Rough volatility

Mathieu Rosenbaum

École Polytechnique

Fall 2024

Table of contents

Table of contents

Main classes of volatility models

Prices are often modeled as continuous semi-martingales of the form

$$dP_t = P_t(\mu_t dt + \sigma_t dW_t).$$

The volatility process σ_s is the most important ingredient of the model. Practitioners consider essentially three classes of volatility models :

- Deterministic volatility (Black and Scholes 1973),
- Local volatility (Dupire 1994),
- Stochastic volatility (Hull and White 1987, Heston 1993, Hagan et al. 2002,...).

In term of regularity, in these models, the volatility is either very smooth or with a smoothness similar to that of a Brownian motion.

Fractional Brownian motion (I)

To allow for a wider range of smoothness, we can consider the fractional Brownian motion in volatility modeling.

Definition

The fractional Brownian motion (fBm) with Hurst parameter H is the only process W^H to satisfy :

- Self-similarity : $(W^H_{at}) \stackrel{\mathcal{L}}{=} a^H (W^H_t)$.
- Stationary increments : $(W^H_{t+h} - W^H_t) \stackrel{\mathcal{L}}{=} (W^H_h)$.
- Gaussian process with $\mathbb{E}[W^H_1] = 0$ and $\mathbb{E}[(W^H_1)^2] = 1$.

Fractional Brownian motion (II)

Proposition

For all $\varepsilon > 0$, W^H is $(H - \varepsilon)$-Hölder a.s.

Proposition

The absolute moments of the increments of the fBm satisfy

$$\mathbb{E}[|W^H_{t+h} - W^H_t|^q] = K_q h^{Hq}.$$

Proposition

If $H > 1/2$, the fBm exhibits long memory in the sense that

$$\mathrm{Cov}[W^H_{t+1} - W^H_t, W^H_1] \sim \frac{C}{t^{2-2H}}.$$

Fractional Brownian motion

Mandelbrot-van Ness representation

We have

$$W_t^H = \int_0^t \frac{dW_s}{(t-s)^{\frac{1}{2}-H}} + \int_{-\infty}^0 \left(\frac{1}{(t-s)^{\frac{1}{2}-H}} - \frac{1}{(-s)^{\frac{1}{2}-H}}\right) dW_s.$$

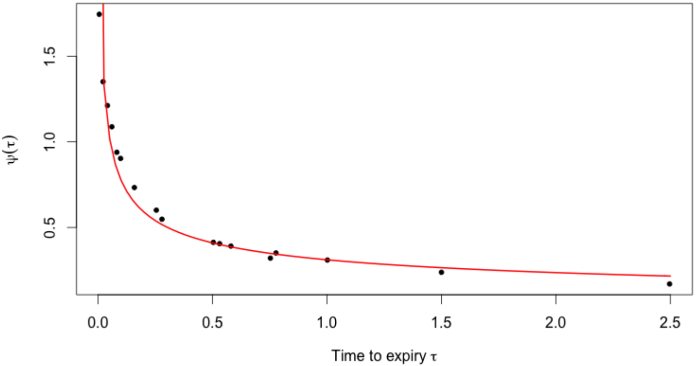
About option data

- Classical stochastic volatility models generate reasonable dynamics for the volatility surface.
- However they do not allow to fit the volatility surface, in particular the term structure of the ATM skew :

$$\psi(\tau) := \left| \frac{\partial}{\partial k} \sigma_{\mathrm{BS}}(k, \tau) \right|_{k=0},$$

where k is the log-moneyness and τ the maturity of the option.

About option data : the volatility skew



The black dots are non-parametric estimates of the S&P ATM volatility skews as of June 20, 2013 ; the red curve is the power-law fit $\psi(\tau) = A\,\tau^{-0.4}$.

About option data : fractional volatility

- The skew is well-approximated by a power-law function of time to expiry τ. In contrast, conventional stochastic volatility models generate a term structure of ATM skew that is constant for small τ.
- Models where the volatility is driven by a fBm generate an ATM volatility skew of the form $\psi(\tau) \sim \tau^{H-1/2}$, at least for small τ.

Table of contents

Intraday volatility estimation

We are interested in the dynamics of the (log)-volatility process. We use two proxies for the spot (squared) volatility of a day.

- A 5 minutes-sampling realized variance estimation taken over the whole trading day (8 hours).
- A one hour integrated variance estimator based on the model with uncertainty zones (Robert and R. 2012).

Note that we are not really considering a "spot" volatility but an "integrated" volatility. This might lead to some slight bias in our measurements (which can be controlled).

From now on, we consider realized variance estimations on the S&P over 3500 days, but the results are fairly "universal".

The log-volatility

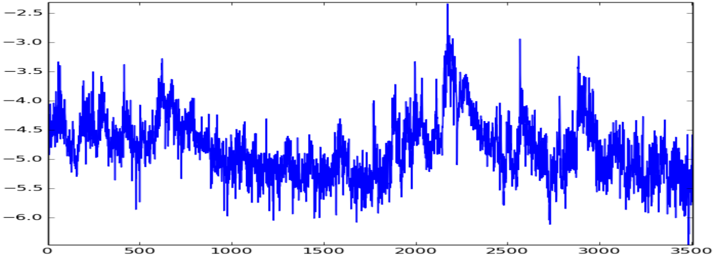


FIGURE – The log volatility $\log(\sigma_t)$ as a function of t, S&P.

Measure of the regularity of the log-volatility

The starting point of this work is to consider the scaling of the moments of the increments of the log-volatility. Thus we study the quantity

$$m(\Delta, q) = \mathbb{E}[|\log(\sigma_{t+\Delta}) - \log(\sigma_t)|^q],$$

or rather its empirical counterpart.

The behavior of $m(\Delta, q)$ when Δ is close to zero is related to the smoothness of the volatility (in the Hölder or even the Besov sense). Essentially, the regularity of the signal measured in l^q norm is s if $m(\Delta, q) \sim c\Delta^{qs}$ as Δ tends to zero.

Scaling of the moments

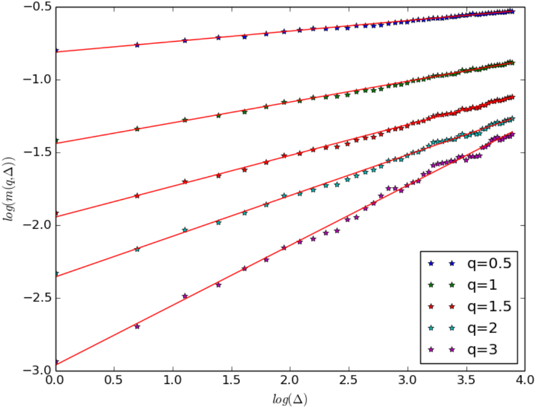


FIGURE – $\log(m(q,\Delta)) = \zeta_q \log(\Delta) + C_q$. The scaling is not only valid as Δ tends to zero, but holds on a wide range of time scales.

Monofractality of the log-volatility

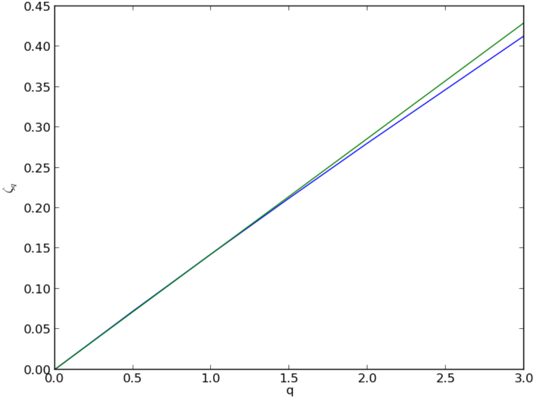


FIGURE – Empirical ζ_q and $q \to Hq$ with $H = 0.14$ (similar to a fBm with Hurst parameter H).

