $M = 2^9$ components, and $\sigma = r = 0$, $V_0 = 0$, $S_0 = 100$.

The numerical sensitivity analysis shown in Figures 5.4 and 5.5 highlights a economically consistent impact of the model parameters on the shape of the implied volatility surface. Each parameter influences a distinct feature of the surface, either by shifting its level or by modifying its curvature:

- First, an increase in the mean-reversion speed κ leads to an upward shift of the implied volatility surface across all strikes in our numerical results. A higher value of κ forces the variance process V_t to revert more rapidly toward its long-run mean θ.

  In the case where the initial variance satisfies $V_0 < \theta$, increasing κ accelerates the adjustment of volatility toward the higher long-run level, which results in an increase in implied volatility, particularly at short maturities.

  This behavior is not fully aligned with the standard intuition. The influence of the initial variance level V_0 is progressively dampened over time, and a larger mean-reversion speed κ is typically associated with a lower persistence of volatility shocks in a stationary regime, which generally tends to reduce implied volatility.

- Second, the volatility-of-volatility parameter σ has a heterogeneous impact across strikes. It mainly increases implied volatilities for out-of-the-money options, while having a much weaker or slightly negative effect on in-of-the-money options.

- Third, the long-run variance parameter θ induces an almost parallel upward shift of the entire implied volatility surface. Since θ defines the equilibrium level toward which the variance process mean-reverts, increasing θ raises the expected variance at all maturities. The effect is therefore essentially uniform across strikes, affecting the level of the surface rather than its shape.

- Finally, the correlation parameter ρ primarily affects the curvature and skew of the implied volatility surface.

We observe in Figure 5.6 that both models (4.4) and (4.5) produce similar out-of-the-money option values; however, the second specification (4.5) tends to generate higher in-the-money values. Consequently, the volatility smile appears more pronounced under this second setting.

For a three-month horizon, the implied volatility surface is shifted upward when increasing the initial variance V_0. However, the influence of V_0 is progressively dampened as the maturity increases, consistent with the mean-reverting nature of the variance process.

