

Quadratic Hawkes processes for financial prices

Pierre Blanc, Jonathan Donier, Jean-Philippe Bouchaud

September 28, 2015

Abstract

We introduce and establish the main properties of QHawkes (“Quadratic” Hawkes) models. QHawkes models generalize the Hawkes price models introduced in E. Bacry et al. (2014), by allowing all feedback effects in the jump intensity that are linear and quadratic in past returns. A non-parametric fit on NYSE stock data shows that the off-diagonal component of the quadratic kernel indeed has a structure that standard Hawkes models fail to reproduce. Our model exhibits two main properties, that we believe are crucial in the modelling and the understanding of the volatility process: first, the model is time-reversal asymmetric, similar to financial markets whose time evolution has a preferred direction. Second, it generates a multiplicative, fat-tailed volatility process, that we characterize in detail in the case of exponentially decaying kernels, and which is linked to Pearson diffusions in the continuous limit. Several other interesting properties of QHawkes processes are discussed, in particular the fact that they can generate long memory without necessarily be at the critical point. Finally, we provide numerical simulations of our calibrated QHawkes model, which is indeed seen to reproduce, with only a small amount of quadratic non-linearity, the correct magnitude of fat-tails and time reversal asymmetry seen in empirical time series.

1 Introduction: fBMs, GARCHs and Hawkes

The hunt for a “perfect” statistical model of financial markets is still going on. Since the primitive Brownian motion model first proposed by Bachelier, droves of more and more sophisticated mathematical frameworks have been devised to describe the salient stylized facts of financial time series, namely: fat (power-law) tails of the return distribution, volatility (or trading activity) clustering with slow decay of correlations, negative return-volatility correlations (the so-called leverage effect), etc. The two most successful family of models to date are: a) GARCH-like models with slowly decaying memory kernels (e.g. FIGARCH models) and b) stochastic volatility models where the log-volatility follows a fractional Brownian motion with a small Hurst exponent (e.g. the Multifractal Random Walk [4] or, more recently, the “rough volatility” model of Gatheral, Jaisson and Rosenbaum [19]). Although these models are remarkably parsimonious and convincingly capture many features of financial time series, they are still unsatisfactory on several counts. First, the returns in these models are conditionally Gaussian and therefore never “fat enough”, even with a fluctuating volatility. Non-Gaussian residuals (or jumps) must be introduced by hand to match empirical probability distributions. Second, these models are not derived from deeper assumptions on the underlying mechanisms giving rise to fat-tails and volatility clustering. The theorist dream would be to start from, e.g. agents with simple trading rules or behavioral biases, and find that upon aggregation, their collective actions lead to a certain class of stochastic model. Many attempts in this direction have been documented, in particular agent-based models of markets, stylized population dynamics models, or “Minority Games” – for reviews see e.g. [12, 16]. Still, it is fair to say that none of these proposals has yet been widely accepted as a convincing “micro-based” explanation of the stylized facts recalled above.

A further, less discussed, but in our eyes highly relevant stylized fact is related to the time-reversal (a)symmetry (TRS/TRA) of financial time series. As initially emphasized by Zumbach [37] (following earlier ideas [29, 30, 31]), financial time series are *not* statistically symmetrical when past and future are interchanged; see [36]. There are (at least) two distinct effects that break this symmetry: one is the leverage effect alluded to above: *past* returns r affect (negatively) *future* volatilities σ , but not the other way round. This is an effect that breaks both TRS and

the up-down symmetry $r \rightarrow -r$. There is another effect though, that *is* invariant under $r \rightarrow -r$, namely: past large scale realized volatilities are more correlated with future small scale realized volatilities than vice-versa [37]. A more transparent way to explain this rather abstract notion is as follows: take r to be daily returns (say) and σ to be an estimator of volatility based on (say) five minute returns. Then consider, as in [14], the average $\langle r_t^2 \sigma_{t+\tau}^2 \rangle_t$ with $\tau > 0$, which measures the correlation between past daily volatilities with future five minutes volatilities. The Zumbach effect, rephrased and empirically confirmed in [14], is that $\langle r_t^2 \sigma_{t+\tau}^2 \rangle_t > \langle r_{t+\tau}^2 \sigma_t^2 \rangle_t$. It is clear that this criterion is invariant under $r \rightarrow -r$, and is thus unrelated to the leverage effect. Where does such an asymmetry come from and what are the models consistent with TRA?