Distribution of the log-volatility increments

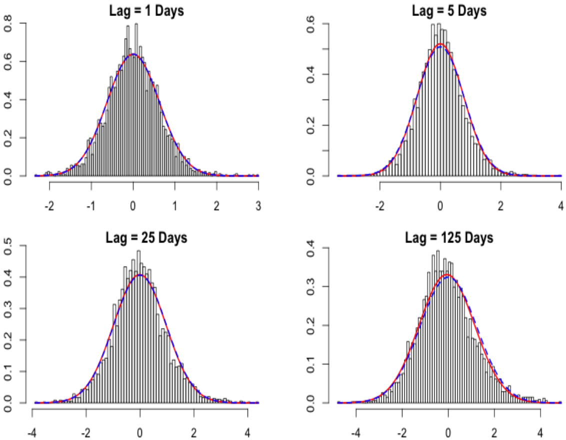


FIGURE – The distribution of the log-volatility increments is close to Gaussian.

A geometric fBm model

The RFSV model

These empirical findings suggest we model the log-volatility as a fractional Brownian motion :

$$\sigma_t = \sigma e^{\nu W_t^H}.$$

Multiscaling in finance

- An important property of volatility time series is their multiscaling behavior, see Mantegna and Stanley 2000 and Bouchaud and Potters 2003.
- This means one observes essentially the same law whatever the time scale.
- In particular, there are periods of high and low market activity at different time scales.
- Very few models reproduce this property, see multifractal models.

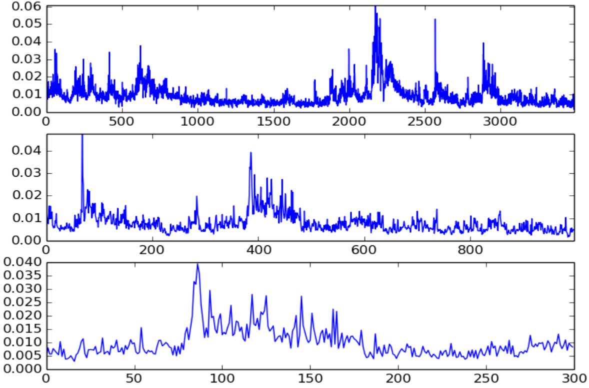


FIGURE – Empirical volatility over 10, 3 and 1 years.

Our model on different time intervals

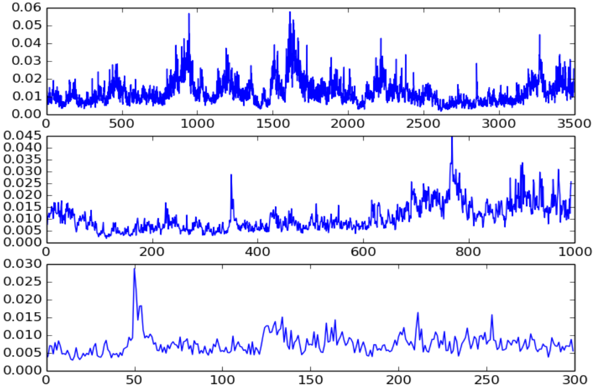


FIGURE – Simulated volatility over 10, 3 and 1 years. We observe the same alternations of periods of high market activity with periods of low market activity.

Apparent multiscaling in our model

- Let $L^{H,\nu}$ be the law on $[0,1]$ of the process $e^{\nu W_t^H}$.
- Then the law of the volatility process on $[0, T]$ renormalized on $[0,1]$: σ_{tT}/σ_0 is $L^{H,\nu T^H}$.
- If one observes the volatility on $T = 10$ years (2500 days) instead of $T = 1$ day, the parameter νT^H defining the law of the volatility is only multiplied by $2500^H \sim 3$.
- Therefore, one observes quite the same properties on a very wide range of time scales.
- The roughness of the volatility process ($H = 0.14$) implies a multiscaling behavior of the volatility.

Table of contents

Prediction of a fractional Brownian motion

There is a nice prediction formula for the fractional Brownian motion.

Proposition (Nuzman and Poor 2000)

For $H < 1/2$

$$\mathbb{E}[W^H_{t+\Delta}|\mathcal{F}_t] = \frac{\cos(H\pi)}{\pi}\Delta^{H+1/2}\int_{-\infty}^t \frac{W^H_s}{(t-s+\Delta)(t-s)^{H+1/2}}ds.$$

Our prediction formula

We apply the previous formula to the prediction of the log-volatility :

$$\mathbb{E}\left[\log \sigma^2_{t+\Delta} | \mathcal{F}_t\right] = \frac{\cos(H\pi)}{\pi} \Delta^{H+1/2} \int_{-\infty}^{t} \frac{\log \sigma^2_s}{(t-s+\Delta)(t-s)^{H+1/2}} ds$$

or more precisely its discrete version :

$$\mathbb{E}\left[\log \sigma^2_{t+\Delta} | \mathcal{F}_t\right] = \frac{\cos(H\pi)}{\pi} \Delta^{H+1/2} \sum_{k=0}^{N} \frac{\log \sigma^2_{t-k}}{(k+\Delta+1/2)(k+1/2)^{H+1/2}}.$$

We compare it to usual predictors using the criterion

$$P = \frac{\sum_{k=1}^{N-\Delta} (\widehat{\log(\sigma^2_{k+\Delta})} - \log(\sigma^2_{k+\Delta}))^2}{\sum_{k=1}^{N-\Delta} (\log(\sigma^2_{k+\Delta}) - \mathbb{E}[\log(\sigma^2_{t+\Delta})])^2}.$$

	AR(5)	AR(10)	HAR(3)	RFSV
SPX2.rv $\Delta = 1$	0.317	0.318	0.314	**0.313**
SPX2.rv $\Delta = 5$	0.459	0.449	0.437	**0.426**
SPX2.rv $\Delta = 20$	0.764	0.694	0.656	**0.606**
FTSE2.rv $\Delta = 1$	0.230	0.229	0.225	**0.223**
FTSE2.rv $\Delta = 5$	0.357	0.344	0.337	**0.320**
FTSE2.rv $\Delta = 20$	0.651	0.571	0.541	**0.472**
N2252.rv $\Delta = 1$	0.357	0.358	0.351	**0.345**
N2252.rv $\Delta = 5$	0.553	0.533	0.513	**0.504**
N2252.rv $\Delta = 20$	0.875	0.795	0.746	**0.714**
GDAXI2.rv $\Delta = 1$	0.237	0.238	0.234	**0.231**
GDAXI2.rv $\Delta = 5$	0.372	0.362	0.350	**0.339**
GDAXI2.rv $\Delta = 20$	0.661	0.590	0.550	**0.498**
FCHI2.rv $\Delta = 1$	0.244	0.244	0.241	**0.238**
FCHI2.rv $\Delta = 5$	0.378	0.373	0.366	**0.350**
FCHI2.rv $\Delta = 20$	0.669	0.613	0.598	**0.522**

Regression window and horizon

After a simple change of variable, the prediction of the log-volatility can be written :

$$\mathbb{E}[\log(\sigma^2_{t+\Delta})|\mathcal{F}_t] \sim \frac{\cos(H\pi)}{\pi} \int_0^1 \frac{\log(\sigma^2_{t-\Delta u})}{(u+1)\, u^{H+1/2}} du.$$

The only time scale that appears in the above regression is the horizon Δ.

As it is known by practitioners :

If trying to predict volatility one week ahead, one should essentially look at the volatility over the last week. If trying to predict the volatility one month ahead, one should essentially look at the volatility over the last month.