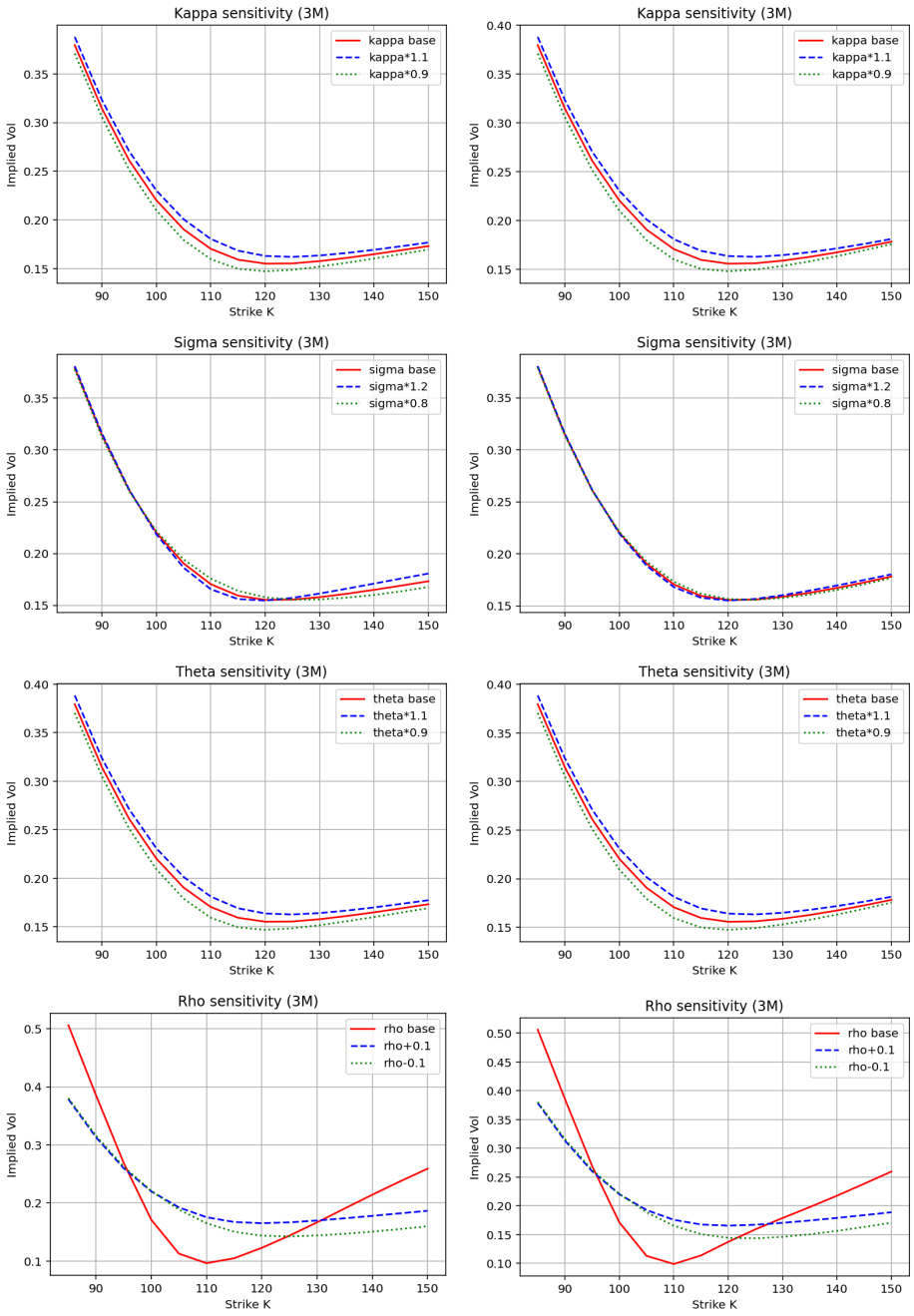

Figure 5.4: Sensitivity of the implied volatility ($T = 3$ months) to different parameters, for the Heston model 4.4 (on the left) and Heston model 4.5 (on the right). We observed nearly no sensitivity to rough process parameters (α, v, ζ), so here we show only significant impacts of parameters $(\kappa, \sigma, \theta, \rho)$, with initial $V_0 = 0$.

