Volatility is rough: microstructural foundations

Jim Gatheral[1], Thibault Jaisson[2] and Mathieu Rosenbaum[3]

[1]City University of New York, [2]École Polytechnique Paris, [3]University Paris 6

27 August 2015

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.

Long memory in volatility

Definition

A stationary process is said to be long memory if its autocovariance function decays slowly, more precisely :

$$\sum_{t=1}^{+\infty} \mathrm{Cov}[\sigma_{t+x}, \sigma_x] = +\infty.$$

Power law long memory for the volatility :

$$\mathrm{Cov}[\sigma_{t+x}, \sigma_x] \sim C/t^{\gamma},$$

with $\gamma < 1$, is considered a stylized fact and has been notably reported in Ding and Granger 1993 (using extra day data) and Andersen *et al.*, 2001 and 2003 (using intra day data).

Fractional Brownian motion (I)

To take into account the long memory property and to allow for a wider range of smoothness, some authors have introduced 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

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

Long memory volatility models

Some models have been built using fractional Brownian motion with Hurst parameter $H > 1/2$ to reproduce the long memory property of the volatility :

- Comte and Renault 1998 (FSV model) :

$$d\log(\sigma_t) = \nu dW_t^H + \alpha(m - \log(\sigma_t))dt.$$

- Comte, Coutin and Renault 2012, where they define a kind of fractional CIR process.

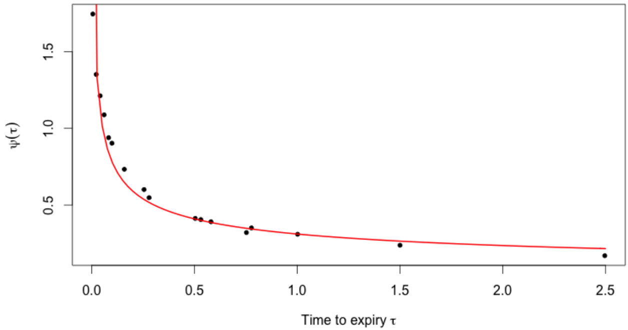
About option data

- 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.

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

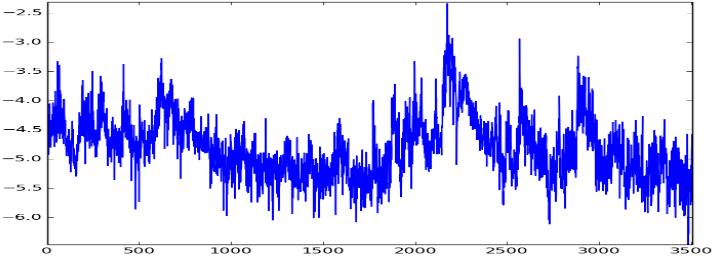


FIGURE : The log volatility $\log(\sigma_t)$ as a function of t, S&P.

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

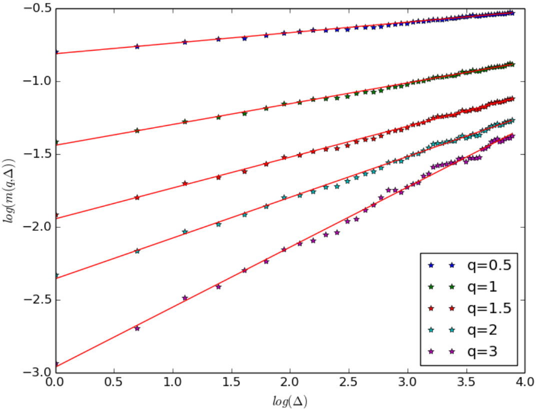


FIGURE : $\log(m(q,\Delta)) = \zeta_q \log(\Delta) + C_q$. The scaling is not only valid as Δ tends to zero, but holds on a wide range of time scales.