Conditional distribution of the fractional Brownian motion and prediction of the variance

Proposition (Nuzman and Poor 2000)

In law,

$$W^H_{t+\Delta}|\mathcal{F}_t = \mathcal{N}(\mathbb{E}[W^H_{t+\Delta}|\mathcal{F}_t], c\Delta^{2H})$$

with

$$c = \frac{\sin(\pi(1/2-H))\Gamma(3/2-H)^2}{\pi(1/2-H)\Gamma(2-2H)}.$$

Therefore, our predictor of the variance writes :

$$\mathbb{E}[\sigma^2_{t+\Delta}|\mathcal{F}_t] = e^{\mathbb{E}\left[\log(\sigma^2_{t+\Delta})|\mathcal{F}_t\right]+2\nu^2 c\Delta^{2H}}.$$

	AR(5)	AR(10)	HAR(3)	RFSV
SPX2.rv $\Delta = 1$	0.520	0.566	0.489	**0.475**
SPX2.rv $\Delta = 5$	0.750	0.745	0.723	**0.672**
SPX2.rv $\Delta = 20$	1.070	1.010	1.036	**0.903**
FTSE2.rv $\Delta = 1$	0.612	0.621	0.582	**0.567**
FTSE2.rv $\Delta = 5$	0.797	0.770	0.756	**0.707**
FTSE2.rv $\Delta = 20$	1.046	0.984	0.935	**0.874**
N2252.rv $\Delta = 1$	0.554	0.579	**0.504**	0.505
N2252.rv $\Delta = 5$	0.857	0.807	0.761	**0.729**
N2252.rv $\Delta = 20$	1.097	1.046	1.011	**0.964**
GDAXI2.rv $\Delta = 1$	0.439	0.448	0.399	**0.386**
GDAXI2.rv $\Delta = 5$	0.675	0.650	0.616	**0.566**
GDAXI2.rv $\Delta = 20$	0.931	0.850	0.816	**0.746**
FCHI2.rv $\Delta = 1$	0.533	0.542	0.470	**0.465**
FCHI2.rv $\Delta = 5$	0.705	0.707	0.691	**0.631**
FCHI2.rv $\Delta = 20$	0.982	0.952	0.912	**0.828**

Table of contents

Understanding rough volatility

Summary of what we have seen and objectives

- We know that : **Volatility is rough** !
- On any asset, using any reasonable volatility proxy/statistical method (realized volatility, realized kernels, uncertainty zones, Garman-Klass, implied volatility, power variations, autocorrelations, Whittle,...), one concludes that volatility is rough.
- It cannot be just coincidence...
- We want to show that typical behaviors of market participants at the high frequency scale naturally lead to rough volatility.
- Our modeling tool : **Hawkes processes**.

Definition

Hawkes process

- A Hawkes process $(N_t)_{t\geq 0}$ is a self-exciting point process, whose intensity at time t, denoted by λ_t, is of the form

$$\lambda_t = \mu + \sum_{0<J_i<t} \phi(t - J_i) = \mu + \int_{(0,t)} \phi(t-s)dN_s,$$

  where μ is a positive real number, ϕ a regression kernel and the J_i are the points of the process before time t.

- These processes have been introduced in 1971 by Hawkes in the purpose of modeling earthquakes and their aftershocks.

Definition

Order flow and volatility

- Thus, it is nowadays classical to model the order flow (number of trades) thanks to Hawkes processes.
- It is known from financial economics theory (see for example Madhavan, Richardson and Roomans (97)) that the order flow is essentially the same thing as the integrated volatility (variance) if the time scale is large enough :

$$N_t \approx \int_0^t \sigma^2(s)ds.$$

Popularity of Hawkes processes in finance

Two main reasons for the popularity of Hawkes processes

- These processes represent a very natural and tractable extension of Poisson processes. In fact, comparing point processes and conventional time series, Poisson processes are often viewed as the counterpart of iid random variables whereas Hawkes processes play the role of autoregressive processes.
- Another explanation for the appeal of Hawkes processes is that it is often easy to give a convincing interpretation to such modeling. To do so, the branching structure of Hawkes processes is quite helpful.

Hawkes processes as a population model

Poisson cluster representation

- Under the assumption $\|\phi\|_1 < 1$, where $\|\phi\|_1$ denotes the L^1 norm of ϕ, Hawkes processes can be represented as a population process where migrants arrive according to a Poisson process with parameter μ.
- Then each migrant gives birth to children according to a non homogeneous Poisson process with intensity function ϕ, these children also giving birth to children according to the same non homogeneous Poisson process, see Hawkes (74).
- Now consider for example the classical case of buy (or sell) market orders. Then migrants can be seen as exogenous orders whereas children are viewed as orders triggered by other orders.

Stability condition

The condition $\|\phi\|_1 < 1$

- The assumption $\|\phi\|_1 < 1$ is crucial in the study of Hawkes processes.
- If one wants to get a stationary intensity with finite first moment, then the condition $\|\phi\|_1 < 1$ is required (similar condition as for the AR(1) process).
- This condition is also necessary in order to obtain classical ergodic properties for the process.
- For these reasons, this condition is often called stability condition in the Hawkes literature.

$\|\phi\|_1$ in practice

Degree of endogeneity of the market

- From a practical point of view, a lot of interest has been recently devoted to the parameter $\|\phi\|_1$.
- For example, Hardiman, Bercot and Bouchaud (13) and Filimonov and Sornette (12,13) use the branching interpretation of Hawkes processes on midquote data in order to measure the so-called degree of endogeneity of the market, defined by $\|\phi\|_1$.

$\|\phi\|_1$ in practice

Degree of endogeneity of the market

- The parameter $\|\phi\|_1$ corresponds to the average number of children of an individual, $\|\phi\|_1^2$ to the average number of grandchildren of an individual,... Therefore, if we call cluster the descendants of a migrant, then the average size of a cluster is given by $\sum_{k\geq 1} \|\phi\|_1^k = \|\phi\|_1/(1-\|\phi\|_1)$.
- Thus, the average proportion of endogenously triggered events is $\|\phi\|_1/(1-\|\phi\|_1)$ divided by $1+\|\phi\|_1/(1-\|\phi\|_1)$, which is equal to $\|\phi\|_1$.

$\|\phi\|_1$ in practice

Unstable Hawkes processes

- This branching ratio can be measured using parametric and non-parametric estimation methods for Hawkes processes, see Ogata (78,83) for likelihood based methods and Reynaud-Bouret and Schbath (10) and Al Dayri *et al.* (11) for functional estimators of the function ϕ.
- In Hardiman, Bercot and Bouchaud (13), very stable estimations of $\|\phi\|_1$ are reported for the E-mini S&P futures between 1998 and 2012, the results being systematically close to one.
- This is also the case for Bund and Dax futures in Al Dayri *et al.* (11) and various other assets in Filimonov and Sornette (12).

Aim of our study

Limiting behavior of Hawkes processes

- Our aim is to study the behavior at large time scales of so-called **nearly unstable Hawkes processes**, which correspond to these estimations of $\|\phi\|_1$, close to 1.
- This will give us insights on the properties of the integrated volatility.
- Furthermore, we want to take into account another stylized fact : The function ϕ has typically a power law tail :

$$\phi(x) \underset{x \to +\infty}{\sim} \frac{K}{x^{1+\alpha}},$$

  with α of order 0.5-0.7.
- This memory effect is likely due to metaorders splitting.