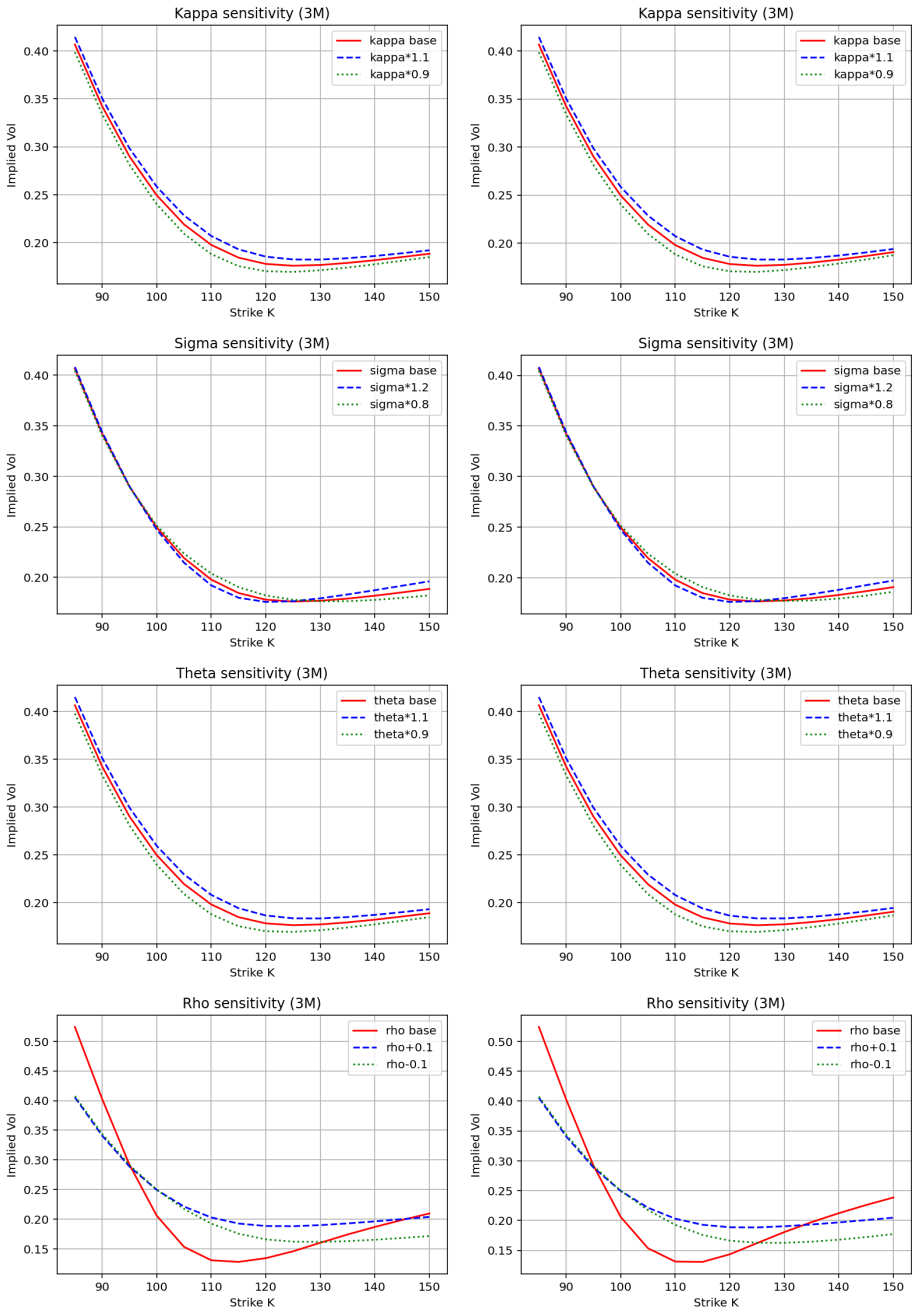

Figure 5.5: Sensitivity of the implied volatility ($T = 3$ months) to different parameters, for the Heston model 4.4 (on the left) and Heston model 4.5 (on the right). We observed nearly no sensitivity to rough process parameters (α, v, ζ), so here we show only significant impacts of parameters $(\kappa, \sigma, \theta, \rho)$, with initial $V_0 = 0.01$.

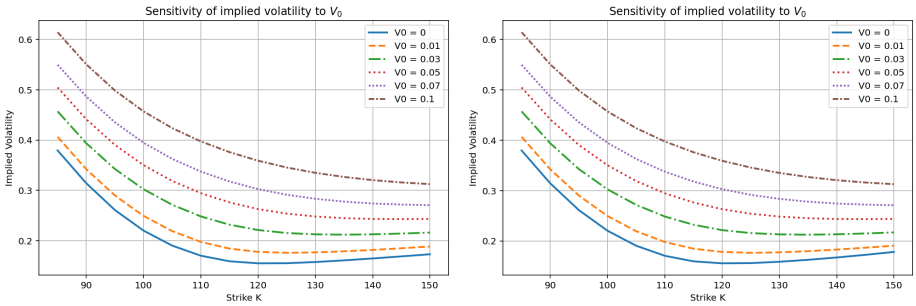

Figure 5.6: Sensitivity of the implied volatility ($T = 3$ months) to the parameter V_0, for the rough Hawkes-Heston model 4.4 (on the left) and mutually excited stock-volatility rough Hawkes-Heston model 4.5 (on the right).

5.4 Further considerations

In case these rough Hawkes-Heston models are too limited to capture the interaction between jumps in the volatility and the underlying, several extensions could be considered.