Interestingly, all continuous time stochastic volatility models, from the famous CIR-Heston model [15, 23] to the Multifractal Random Walk model alluded to above, obey TRS by construction, and therefore *cannot* account for the empirical TRA of financial time series. GARCH-like models, on the other hand, do lead to strong TRA [37], in fact stronger than seen in data [14]. This is expected; GARCH models do encode a feedback from past to future: large past realized returns lead to large future volatilities. This self-exciting mechanism is actually very similar to the one underlying “Hawkes processes” (invented in the context of earthquake statistics), that have attracted a considerable amount of interest recently (for recent reviews, see [5, 28]). In a financial context, Hawkes processes can be seen as a mid-way between purely stochastic models and agent based models. One postulates that the activity rate at time t , λ_t , depends on the history of the point process itself $N_{s < t}$ via the auto-regressive relation

$$\lambda_t = \lambda_\infty + \int_{-\infty}^t \phi(t-s) dN_s, \quad (1)$$

where λ_∞ is a baseline intensity and ϕ is a non-negative, measurable function such that $\|\phi\|_1 = \int_0^\infty ds \phi(s) \leq 1$. Hawkes processes are called “self-exciting”, because every jump $dN_s \neq 0$ increases the probability of future events for $t > s$ via the kernel ϕ ; this in turn leads to activity clustering with an enticing causal interpretation: each event is a new signal for the rest of the market, triggering more activity. When calibrated on financial data, two remarkable features of the Hawkes process are found [10, 20, 21, 5]: its kernel $\phi(s)$ shows long-range (power-law) decay $s^{-1-\epsilon}$, and its L1 norm $\|\phi\|_1$ is very close to unity, meaning that the process is on the verge of becoming unstable (see however [17]). This is quite interesting since this is precisely the regime where the corresponding continuous time limit for the squared volatility (identified here with the activity) is a fractional CIR-Heston process [26], with local Hurst exponent $H = \epsilon - 1/2$. This seems to close the loop: since ϵ is empirically found to be close to $1/2$ [20], one has at hand a “micro-model” (the Hawkes process) that generates on coarse-grained scales a rough volatility process, which generalizes the CIR-Heston model to account for a slow, multi-timescale decay of volatility. Unfortunately, the situation is not as rosy yet: first, the fractional CIR-Heston process has tails that are much too thin (exponentially decaying) to account for the empirical distribution of volatility. Jaisson and Rosenbaum [26] therefore suggest to interpret the Hawkes process as a model for the *log*-volatility, but this is not natural. Second, following our discussion on TRS above, the fractional CIR-Heston process (as on fact the normal CIR-Heston one) is strictly TRS, and therefore fails to capture the observed TRA of financial time series!

The long story above sets the stage for our contribution, which is in an attempt to address the above deficiencies of the Hawkes formalism – when applied to financial time series – and take a step closer to the “perfect” model alluded to in our opening sentence. We propose a generalized version of the Hawkes process (called the QHawkes below) that includes features of the QARCH model introduced by Sentana [33] and revisited in depth in [14]. The idea is that the self-exciting mechanism is not only from market activity onto market activity but also from actual price changes onto market activity. To make our motivation clear, consider a sequence of price moves, all with the same amplitude $|r| := \psi$. One expects that local trends, i.e. a succession of price moves in the same direction (up or down), triggers more volatility than a succession of compensated price moves, even though the high-frequency activity – the number of price moves – is exactly the same. The extra term we need to include in our generalized Hawkes process, beyond being motivated by empirical data, will encode mathematically this effect. We will see how this modification not only generates the needed fat tails in the return distribution (coming from the fact that the log-activity will indeed appear as a natural variable), but also accounts quantitatively for the TRA of

returns, at least on the intraday time scales on which we calibrate the model. We will see that in a particular case, the continuous-time limit of our model boils down to a simple, tractable two-dimensional Pearson diffusion, which can then be used as a low-frequency proxy for the volatility process.

The outline of the paper is as follows: we first introduce our general model in Section 2, and highlight some of its core properties. Section 2.2 introduces a particular sub-family of factorized QHawkes processes, which we call ZHawkes after Zumbach, since it captures the Zumbach mechanism for generating TRA discussed above. Section 3 works out the parallel with QARCH models, which we calibrate on intra-day US stock data using the methodology similar to [14], showing a non-zero off-diagonal structure. We show in Section 4 that in the case of exponential kernels the process is Markovian, and we write the corresponding stochastic differential equation as well as its continuous counterpart. Finally, we show in Section 5, using numerical simulations, that the order of magnitude of the TRA generated by our ZHawkes process matches data quite well, and produce a volatility process with the right amount of fat-tails. Section 6 then concludes.