The model

Sequence of Hawkes processes

- We consider a sequence of Hawkes processes $(N_t^T)_{t\geq 0}$ indexed by $T \to \infty$ with
$$\lambda_t^T = \mu^T + \int_0^t \phi^T(t-s)dN_s^T.$$
- For some sequence $a_T < 1$, $a_T \to 1$, $K > 0$ and $\alpha \in (0,1)$:
$$\phi^T(t) = a_T \phi(t),\ \alpha x^\alpha \big(1 - F(x)\big) \underset{x \to +\infty}{\to} K,$$
with $\|\phi\|_1 = 1$ and
$$F(x) = \int_0^x \phi(s)ds.$$

Non-degenerate limit for nearly unstable Hawkes processes

Martingale process

- Let M^T be the martingale process associated to N^T, that is, for $t \geq 0$,

$$M_t^T = N_t^T - \int_0^t \lambda_s^T ds.$$

- We also set ψ^T the function defined on $\mathbb{R}^+$ by

$$\psi^T(t) = \sum_{k=1}^{\infty} (\phi^T)^{*k}(t).$$

- We can show that

$$\lambda_t^T = \mu^T + \int_0^t \psi^T(t-s)\mu^T ds + \int_0^t \psi^T(t-s) dM_s^T.$$

Non-degenerate limit for nearly unstable Hawkes processes

Rescaling

- We rescale our processes so that they are defined on $[0, 1]$. To do that, we consider for $t \in [0, 1]$

$$\lambda_{tT}^T = \mu^T + \int_0^{tT} \psi^T(Tt - s)\mu^T ds + \int_0^{tT} \psi^T(Tt - s)dM_s^T.$$

- For the scaling in space, a natural multiplicative factor is $(1 - a_T)/\mu^T$. Indeed, in the stationary case,

$$\mathbb{E}[\lambda_t^T] = \mu^T/(1 - \|\phi^T\|_1).$$

Thus, the order of magnitude of the intensity is $\mu^T(1 - a_T)^{-1}$. This is why we define

$$C_t^T = \lambda_{tT}^T(1 - a_T)/\mu^T.$$

Decomposition of C_t^T

- Then we easily get :

$$C_t^T = (1-a_T) + \int_0^t T(1-a_T)\psi^T(Ts)ds + \sqrt{\frac{T(1-a_T)}{\mu^T}}\int_0^t \psi^T(T(t-s))\sqrt{C_s^T}dB_s^T,$$

with

$$B_t^T = \frac{1}{\sqrt{T}}\int_0^{tT}\frac{dM_s^T}{\sqrt{\lambda_s^T}}.$$

Non-degenerate limit for nearly unstable Hawkes processes

The function ψ^T

- The asymptotic behavior of C_t^T is closely linked to that of ψ^T.
- Remark that the function defined for $x \geq 0$ by

$$\rho^T(x) = T\frac{\psi^T(Tx)}{\|\psi^T\|_1}$$

is the density of the random variable

$$X^T = \frac{1}{T}\sum_{i=1}^{I^T} X_i,$$

where the (X_i) are iid random variables with density ϕ and I^T is a geometric random variable with parameter $1 - a_T$.

Non-degenerate limit for nearly unstable Hawkes processes

The function ψ^T

- The Laplace transform of the random variable X^T, denoted by $\widehat{\rho}^T$, satisfies :

$$\widehat{\rho}^T(z) = \frac{\hat{\phi}(\frac{z}{T})}{1 - \frac{a_T}{1-a_T}(\hat{\phi}(\frac{z}{T}) - 1)},$$

  where $\hat{\phi}$ denotes the Laplace transform of X_1.

- Due to the assumptions on ϕ, we have

$$\widehat{\phi}(z) = 1 - K\frac{\Gamma(1-\alpha)}{\alpha}z^\alpha + o(z^\alpha).$$

Non-degenerate limit for nearly unstable Hawkes processes

The function ψ^T

- Set $\delta = K\frac{\Gamma(1-\alpha)}{\alpha}$ and $v_T = \delta^{-1}T^{\alpha}(1-a_T)$.
- Using that a_T and $\hat{\phi}(\frac{z}{T})$ both tend to 1 as T goes to infinity, $\widehat{\rho}^T(z)$ is equivalent to
$$\frac{v_T}{v_T + z^{\alpha}}.$$
- The function whose Laplace transform is equal to this last quantity is given by
$$v_T x^{\alpha-1} E_{\alpha,\alpha}(-v_T x^{\alpha}),$$
with $E_{\alpha,\beta}$ the (α,β) Mittag-Leffler function.

Non-degenerate limit for nearly unstable Hawkes processes

Expected limit for C_t^T

- Putting everything together, we can expect (for $\alpha > 1/2$)

$$C_t^T \sim v_T \int_0^t s^{\alpha-1} E_{\alpha,\alpha}(-v_T s^\alpha) ds + \gamma_T v_T \int_0^t (t-s)^{\alpha-1} E_{\alpha,\alpha}(-v_T (t-s)^\alpha) \sqrt{C_s^T} dB_s^T,$$

with

$$\gamma_T = \frac{1}{\sqrt{\mu^T T (1-a_T)}}.$$

- The process B^T can be shown to converge to a Brownian motion B.

Non-degenerate limit for nearly unstable Hawkes processes

Expected limit for C_t^T

- We need that both v_T and γ_T converge to positive constants so we assume :

$$T^{\alpha}(1-a_T) \to \lambda\delta, \;\; T^{1-\alpha}\mu^T \to \mu^*\delta^{-1}.$$

- Passing to the limit, we obtain (for $\alpha > 1/2$)

$$C_t^{\infty} \sim \lambda \int_0^t s^{\alpha-1} E_{\alpha,\alpha}(-\lambda s^{\alpha}) ds$$
$$+ \sqrt{\frac{\lambda}{\mu^*}} \int_0^t (t-s)^{\alpha-1} E_{\alpha,\alpha}(-\lambda(t-s)^{\alpha}) \sqrt{C_s^{\infty}} dB_s.$$

Non-degenerate limit for nearly unstable Hawkes processes

Limit theorem

For $\alpha > 1/2$, the sequence of renormalized Hawkes processes converges to some process which is differentiable on $[0, 1]$. Moreover, the law of its derivative V satisfies

$$V_t = F^{\alpha,\lambda}(t) + \frac{1}{\sqrt{\mu^*\lambda}} \int_0^t f^{\alpha,\lambda}(t-s)\sqrt{V_s}dB_s^1,$$

with B^1 a Brownian motion and

$$f^{\alpha,\lambda}(x) = \lambda x^{\alpha-1} E_{\alpha,\alpha}(-\lambda x^\alpha).$$

Non-degenerate limit for nearly unstable Hawkes processes

Rough Heston model

Using fractional integration, we easily get that the equation for V_t on the preceding slide is equivalent to

$$V_t = \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\lambda(1-V_s)ds + \frac{1}{\Gamma(\alpha)}\sqrt{\frac{\lambda}{\mu^*}}\int_0^t (t-s)^{\alpha-1}\sqrt{V_s}dB_s.$$

Now recall Mandelbrot-van-Ness representation :

$$W_t^H = \int_0^t \frac{dW_s}{(t-s)^{\frac{1}{2}-H}} + \int_{-\infty}^0 \left(\frac{1}{(t-s)^{\frac{1}{2}-H}} - \frac{1}{(-s)^{\frac{1}{2}-H}}\right) dW_s.$$

Therefore we have a rough Heston model with $H = \alpha - 1/2$. Furthermore, for any $\varepsilon > 0$, Y has Hölder regularity $\alpha - 1/2 - \varepsilon$.

Agent based explanation for rough volatility

Microstructural foundations for the RFSV model

- It is clearly established that there is a linear relationship between cumulated order flow and integrated variance.
- Consequently the "derivative" of the order flow corresponds to the spot variance.
- Thus endogeneity of the market together with order splitting lead to a superposition effect which explains (at least partly) the rough nature of the observed volatility.
- Near instability together with a tail index $\alpha \sim 0.6$ correspond to $H \sim 0.1$.

Table of contents

Building the model

Necessary conditions for a good microscopic price model

We want :

- A tick-by-tick model.
- A model reproducing the stylized facts of modern electronic markets in the context of high frequency trading.
- A model helping us to understand the rough dynamic of volatility from the high frequency behavior of market participants.

Building the model

Stylized facts 1-2

- Markets are highly endogenous, meaning that most of the orders have no real economic motivations but are rather sent by algorithms in reaction to other orders, see Bouchaud *et al.*, Filimonov and Sornette.
- Mechanisms preventing statistical arbitrages take place on high frequency markets, meaning that at the high frequency scale, building strategies that are on average profitable is hardly possible.