First of all, following empirical observations of links between jumps in volatility and underlying, one could still model a single intensity occurrence for both dynamics. However, the link in magnitude of jumps in both dynamics may differ more than through a single scaling parameter δ. Instead, one could consider different jump distributions for J, in the following model:

$$\begin{cases} dX_t = (\mu - \frac{V_t}{2})dt + \sqrt{V_t}dW_t^X - \sum_{k=1}^{N_t} J_{k,1} \\ dV_t = \kappa(\theta - V_t)dt + \sigma\sqrt{V_t}dW_t^V + \sum_{k=1}^{N_t} J_{k,2} \\ d\lambda_t = \zeta \int_{\mathbb{R}_+} dY_t(z)\mu(dz) \\ dY_t(z) = \alpha Y_t(z)dt + dN_t \\ N_t \sim \text{Poi}(\lambda_t), \quad J_{k,1} \sim J_1, \quad J_{k,2} \sim J_2 \\ \langle dW_t^V, dW_t^X \rangle = \rho dt \end{cases}$$

Secondly, motivated by the performance of the mutually exciting rough Hawkes jumps [1] in Chapter 3 in comparison to the uni-dimensional rough Hawkes model, one could introduce mutual jumps between the volatility process and its underlying, through a similar system than (3.16).

Chapter 6

Conclusion

Rough Hawkes models have already been presented in different paper in literature, notably by my supervisor Donatien Hainaut, in which its link with the affine Volterra analysis have been studied.

This thesis has confirmed that rough kernels may be less useful than expected for modeling equities. Hainaut [1] has drawn a promising conclusion for cryptocurrency log-returns. Nevertheless, I outweigh his conclusion, adding that rough kernels may be suitable if they are sufficiently flexible to capture the different memory patterns and dependencies of the process. I highlight the fact that modeling separate memory structure for positive and negative jumps is necessary for BTC log-returns, since unidimensional rough model fails providing better results than Markovian models.

I have provided a clear framework applicable to all rough kernels problems, covering the most cases I envisaged. We have argued that the affine Volterra setting has strong intuitive patterns.

Finally, I presented a meaningful pattern to include roughness in Heston model. This counterparty is the introduction of a lot of parameters we need to calibrate.

Calibration under the real-measure $\mathbb{P}$ has been investigated, and gives consistent values of parameters.

The mutually excited stock-volatility rough Hawkes-Heston model (4.5) has shown the potential to replicate the volatility smile.

Further research on the calibration based on implied volatilities in this setting should be led for SPX/SP500 at first, and for different underlying, for which the volatility is not observed.

Use of AI In this work, I occasionally used artificial intelligence tools (notably ChatGPT and GitHub Copilot) to rephrase non-scientific sentences, translate technical texts, and improve code formatting or translation. These tools were used solely as linguistic and technical assistance, without influencing methodological choices or scientific analysis.

Chapter 7

Appendix

7.1 Dynamic of the volatility under CIR process

This section is based on results presented in [9].

Two first conditional moments of variables modeled with CIR dynamic

Using the squared Bessel representation of the CIR process,

$$dy_t = \kappa(\theta - y_t)dt + \sigma\sqrt{y_t}dW_t,$$

it can be shown that the exact transition distribution is a scaled non-central chi-squared distribution:

$$y_{t+\Delta t} \mid y_t = x \sim c \cdot \chi'^2(d, h),$$

where

$$\begin{cases} d = \frac{4\kappa\theta}{\zeta^2}, \\ h = \frac{4\kappa e^{-\kappa\Delta t} y_t}{\zeta_x^2(1-e^{-\kappa\Delta t})}, \\ c = \frac{\zeta^2}{4\kappa}(1 - e^{-\kappa\Delta t}). \end{cases}$$

In this setting,

$$\mathbb{E}(y_{t+\Delta t} \mid y_t) = e^{-\kappa\Delta t} y_t^x + \theta\left(1 - e^{-\kappa\Delta t}\right) \text{ and}$$

$$\mathbb{V}(y_{t+\Delta t} \mid y_t) = \zeta^2 \left[\frac{y_t}{\kappa}\left(e^{-\kappa\Delta t} - e^{-2\kappa\Delta t}\right) + \frac{\theta}{2\kappa}\left(1 - e^{-\kappa\Delta t}\right)^2\right].$$