2 The QHawkes model

2.1 A general model

Similarly to Hawkes processes (1), the QHawkes (Quadratic Hawkes) process $(P_t)_{t \geq 0}$ is a self-exciting point process, whose intensity λ_t is dependent on the past realization of the process itself. As the name suggests, we model the intensity of price changes as the most general self-exciting point process that is Quadratic in $dP_{s < t}$:

$$\lambda_t = \lambda_\infty + \frac{1}{\psi} \int_{-\infty}^t L(t-s) dP_s + \frac{1}{\psi^2} \int_{-\infty}^t \int_{-\infty}^t K(t-s, t-u) dP_s dP_u, \quad (2)$$

where P is the high-frequency price, which is a pure jump process with signed increments. More precisely, whenever an event occurs between t and $t + dt$, with probability $\lambda_t dt$, the price jumps by an amount ξ , where ξ is a random variable of zero mean and variance ψ^2 . A simple case that we will consider below is $\xi = \pm\psi$ with probability $\frac{1}{2}, \frac{1}{2}$, where ψ can be seen as the tick size. In the above equation, $L : \mathbb{R}^+ \rightarrow \mathbb{R}$ is a “leverage” kernel, coupling linearly price changes to market activity and $K : \mathbb{R}^+ \times \mathbb{R}^+ \rightarrow \mathbb{R}$ is a quadratic feedback kernel. λ_∞ is again the baseline intensity of the process (in the absence of any feedback). Note that the above equation can be seen as a systematic expansion of the intensity of price changes in powers of past price changes, truncated to second order. One could of course generalize the model further by adding, e. g. a third order term as $\int_{-\infty}^t \int_{-\infty}^t \int_{-\infty}^t K_3(t-s, t-u, t-v) dP_s dP_u dP_v$, etc., but we will not consider this path further in the following.

Although it is necessary to account for the leverage effect on daily time scales, we will find later that on intra-day scales, the kernel L is not significant, so for many applications one can focus on the quadratic kernel only and write

$$\lambda_t = \lambda_\infty + \frac{1}{\psi^2} \int_{-\infty}^t \int_{-\infty}^t K(t-s, t-u) dP_s dP_u. \quad (3)$$

It is easy to see that model (3) is a generalisation of the simple Hawkes process for prices introduced in [3]: when choosing unit price jumps $dP_t = \pm\psi dN_t$ where ψ can be seen as the tick and discarding any off-diagonal quadratic effects (so that $K(t, s) = \phi(t)\delta_{t-s}$), we recover a Hawkes process of kernel $\phi(s) = K(s, s)$.¹

It is well known that the linear Hawkes process (1) can be seen as a branching process, where each “immigrant” event from the exogenous intensity λ_∞ gives birth to a number of “children” events distributed as a Poisson law of parameter $n_H = \|\phi\|_1$, where $\|\phi\|_1$ is the L^1 norm of the kernel ϕ . Each of these children in turn gives birth to a second generation of children with

¹In fact, if the kernels K, K_3 , etc. to arbitrary order are all diagonal, the model boils down to a Hawkes process with leverage, i.e. $\lambda_t = \lambda_\infty + \int_{-\infty}^t \phi(t-s) dN_s + \psi^{-1} \int_{-\infty}^t L(t-s) dP_s$, with adequately redefined kernels ϕ and L , such that $\phi(s) - |L(s)| + \lambda_\infty \geq 0$ to ensure positivity of the intensity.

the same probability law and so on. When $n_H < 1$, each immigrant gives birth on average to $n_H/(1 - n_H) < \infty$ descendants. n_H can thus be seen as a measure of endogeneity of the process, since it corresponds to the fraction of events that are triggered internally, reaching zero in the case of simple Poisson process and one in the special case of Hawkes process without ancestors [10]. The intuition behind the QHawkes in terms of a branching process is very similar, except that now the rate of events also depends on the interaction between the pairs of events. We will consider a positive feedback $K(s, t) \geq 0$ such that two mother events with the same sign (i.e. two prices moves in the same direction) increase the probability of a new event to be triggered in the future (i.e. increase future volatility), whereas compensated events have inhibiting effects, in line with (and directly motivated by) the empirical observations of [14], as emphasized in the introduction.