Building the model

Stylized facts 3-4

- There is some asymmetry in the liquidity on the bid and ask sides of the order book. In particular, a market maker is likely to raise the price by less following a buy order than to lower the price following the same size sell order.
- A large proportion of transactions is due to large orders, called metaorders, which are not executed at once but split in time.

Building the model

Hawkes processes

- Our tick-by-tick price model is based on Hawkes processes in dimension two.
- A two-dimensional Hawkes process is a bivariate point process $(N_t^+, N_t^-)_{t\geq 0}$ taking values in $(\mathbb{R}^+)^2$ and with intensity $(\lambda_t^+, \lambda_t^-)$ of the form :

$$\begin{pmatrix} \lambda_t^+ \\ \lambda_t^- \end{pmatrix} = \begin{pmatrix} \mu^+ \\ \mu^- \end{pmatrix} + \int_0^t \begin{pmatrix} \varphi_1(t-s) & \varphi_3(t-s) \\ \varphi_2(t-s) & \varphi_4(t-s) \end{pmatrix} . \begin{pmatrix} dN_s^+ \\ dN_s^- \end{pmatrix}.$$

Building the model

The microscopic price model

- Our model is simply given by

$$P_t = N_t^+ - N_t^-.$$

- N_t^+ corresponds to the number of upward jumps of the asset in the time interval $[0, t]$ and N_t^- to the number of downward jumps. Hence, the instantaneous probability to get an upward (downward) jump depends on the location in time of the past upward and downward jumps.
- By construction, the price process lives on a discrete grid.
- Statistical properties of this model have been studied in details.

Encoding the stylized facts

The right parametrization of the model

- Recall that

$$\begin{pmatrix} \lambda_t^+ \\ \lambda_t^- \end{pmatrix} = \begin{pmatrix} \mu^+ \\ \mu^- \end{pmatrix} + \int_0^t \begin{pmatrix} \varphi_1(t-s) & \varphi_3(t-s) \\ \varphi_2(t-s) & \varphi_4(t-s) \end{pmatrix} \cdot \begin{pmatrix} dN_s^+ \\ dN_s^- \end{pmatrix}.$$

- High degree of endogeneity of the market$\rightarrow$ L^1 norm of the largest eigenvalue of the kernel matrix close to one (nearly unstable regime).
- No arbitrage$\rightarrow$ $\varphi_1 + \varphi_3 = \varphi_2 + \varphi_4$.
- Liquidity asymmetry$\rightarrow$ $\varphi_3 = \beta\varphi_2$, with $\beta > 1$.
- Metaorders splitting$\rightarrow$ $\varphi_1(x),\ \varphi_2(x) \underset{x\to\infty}{\sim} K/x^{1+\alpha}$, $\alpha \approx 0.6$.

Rough Heston model

Limit theorem

After suitable scaling in time and space, the long term limit of our price model satisfies the following **rough Heston** dynamics :

$$P_t = \int_0^t \sqrt{V_s} dW_s - \frac{1}{2} \int_0^t V_s ds,$$

$$V_t = V_0 + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \lambda(\theta - V_s) ds + \frac{\lambda\nu}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \sqrt{V_s} dB_s,$$

with

$$d\langle W, B\rangle_t = \frac{1-\beta}{\sqrt{2(1+\beta^2)}} dt.$$

The Hurst parameter H satisfies $H = \alpha - 1/2$.

An even deeper result

No-arbitrage implies rough volatility and power law market impact

- We have shown that combining typical behaviours of market participants at the high frequency scale automatically generates rough volatility.
- We can actually prove that only assuming no-statistical arbitrage implies rough volatility.
- The key phenomenon to obtain this result is the **market impact**.

Table of contents

Deriving the characteristic function of the rough Heston model

Strategy

- From our last theorem, we are able to derive the characteristic function of our high frequency Hawkes-based price model.
- We then pass to the limit.

Characteristic function of rough Heston models

We write :

$$I^{1-\alpha} f(x) = \frac{1}{\Gamma(1-\alpha)} \int_0^x \frac{f(t)}{(x-t)^\alpha} dt, \;\; D^\alpha f(x) = \frac{d}{dx} I^{1-\alpha} f(x).$$

Theorem

The characteristic function at time t for the rough Heston model is given by

$$\exp\Big(\int_0^t g(a,s) ds + \frac{V_0}{\theta\lambda} I^{1-\alpha} g(a,t) \Big),$$

with $g(a,)$ the unique solution of the fractional Riccati equation :

$$D^\alpha g(a,s) = \frac{\lambda\theta}{2}(-a^2 - ia) + \lambda(ia\rho\nu - 1) g(a,s) + \frac{\lambda\nu^2}{2\theta} g^2(a,s).$$

Comments on the theorem

The rough Heston formula

- The formula is the very same as the celebrated Heston formula, up to the replacement of a classical time derivative by a fractional derivative.
- This formula allows for fast derivatives pricing and risk management.
- Thanks to this approach, we can derive the infinite dimensional Markovian structure underlying rough Heston models, leading to explicit hedging formulas.

Table of contents

Market impact

Some definitions

- Market impact is the link between the volume of an order (either market order or metaorder) and the price moves during and after the execution of this order.
- We focus here on the impact function of metaorders, which is the expectation of the price move with respect to time during and after the execution of the metaorder.
- We call permanent market impact of a metaorder the limit in time of the impact function (that is the average price move between the start of the metaorder and a long time after its execution).

Market impact in practice

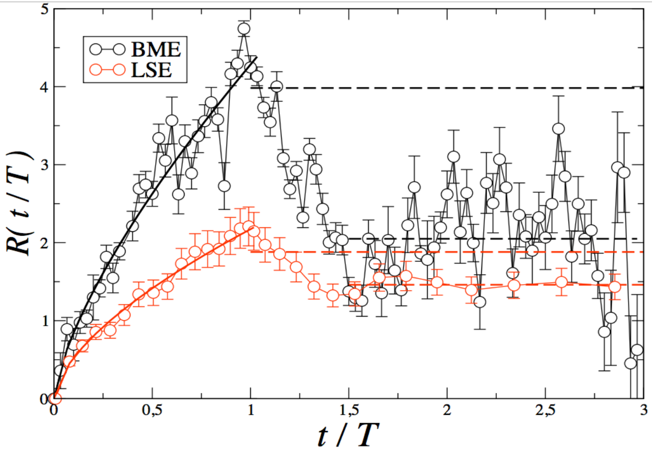


FIGURE – Market impact curves.

Market impact

Linear permanent impact

- Let P_t be the asset price at time t. Consider a metaorder with total volume V.
- $$PMI(V) = \lim_{s \to +\infty} \mathbb{E}[P_s - P_0 | V].$$
- Price manipulation is a roundtrip with negative average cost.
- From Huberman and Stanzl and Gatheral : Only linear permanent market impact can prevent price manipulation : $PMI(V) = kV$.

Dynamics

Assumptions

- All market orders are part of metaorders.
- Let $[0, S]$ be the time during which metaorders are being executed (which can be thought of as the trading day). Let v_i^a (resp. v_i^b) be the volume of the i-th buy (resp. sell) metaorder and N_S^a (resp. N_S^b) be the number of buy (resp. sell) metaorders up to time S. Finally, write V_S^a and V_S^b for cumulated buy and sell order flows up to time S.
- We assume

$$P_S = P_0 + k\Big(\sum_{i=1}^{N_S^a} v_i^a - \sum_{i=1}^{N_S^b} v_i^b\Big) + Z_S = P_0 + k(V_S^a - V_S^b) + Z_S,$$

with Z a martingale term that we neglect.

Dynamics

Martingale assumption

- We furthermore assume that the price P_t is a martingale. We obtain

$$P_t = P_0 + \mathbb{E}\big[k(V_S^a - V_S^b)|F_t\big].$$

- We suppose that $\lim_{S\to+\infty} \mathbb{E}\big[k(V_S^a - V_S^b)|F_t\big]$ is well defined. This means

$$\mathbb{E}\big[(V_{S+h}^a - V_{S+h}^b) - (V_S^a - V_S^b)|F_t\big] \to 0,$$

  that is the order flow imbalance between S and $S+h$ is asymptotically (in S) not predictable at time t.