In consequence, the stationary (long-term) first two centered moments are constant :

$$\mathbb{E}(y_\infty) = \theta,$$
$$\mathbb{V}(y_\infty) = \frac{\zeta^2\theta}{2\kappa}.$$

Proof. Define the standardized variable

$$z_{t+1} = \frac{y_{t+\Delta t}}{c} \quad \Rightarrow \quad z_{t+1} \sim \chi'^2(d, h).$$

The exact likelihood function of the observed $\{y_0^x, \dots, y_n^x\}$ is given by

$$\mathcal{L}(\kappa, \theta, \zeta) = \prod_{i=0}^{n-1} f_Y\big(y_{i+1} \mid y_i\big),$$

and the corresponding log-likelihood is

$$\ell(\kappa, \theta, \zeta) = \sum_{i=0}^{n-1} \log f_Y\big(y_{i+1} \mid y_i\big).$$

Using the change of variable formula, the conditional density is

$$f_Y(y_{i+1} \mid y_i) = \frac{1}{c}\, f_{\chi'^2}\Big(\frac{y_{i+1}}{c}; d, h\Big),$$

where $f_{\chi'^2}(\cdot)$ denotes the density of the non-central chi-squared distribution.

Hence, the log-likelihood becomes

$$\ell(\kappa, \theta, \zeta) = \sum_{i=0}^{n-1} \Big[-\log c + \log f_{\chi'^2}\Big(\frac{y_{i+1}}{c}; d, h_i\Big)\Big].$$

Finally, the maximum likelihood estimator is defined as

$$(\hat{\kappa}, \hat{\theta}, \hat{\zeta}) = \arg\max_{\kappa,\theta,\zeta} \ell(\kappa, \theta, \zeta).$$

This estimator is exact, since it relies on the true transition density of the CIR process.

In comparison to OU/G1++ model which implies a normal distribution, CIR has a Chi-squared distribution (that is, essentially a normal squared distribution). In consequence, we are able to define a conditional mean and variance. In opposition to a normal distribution, the initial guess choice influences mean and variance in short-term. We would like to compute an exact expression for y_t and derive the two first conditional centered moments. Firstly, take

$$\begin{aligned} d(\exp(\kappa t) y_t) &= \kappa \exp(\kappa t) y_t dt + \exp(\kappa t) dy_t \\ &= \kappa \exp(\kappa t) y_t dt + \exp(\kappa t) \Big[\kappa(\theta - y_t)dt + \zeta\sqrt{y_t} dW_t\Big] \\ &= \kappa\theta \exp(\kappa t) dt + \zeta\sqrt{y_t}\exp(\kappa t) dW_t. \end{aligned}$$

Then, we take the integral on both sides between t and $t + \Delta t$. This gives :

$$y_{t+\Delta t} = \exp(-\kappa\Delta t) y_t + \theta(1 - \exp(-\kappa\Delta t)) + \zeta \int_t^{t+\Delta t} \exp(-\kappa(t + \Delta t - s))\sqrt{y_s} dW_s.$$

Consequently :

$$\begin{cases} \mathbb{E}(y_{t+\Delta t} \mid y_t) = e^{-\kappa\Delta t} y_t^x + \theta\left(1 - e^{-\kappa\Delta t}\right) \\ \mathbb{V}(y_{t+\Delta t} \mid y_t) = \mathbb{E}\left[\left(\zeta \displaystyle\int_t^{t+\Delta t} e^{-\kappa(t+\Delta t-s)}\sqrt{y_s}\, dW_s\right)^2 \middle| y_t\right] \\ \qquad\qquad = \mathbb{E}\left[\zeta^2 \displaystyle\int_t^{t+\Delta t} e^{-2\kappa(t+\Delta t-s)} y_s\, ds \middle| y_t\right] \\ \qquad\qquad = \zeta^2 \displaystyle\int_t^{t+\Delta t} e^{-2\kappa(t+\Delta t-s)} \mathbb{E}[y_s | \, y_t]\, ds \\ \qquad\qquad = \zeta^2 \displaystyle\int_t^{t+\Delta t} e^{-2\kappa(t+\Delta t-s)} \left[y_t e^{-\kappa(s-t)} + \theta\left(1 - e^{-\kappa(s-t)}\right)\right] ds \end{cases}$$