2.2 A special case: the ZHawkes model

Motivated by the discussion in the introduction and by the empirical (intraday) results presented below, we will specialize the QHawkes model to the case where there is no leverage ($L \equiv 0$) and the quadratic feedback kernel K is of the form

$$K(t, s) = \phi(t)\delta_{t-s} + k(t)k(s),$$

i.e. the sum of a diagonal Hawkes component and of a factorisable, rank one kernel. We assume that $\phi, k : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ are two positive, measurable functions that satisfy

$$\|\phi\|_1 \equiv \int_0^{+\infty} \phi(u) \, du < +\infty \quad , \quad \|k^2\|_1 \equiv \int_0^{+\infty} k(u)^2 \, du < +\infty.$$

Equation (2) becomes in that case

$$\lambda_t = \lambda_\infty + H_t + Z_t^2, \tag{4}$$

where

- The ‘‘Hawkes term’’ is given by

$$H_t := \int_{-\infty}^t \phi(t-s) \, dN_s; \quad N_t - N_{t-} := \frac{1}{\psi^2} (P_t - P_{t-})^2$$

- The ‘‘Zumbach term’’ given by Z_t^2 where

$$Z_t = \frac{1}{\psi} \int_{-\infty}^t k(t-s) \, dP_s.$$

We call Z_t the Zumbach term since it is directly inspired by the series of empirical observations made by G. Zumbach on the volatility process ([37],[36]). Indeed, Z_t is simply a moving average of the past returns (with positive un-normalized weights $k(\tau)$). Therefore, this term will indeed be such that a sequence of returns in the same direction triggers more future volatility than compensated returns, as empirically observed [36].²

Besides its empirical motivations, the factorization property of the ZHawkes kernel significantly reduces the risk of over-fitting, since we will be left with two one-dimensional kernels instead of the two-dimensional kernel in Eq. 2. As we see below, this simplified setup still captures the main phenomenology of price volatility, with in particular time-reversal asymmetry and fat tails, *even for short-ranged kernels*.

2.3 Mathematical framework

Let us start by specifying the mathematical definition of the objects present in Equation (2):

²Although Zumbach describes this effect at the daily time scale, whereas we will here study intra-day time scales.

- $(P_t)_{t \in \mathbb{R}}$ is a pure jump process of stochastic intensity $(\lambda_t)_{t \in \mathbb{R}}$, with unpredictable i.i.d. jump sizes ξ of common law p on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We assume that $\int_{\mathbb{R}} \xi p(d\xi) = 0$ and $\int_{\mathbb{R}} \xi^2 p(d\xi) = \psi^2 < +\infty$, i.e. that jumps are centered and have a finite variance.
- $\mathcal{F}_t = \sigma(P_s, s \leq t)$ is the natural filtration of P .
- $m(dt, d\xi)$ is the Punctual Poisson Measure associated to P , such that for all $t \in \mathbb{R}$ and $A \in \mathcal{B}(\mathbb{R})$,

$$\lim_{h \rightarrow 0} \frac{1}{h} \mathbb{E} [m([t, t+h[, A) | \mathcal{F}_t] = \lambda_t p(A).$$

The quadratic kernel $K : \mathbb{R}^+ \times \mathbb{R}^+ \rightarrow \mathbb{R}$ is assumed to satisfy

- Symmetry: $\forall s, t \geq 0, K(t, s) = K(s, t)$,
- Positivity: $\forall f \in L^2(\mathbb{R}^+), \int_0^{+\infty} \int_0^{+\infty} K(t, s) f(t) f(s) dt ds \geq 0$,
- Non-explosion: $\int_0^{+\infty} |K(t, t)| dt < +\infty$.

K defines an integral operator $T_K : L^2(\mathbb{R}^+) \rightarrow L^2(\mathbb{R}^+)$ which maps $f \in L^2(\mathbb{R}^+)$ to $T_K f : t \mapsto \int_0^{+\infty} K(t, s) f(s) ds$. If K is continuous, this operator is Hilbert-Schmidt and thus compact and one has $K(t, t) \geq 0$ for all $t \geq 0$ (see [11]). We define the trace of K

$$\text{Tr}(K) = \int_0^{+\infty} K(t, t) dt < +\infty.$$

The leverage kernel $L : \mathbb{R}^+ \rightarrow \mathbb{R}$ is assumed to be a measurable function. By analogy with QARCH models (see [14]) it should be dominated by K in some way to ensure the positivity of the intensity λ_t (see footnote 1 above). Since the leverage kernel is found empirically negligible in the sequel, we leave this positivity condition for future research.