Dynamics

Price dynamics

- Under the preceding assumptions, we finally get

$$P_t = P_0 + k \lim_{S \to +\infty} \mathbb{E}\big[(V_S^a - V_S^b)|F_t\big].$$

- Martingale price.
- Linear permanent impact, independent of execution mode.
- The price process only depends on the global market order flow and not on the individual executions of metaorders. We thus do not need to assume that the market sees the execution of metaorders as it is usually done.
- Market orders move the price because they change the anticipation that market makers have about the future of the order flow.

Hawkes specification

Hawkes propagator

- We now assume that buy and sell order flows are modeled by independent Hawkes processes N^a and N^b with same parameters μ and ϕ. All orders have same unit volume.
- Later on we will consider an asymptotic setting so that the flows are defined on $[0, T]$ with $T \to +\infty$.
- To be very general, we allow the parameters to depend on T (but do not assume they do). So we write $N^{a,T}$, $N^{b,T}$, μ^T, $\phi^T = a^T \phi$ with $a^T < 1$ and $\int \phi = 1$ (stability condition).
- Note that the average intensity of our processes is essentially $\beta^T = \mu^T (1 - a^T)^{-1}$ (stationary case).

Price dynamic under Hawkes specification

Price equation

- In this case, the general equation above rewrites as the following propagator dynamic

$$P_t = P_0 + \int_0^t \zeta^T(t-s)(dN_s^{a,T} - dN_s^{b,T}),$$

with $\zeta^T(t) = \big(1 + \int_t^{+\infty} \psi^T(u) - \int_0^t \psi^T(u-s)\phi^T(s)dsdu\big)$.

- The propagator kernel compensates the correlation of the order flow implied by the Hawkes dynamics to recover a martingale price. Note that the kernel does not tend to 0 since there is permanent impact.

Adding our own transactions

Labeled order

- In the above framework, $N^{a,T}$ and $N^{b,T}$ are the flows of anonymous market orders.
- Now assume we arrive on the market, executing our own (buy) metaorder. Our flow is a Poisson process n on $[0, T]$ (can be generalized) with intensity $I^T = \gamma\beta^T$, $\gamma < 1$ (proportion γ of the total flow).
- According to the propagator approach, we get

$$P_t = P_0 + \int_0^t \zeta^T(t-s)(dN_s^{a,T} - dN_s^{b,T}) + \int_0^t \zeta^T(t-s)dn_s.$$

Impact function

Explicit market impact

- We get that the impact function of a metaorder executed between 0 and T is for $0 \leq t \leq T$

$$MI(t) := \mathbb{E}[P_t - P_0] = I^T \int_0^t \zeta^T(t-s)ds.$$

- We define

$$\overline{MI}^T(t) = \frac{MI^T_{tT}}{T\beta^T} = \int_0^t \chi^T(t-s)\mathrm{d}s,$$

with

$$\chi^T(s) = \gamma \frac{\zeta^T(Ts)}{1-a^T}.$$

Decomposing the impact

Transient and permanent market impact

- We have

$$\overline{MI}^T(t) = \int_0^t \chi^T(t-s)\mathrm{d}s,$$

$$\chi^T(s) = \gamma\big(1 + (1-a^T)^{-1}\int_{Ts}^{+\infty}\phi\big).$$

- The market impact kernel is the sum of a linear market impact representing the permanent component and of a transient term vanishing after the metaorder completion.
- Existence of transient part is equivalent (asymptotically) to the existence of a limit for $(1-a^T)^{-1}\int_0^t\int_{T(t-s)}^{+\infty}\phi(u)\mathrm{d}u\mathrm{d}s$.

Shape of the market impact

Power-law market impact

Assume the transient part of the market impact exists. Then for $t < 1$,

$$\lim_{T \to +\infty} \overline{MI}^T(t) - \gamma t = \gamma K t^{1-\alpha}$$

for some $K > 0$ and $\alpha \in (0,1)$. Furthermore, we necessarily have $a^T \to 1$ (highly endogenous market) and the tail of the Hawkes kernel is power-law of order $x^{-(1+\alpha)}$.

Note that the celebrated square-root law corresponds to $\alpha = 1/2$.

Market impact decomposition

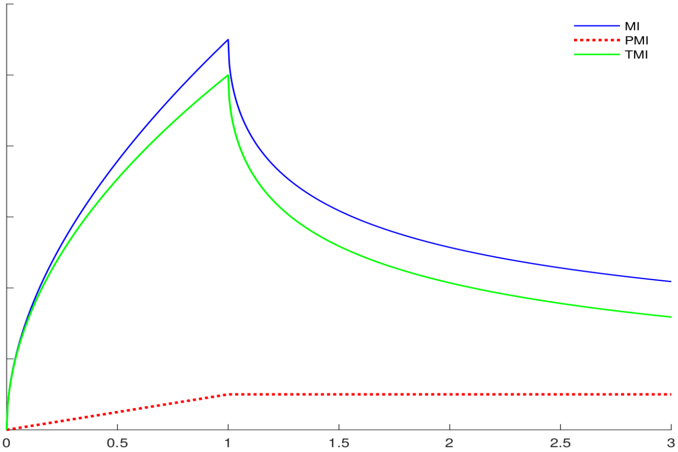


FIGURE – Permanent and temporary market impact

Limiting price process

Emergence of (hyper-)rough processes

Let $\bar{P}_t^T = \frac{1}{T\beta^T} P_t^T$ and assume $\mu^T(1 - a^T)T$ tends to δ. As T goes to infinity, the limit P_t of $\bar{P}_t^T$ satisfies

$$P_t = B_{X_t},$$

$$X_t = \frac{2}{\delta} \int_0^t F^{\alpha,\lambda}(s)\mathrm{d}s + \frac{1}{\delta\sqrt{\lambda}} \int_0^t F^{\alpha,\lambda}(t-s)\mathrm{d}W_{X_s},$$

where B and W are Brownian motions, $\lambda = K\Gamma(1-\alpha)^{-1}$ and $F^{\alpha,\lambda}(t) = \int_0^t f^{\alpha,\lambda}(s)\mathrm{d}s$ with $f^{\alpha,\lambda}$ the density of the Mittag-Leffler distribution. Furthermore, X has Hölder regularity $\min(2\alpha, 1) - \varepsilon$.

The case $\alpha > 1/2$

Rough Heston limit

When $\alpha > \frac{1}{2}$, the rescaled price process variance is almost surely differentiable. Furthermore

$$P_t = \int_0^t \sqrt{Y_s} \mathrm{d}B_s,$$

$$Y_t = \frac{\lambda}{\Gamma(\alpha)}\Big(\int_0^t (t-s)^{\alpha-1}(\frac{2}{\delta} - \lambda Y_s)\mathrm{d}s + \frac{1}{\delta\sqrt{\lambda}} \int_0^t (t-s)^{\alpha-1}\sqrt{Y_s}\mathrm{d}W_s\Big).$$

Therefore we have a rough Heston model with $H = \alpha - 1/2$.

Summary

From no-arbitrage to volatility

- We made two assumptions : Linear permanent impact and martingale price.
- Only modeling assumption : Hawkes dynamics for the order flow (reasonable...).
- This leads to rough volatility.

Table of contents

Further aspects of volatility

Zumbach effect (Zumbach *et al.*) : description

- Feedback of price returns on volatility.
- Price trends induce an increase of volatility.
- In the literature (notably works by J.P. Bouchaud and co-authors), a way to reinterpret the Zumbach effect is to consider that the predictive power of past squared returns on future volatility is stronger than that of past volatility on future squared returns.
- To check this on data, one typically shows that the covariance between past squared price returns and future realized volatility (over a given duration) is larger than that between past realized volatility and future squared price returns.
- We refer to this version of Zumbach effect as *weak Zumbach effect*.