Monofractality of the log-volatility

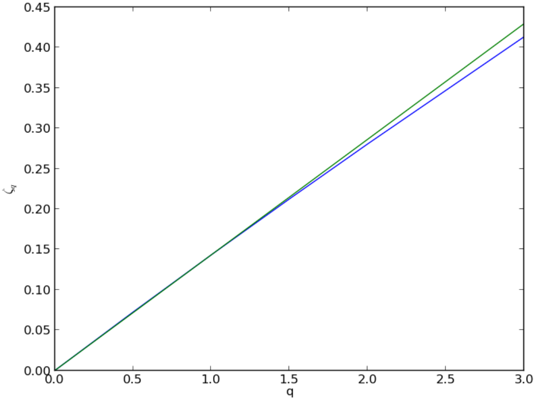


FIGURE : Empirical ζ_q and $q \to Hq$ with $H = 0.14$ (similar to a fBm with Hurst parameter H).

Distribution of the log-volatility increments

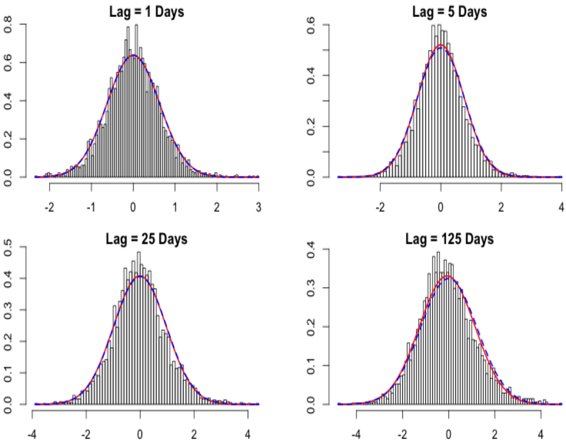


FIGURE : The distribution of the log-volatility increments is close to Gaussian.

A geometric fBm model ?

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

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

However, this model is not stationary. In particular, the empirical autocovariance function of the (log-)volatility (which will be of interest) does not make much sense.

A geometric fOU model

We make it formally stationary by considering a fractional Ornstein-Uhlenbeck model for the log-volatility denoted by X_t

$$dX_t = \nu dW_t^H + \alpha(m - X_t)dt.$$

This process satisfies

$$X_t = \nu \int_{-\infty}^{t} e^{-\alpha(t-s)} dW_t^H + m.$$

We take the reversion time scale $1/\alpha$ very large compared to the observation time scale.

This model is a particular case of the FSV model. However, in strong contrast to FSV, we take H small and $1/\alpha$ large. Thus we call our model Rough FSV (RFSV).

Properties of the RFSV model

Statistical analysis of the RFSV model

- Reproduces very well the correlation structure of the volatility, with explicit formulas.
- No power law long memory property.
- When applied to the RFSV model, statistical tests for long memory behave the same way as for real data and deduce, probably wrongly, the presence of long memory in the volatility.
- Multiscaling behaviour.
- Explicit prediction formulas for the future volatility, depending only on the parameter H, outperforming classical predictors. To forecast the volatility at time $t + \Delta$, one needs to consider the data in the past until $t - \Delta$.

Table of contents

Definition

Hawkes process

- It is nowadays classical to model the order flow (number of trades) thanks to Hawkes processes. The order flow is essentially the same thing as the integrated volatility if the time scale is large enough.
- 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.

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.
- 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 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 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.

Table of contents

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 $\alpha \in (0,1)$:

$$\phi^T(t) = a_T\phi(t),\ \alpha x^\alpha (1 - F(x)) \underset{x \to +\infty}{\to} K,$$

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

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

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 $t \in [0,1]$

$$\lambda^T_{tT} = \mu^T + \int_0^{tT} \psi^T(Tt-s)\mu^T ds + \int_0^{tT} \psi^T(Tt-s)dM^T_s.$$

- 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

$$C^T_t = \lambda^T_{tT}(1-a_T)/\mu^T.$$

Non degenerate limit for nearly unstable Hawkes processes

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 Y satisfies

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

with B^1 a Brownian motion and

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

Therefore $H = \alpha - 1/2$. Furthermore, for any $\varepsilon > 0$, Y has Hölder regularity $\alpha - 1/2 - \varepsilon$.

Agent based explanation for RFSV

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 $H \sim 0.1$.