2.4 Necessary condition for time stationarity

In the case of linear Hawkes processes, it has been shown that stationarity is obtained as soon as the norm of the kernel verifies $\|\phi\|_1 < 1$. Intuitively, this means that each event triggers on average less than one child event, so that the clusters generated by each ancestor eventually die out. If this condition is violated, the probability that an ancestor generates an infinite number of events is non-zero, which can result in a stationary process only in the case $\|\phi\|_1 = 1$ and $\lambda_\infty = 0$ studied in [10], see also [20]. Because of the quadratic feedback, the QHawkes process cannot be interpreted as a simple branching process, making things somewhat trickier. The goal of this section is to find a necessary condition for (first order) time stationarity.

We define the jump process (N_t) that has the same jump times as (P_t) , with $\Delta N_\tau = (\Delta P_\tau)^2 / \psi^2$ ($= 1$ iff $\Delta P_\tau = \pm \psi$) for any jump time τ , and re-write Equation (2) as

$$\lambda_t = \lambda_\infty + \mathcal{L}_t + H_t + 2M_t \tag{5}$$

with the notations

$$\begin{cases} \mathcal{L}_t = \frac{1}{\psi} \int_{-\infty}^t L(t-u) dP_u & \text{(leverage)} \\ H_t = \int_{-\infty}^t K(t-u, t-u) dN_u & \text{(Hawkes/diagonal)} \\ M_t = \frac{1}{\psi^2} \int_{-\infty}^t \Theta_{t,u} dP_u & \text{(off-diagonal)} \end{cases}$$

where $\Theta_{t,u} = \int_{-\infty}^{u-} K(t-u, t-r) dP_r$ is $(\mathcal{F}_u)_{u \leq t}$ -adapted for t fixed. Since P is a martingale, one has $\mathbb{E}[M_t] = 0$ and $\mathbb{E}[\mathcal{L}_t] = 0$. Therefore,

$$\begin{aligned} \mathbb{E}[\lambda_t] &= \lambda_\infty + \frac{1}{\psi^2} \mathbb{E} \left[\int_{\mathbb{R}} \int_{-\infty}^t K(t-s, t-s) \xi^2 m(ds, d\xi) \right] \\ &= \lambda_\infty + \mathbb{E} \left[\int_{-\infty}^t K(t-s, t-s) \lambda_s ds \right] \end{aligned}$$

by definition of the punctual Poisson measure $m(ds, d\xi)$. We obtain

$$\mathbb{E}[\lambda_t] = \lambda_\infty + \int_{-\infty}^t K(t-s, t-s) \mathbb{E}[\lambda_s] ds.$$

A necessary condition for the process $(\lambda_t)_{t \in \mathbb{R}}$ to be in a stationary state is that its expected value $\bar{\lambda} \equiv \mathbb{E}[\lambda_t]$ is constant, positive and finite. This yields $\bar{\lambda} = \lambda_\infty + \bar{\lambda} \text{Tr}(K)$, thus if $\lambda_\infty > 0$,

$$\bar{\lambda} = \frac{\lambda_\infty}{1 - \text{Tr}(K)}.$$

This leads to the necessary stationarity condition³

$$\lambda_\infty > 0 \text{ and } \text{Tr}(K) < 1 \quad (6)$$

$$\text{or } \lambda_\infty = 0 \text{ and } \text{Tr}(K) = 1. \quad (7)$$

In the special case of the ZHawkes process, the endogeneity ratio is then given by:

$$\text{Tr}(K) = \|\phi\|_1 + \|k^2\|_1 \equiv n_H + n_Z,$$

where n_H is the standard ‘‘Hawkes norm’’ while $n_Z \equiv \|k^2\|_1$ is the ‘‘Zumbach norm’’.

The existence of a finite average intensity $\bar{\lambda}$ is of course necessary for the process to reach a stationary state. However, the existence of higher moments of the intensity require stronger conditions on $K(t, s)$, similarly to the QARCH case studied in [14]. In particular, the decay of the off-diagonal part of K must be fast enough to ensure the existence of two-point and three-point correlations of the process (see below).