Further aspects of volatility

Weak and strong Zumbach effect

- It is shown in Gatheral *et al.* that the rough Heston model reproduces the weak form of Zumbach effect.
- However, it is not obtained through feedback effect, which is the motivating phenomenon in the original paper by Zumbach. It is only due to the dependence between price and volatility induced by the correlation of the Brownian motions driving their dynamics.
- In particular in the rough Heston model, the conditional law of the volatility depends on the past dynamic of the price only through the past volatility.
- We speak about *strong Zumbach effect* when the conditional law of future volatility depends not only on past volatility trajectory but also on past returns.

A convenient microscopic model encoding Zumbach effect

Quadratic Hawkes processes

- Inspired by Blanc *et al.*, we model high frequency prices using quadratic Hawkes processes.
- Jump sizes of the price P_t are i.i.d taking values -1 and 1 with probability $1/2$ and jump times are those of a point process N_t with intensity

$$\lambda_t = \mu + \int_0^t \phi(t-s)\mathrm{d}N_s + Z_t^2, \text{ with } Z_t = \int_0^t k(t-s)\mathrm{d}P_s.$$

- The component Z_t is a moving average of past returns.
- If the price has been trending in the past, Z_t is large leading to high intensity. On the contrary if it has been oscillating, Z_t is close to zero and there is no feedback from the returns on the volatility. So Z_t is a (strong) Zumbach term.

One particular scaling limit

Quadratic rough Heston model

$$\mathrm{d}S_t = S_t\sqrt{V_t}dW_t,\ \ V_t = a(Z_t - b)^2 + c,$$

where a, b and c some positive constants and Z_t follows

$$Z_t = \int_0^t f^{\alpha,\lambda}(t-s)\theta_0(s)\mathrm{d}s + \int_0^t f^{\alpha,\lambda}(t-s)\sqrt{V_s}\mathrm{d}W_s,$$

with $\alpha \in (1/2, 1)$, $\lambda > 0$ and θ_0 a deterministic function.

- Z_t is *path-dependent* : a weighted average of past returns.
- c : minimal instantaneous variance.
- $b > 0$: asymmetry of the feedback effect.
- a : sensitivity of the volatility feedback.
- **A log-normal rough volatility model with strong Zumbach effect.**

Table of contents

The Volatility Index

Definition of the VIX

- Introduced in 1993 by the CBOE.
- VIX is the square root of the price of a specific basket of options on the S&P 500 Index (SPX) with maturity $\Delta = 30$ days such that

$$\mathrm{VIX}_t = \frac{2}{\Delta}\sqrt{-\mathbb{E}[\log(S_{t+\Delta}/S_t)|\mathcal{F}_t]} \times 100,$$

  with S the SPX index.
- VIX futures and VIX options exist.

The joint calibration problem

VIX options

- More than 500,000 VIX options traded each day.
- Quite wide spreads for VIX options : non-mature market.
- VIX is by definition a derivative of the SPX, any reasonable methodology must necessarily be consistent with the pricing of SPX options.
- Designing a model that jointly calibrates SPX and VIX options prices is known to be extremely challenging.
- This problem is sometimes considered to be *the holy grail of volatility modeling*.
- We simply refer to it as the *joint calibration problem*.

The joint calibration problem

Attempts to solve the joint calibration problem

- Theoretical approch by J. Guyon : the joint calibration problem is interpreted as a model-free constrained martingale transport problem. Perfect calibration of VIX options smile at time T_1 and SPX options smiles at T_1 and $T_2 = T_1 + 30$ days. Hard to be extended to any set of maturities and high computational cost.
- Models with jumps : most of them fail to reproduce VIX smiles for maturities shorter than one month.
- Continuous models : Unsuccessful so far. Interpretation : the very large negative skew of short-term SPX options, which in continuous models implies a very large volatility of volatility, seems inconsistent with the comparatively low levels of VIX implied volatilities

The VIX conjecture

The joint calibration problem and continuous models

- "So far all the attempts at solving the joint SPX/VIX smile calibration problem [using a continuous time model] only produced imperfect, approximate fits".
- "Joint calibration seems out of the reach of continuous-time models with continuous SPX paths".
- Investigating Guyon's work one can realise the following : a necessary condition for a continuous model to fit simultaneously SPX and VIX smiles is the inversion of convex ordering between volatility and the local volatility implied by option prices.
- The intuition behind this condition could be reinterpreted as some kind of strong Zumbach effect.
- Natural for us to investigate the ability of super-Heston rough volatility models to solve the joint calibration problem.

Parameters calibration with Deep Learning

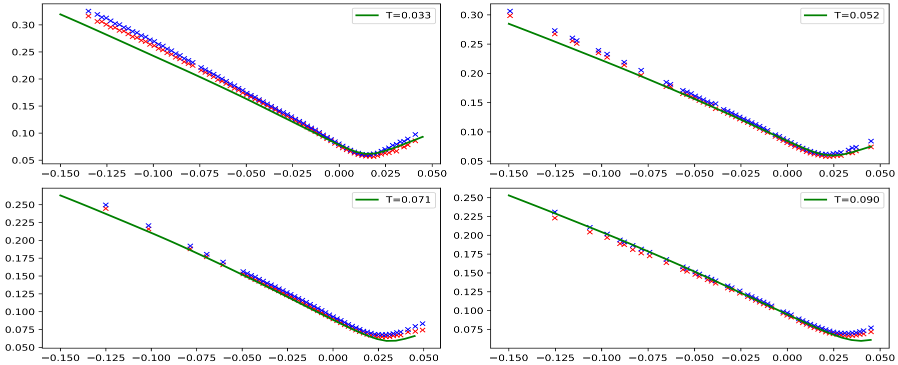


FIGURE – Implied volatility on SPX options for 19 May 2017. Blue and red points are bid and ask of market implied volatilities. Model implied volatility smiles from the model are in green. Strikes are in log-moneyness, maturity in year.

Parameters calibration with Deep Learning

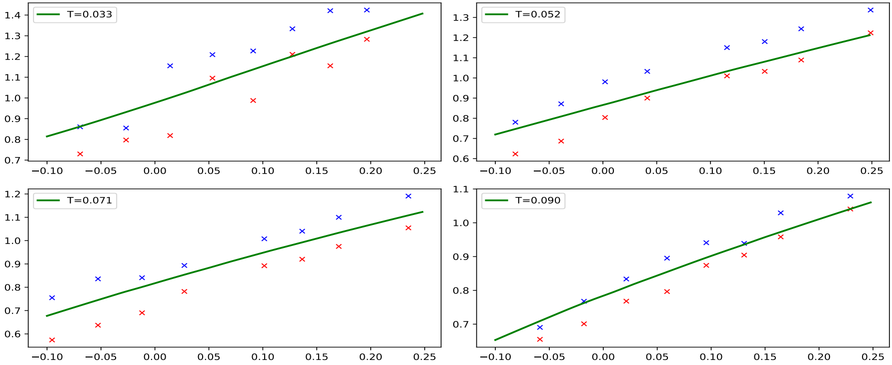


FIGURE – Implied volatility on VIX options for 19 May 2017. Blue and red points are bid and ask of market implied volatilities. Model implied volatility smiles from the model are in green. Strikes are in log-moneyness, maturity in year.

Take home message for the joint calibration problem

Thanks to the quadratic rough Heston model

- 6 parameters.
- VIX smiles in the bid-ask spread.
- Global shape of the implied volatility surface of the SPX very well reproduced
- Very accurate SPX skews of orders -1.5 (shortest maturites), -1 (longer maturities), as for market data.