We split the integral defining the conditional variance in two parts I_1 and I_2 :

$$\mathbb{V}(y_{t+\Delta t} \mid y_t) = \zeta^2(I_1 + I_2).$$

The first one gives :

$$\begin{aligned} I_1 &= \int_t^{t+\Delta t} e^{-2\kappa(t+\Delta t-s)} y_t e^{-\kappa(s-t)} ds \\ &= y_t e^{-2\kappa\Delta t} \int_0^{\Delta t} e^{\kappa u} du \\ &= \frac{y_t}{\kappa} \left(e^{-\kappa\Delta t} - e^{-2\kappa\Delta t}\right). \end{aligned}$$

The second one is :

$$\begin{aligned} I_2 &= \theta \left[\int_0^{\Delta t} e^{-2\kappa(\Delta t-u)} du - \int_0^{\Delta t} e^{-2\kappa(\Delta t-u)} e^{-\kappa u} du\right] \\ &= \theta \left[e^{-2\kappa\Delta t} \int_0^{\Delta t} e^{2\kappa u} du - e^{-2\kappa\Delta t} \int_0^{\Delta t} e^{\kappa u} du\right] \\ &= \theta \left[e^{-2\kappa\Delta t} \cdot \frac{e^{2\kappa\Delta t} - 1}{2\kappa} - e^{-2\kappa\Delta t} \cdot \frac{e^{\kappa\Delta t} - 1}{\kappa}\right] \\ &= \theta_x \left[\frac{1 - e^{-2\kappa\Delta t}}{2\kappa} - \frac{e^{-\kappa\Delta t} - e^{-2\kappa\Delta t}}{\kappa}\right] \\ &= \frac{\theta}{2\kappa} \left(1 - 2e^{-\kappa\Delta t} + e^{-2\kappa\Delta t}\right) \\ &= \frac{\theta}{2\kappa} \left(1 - e^{-\kappa\Delta t}\right)^2. \end{aligned}$$

All results gathered, we finally obtain :

$$\mathbb{V}(y_{t+\Delta t} \mid y_t) = \zeta^2 \left[\frac{y_t}{\kappa} \left(e^{-\kappa\Delta t} - e^{-2\kappa\Delta t}\right) + \frac{\theta}{2\kappa} \left(1 - e^{-\kappa\Delta t}\right)^2\right].$$

□

Bibliography

[1] Donatien Hainaut. A mutually exciting rough jump diffusion for financial modelling. Technical report, Université catholique de Louvain, LIDAM, April 2023.

[2] Peter Tankov. *Financial Modeling with Lévy Processes*. École Polytechnique. Lecture notes.

[3] Donatien Hainaut, Maggie Chen, and Enrico Scalas. The rough hawkes process. Technical report, Université catholique de Louvain, LIDAM, February 2023.

[4] Donatien Hainaut. Switching, self-exciting, fractional and other recent dynamics. In *Continuous Time Processes for Finance*. Springer, 2023.

[5] Donatien Hainaut. Pricing of spread and exchange options in a rough jump-diffusion market. Technical Report 2022/12, LIDAM/ISBA, 2022.

[6] Alex Badescu, Robert J. Elliott, and Tak Kuen Siu. Esscher transforms and consumption-based models. *Insurance: Mathematics and Economics*, 45(3):337–347, 2009.

[7] Gian Marco Oddo. Heston model calibration from option pricing. `https://github.com/GianMarcoOddo/HestonModelCalibrationFFT`, 2024.

[8] Matteo Gambara. Deep calibration: Heston model calibration by machine learning the pricing functional. `https://www.quantlib.org/`, 2024.

[9] Damiano Brigo and Fabio Mercurio. *Interest Rate Models: Theory and Practice*. Springer, 2 edition, 2006.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Faculté des sciences
Place des Sciences, 2 bte L6.06.01, 1348 Louvain-la-Neuve, Belgique | www.uclouvain.be/sc