Table of contents

Forecasting devices for next day realized volatility

Parametric methods

- AR(p) :

$$\hat{\sigma}_t = \alpha_0 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}\,,$$

- HAR :

$$\hat{\sigma}_t = \alpha_0 + \beta_1 \sigma_{t-1} + \beta_2 \frac{1}{5} \sum_{j=1}^{5} \sigma_{t-j} + \beta_3 \frac{1}{22} \sum_{j=1}^{22} \sigma_{t-j}\,,$$

- RFSV ($d \log \sigma_t = \nu dW_t^H$) :

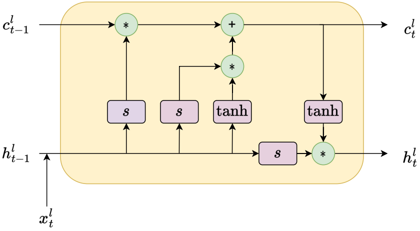
$$\widehat{\log \sigma_t} = \frac{\cos(H\pi)}{\pi} \int_{-\infty}^{t-1} \frac{\log \sigma_s}{(t-s+1)(t-s)^{H+1/2}} ds\,,$$

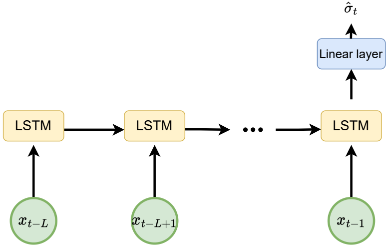
$$\hat{\sigma}_t = c \exp(\widehat{\log \sigma_t}).$$

Forecasting devices for daily realized volatility

A universal non-parametric method

- LSTM recurrent neural network, with similar weights for each asset, trained on a pooled dataset.
- Inputs are $x_t = (\sigma_t^2)$ or $x_t = (\sigma_t^2, r_t)$, where r_t is the daily return at time t, with variable length for history.





Dataset

Description

- 5-minutes intraday prices of Russell 1000 and STOXX Europe 600 constituents, for years between 2010 and 2020.
- 862 names from the US market and 503 names from the European market.
- Scaling for each stock :

$$\sigma_t = \frac{\sigma_t}{\sqrt{\langle \sigma_t^2 \rangle}}, \qquad r_t = \frac{r_t - \langle r_t \rangle}{\sqrt{\langle (r_t - \langle r_t \rangle)^2 \rangle}}.$$

- We focus mostly on the US market. The data of the European market is used for an out-of-sample double-check.
- We use the pooled dataset of 862 stocks over years 2010 - 2015 to train the LSTM network. The period 2016 - 2020 is used for out-of-sample evaluation.

Evaluation metric

Relative mean square error

- $$\mathrm{MSE}(\sigma, \hat{\sigma}) = \frac{1}{T}\sum_{t=1}^{T}(\hat{\sigma}_t - \sigma_t)^2,$$

  where T is the number of trading days of the out-of-sample period.
- We focus on each model's relative performance compared to that of the HAR model so that we compute instead $(\mathrm{MSE}_{\mathrm{m}}/\mathrm{MSE}_{\mathrm{HAR}})$, for $m \in \{\mathrm{AR}(22), \mathrm{RFSV}, \mathrm{LSTM}\}$.

Capturing universality with LSTM

Hurst parameter

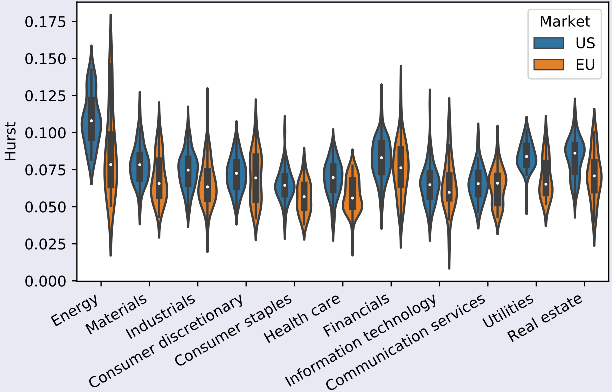


FIGURE – Empirical distribution of the estimated Hurst parameters inside each sector.

Capturing universality with LSTM

Parametric vs non-parametric

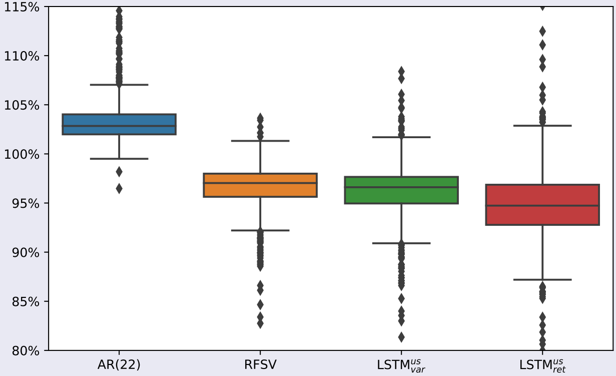


FIGURE – Boxplot showing each model's out-of-sample MSE relative to the HAR model for the stocks of the US market.

Capturing universality with LSTM

Parametric vs non-parametric

- AR(22) underperforms the HAR model (overfitting).
- RFSV outperforms the HAR model. It is remarkable as it involves essentially no parameters ($H = 0.055$, $c = 1.03$).
- LSTM^{us}_{var} and LSTM^{us}_{ret} outperform the other parametric models, especially when we incorporate past returns data. This indicates that the potential universal volatility formation mechanism across assets, relating past volatilities and returns to current volatilities, allows us to calibrate a universal model based on all assets, where the risk of overfitting is reduced due to enriched realized scenarios.
- We check for potential sector/stock (transfer learning)/market specific or time dependent component in the volatility formation process but consistently found that our universal network could not be significantly improved.

Uncovering the universal volatility formation process

A quadratic rough Heston inspired forecast

- Following the idea on Zumbach effect in the QRH model, we propose the following forecasting device :

$$\hat{\sigma}_t^2 = a(Z_{t-1} - b)^2 + c$$

with $Z_t = \int_{-\infty}^{t} \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} \sigma_s \mathrm{d}W_s$.

- We finally consider the following forecast

$$(1-\lambda)\hat{\sigma}^{RFSV} + \lambda\hat{\sigma}^{QRH},$$

with $H = 0.055$, $c = 1.03$, $a = 0.043$, $b = 0.74$, $c = 0.55$.

Uncovering the universal volatility formation process

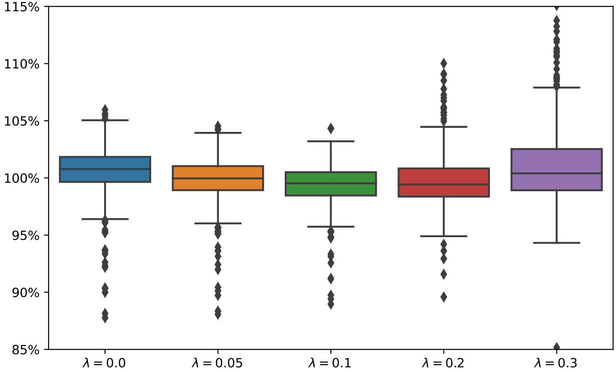


FIGURE – Out-of-sample performance of the forecast $(1-\lambda)\hat{\sigma}^{RFSV} + \lambda\hat{\sigma}^{QRH}$ relative to LSTM^{eu}_{ret} in the EU market.

Conclusion

Universality of the volatility formation process

- The universal LSTM network, trained on a pooled dataset of hundreds of stocks, outperforms consistently the asset-specific parametric models based on past volatilities.
- Similar superior performances hold on assets that are not part of the training set, even on those of a different market. Fine-tuning the universal model with the data of each stock does not help improve the performance.
- These observations suggest the existence of a universal volatility formation mechanism from a nonparametric perspective.
- A simple combination of the RFSV and QRH forecasts with fixed parameters perform similarly to our LSTM network.
- From a parametric perspective, this shows that the main features of this universal volatility formation process can be well described by the rough volatility paradigm boosted with Zumbach effect.