2.5 Auto-correlation structure in the QHawkes model

It is quite useful for such type of models to investigate the relation between the input kernels and the auto-correlation functions of the generated process. Indeed, since the latter is directly observable on the data, the underlying kernels can then be obtained by inverting such relations. For linear Hawkes processes, one finds a Wiener-Hopf equation that relates the two-points correlation function to the 1-d kernel [2]. In our case, one also needs to consider the three-points correlation function, which will lead to a set of closed relations.

2.6 Exact set of equations

We take the model with no leverage, $L \equiv 0$. Equation (5) becomes (see notations above):

$$\lambda_t = \lambda_\infty + H_t + 2M_t.$$

We define for $\tau \neq 0$ and $\tau_1 > 0, \tau_2 > 0, \tau_1 \neq \tau_2$, the correlation functions

$$\begin{aligned} \mathcal{C}(\tau) &\equiv \mathbb{E} \left[\frac{dN_t}{dt} \frac{dN_{t-\tau}}{dt} \right] - \bar{\lambda}^2 = \mathbb{E} \left[\lambda_t \frac{dN_{t-\tau}}{dt} \right] - \bar{\lambda}^2, \\ \mathcal{D}(\tau_1, \tau_2) &\equiv \frac{1}{\psi^2} \mathbb{E} \left[\frac{dN_t}{dt} \frac{dP_{t-\tau_1}}{dt} \frac{dP_{t-\tau_2}}{dt} \right] = \frac{1}{\psi^2} \mathbb{E} \left[\lambda_t \frac{dP_{t-\tau_1}}{dt} \frac{dP_{t-\tau_2}}{dt} \right]. \end{aligned} \quad (8)$$

\mathcal{C} is then extended continuously at zero, as in [22]. Let us note that by construction \mathcal{C} is even and \mathcal{D} is symmetric. One finds the following exact equations between the auto-correlation functions (\mathcal{C}, \mathcal{D}) and the kernel K (cf. the derivation in Appendix A):

$$\mathcal{C}(\tau) = \kappa \bar{\lambda} K(\tau, \tau) + \int_0^\infty du K(u, u) \mathcal{C}(\tau + u) + 2 \int_{0^+}^\infty du \int_{u^+}^\infty dr K(\tau + u, \tau + r) \mathcal{D}(u, r), \quad (9)$$

$$\begin{aligned} \mathcal{D}(\tau_1, \tau_2) &= 2K(\tau_1, \tau_2) [\mathcal{C}(\tau_2 - \tau_1) + \bar{\lambda}^2] + \int_{(\tau_2 - \tau_1)^+}^{\tau_2} du K(\tau_2 - u, \tau_2 - u) \mathcal{D}(u - \tau_2 + \tau_1, u) \\ &\quad + 2 \int_{(\tau_2 - \tau_1)^+}^{+\infty} du K(\tau_1, \tau_2 + u) \mathcal{D}(\tau_2 - \tau_1, \tau_2 - \tau_1 + u), \end{aligned} \quad (10)$$

³In the case of linear Hawkes processes, this condition is also sufficient to obtain stationarity in the case $\text{Tr}(K) < 1$ (whereas the case $\text{Tr}(K) = 1$ is more subtle, see [10]).

where $\kappa = \frac{1}{\psi^4} \int_{\mathbb{R}} \xi^4 p(d\xi)$ is the fourth moment of price jumps ($\kappa = 1$ for constant price jumps). As $\mathcal{C}(\tau)$ and $\mathcal{D}(\tau_1, \tau_2)$ are directly measurable on the data, one can in principle infer the kernel $K(s, t)$ by inverting the above equations.

2.6.1 Asymptotic behaviour

Whereas the above equations (9) and (10) are difficult to solve in general, one can investigate the joint tail behaviour as $\tau \rightarrow \infty$ when both the kernel and the auto-correlation functions have power law decays. Let us assume that:

$$\left\{ \begin{array}{ll} K(\tau, \tau) & \underset{\tau \rightarrow \infty}{\sim} c_0 \tau^{-1-\epsilon} \quad (\text{diagonal, } \epsilon > 0) \\ K(\tau v_1, \tau v_2) & \underset{\tau \rightarrow \infty}{\sim} \tilde{K}(v_1, v_2) \tau^{-2\delta} \quad (\text{off-diagonal, } \delta > 1/2) \\ \mathcal{C}(\tau) & \underset{\tau \rightarrow \infty}{\sim} c_1 \tau^{-\beta} \quad (\text{2-points AC}) \\ \mathcal{D}(\tau, \tau) & \underset{\tau \rightarrow \infty}{\sim} c_2 \tau^{-\beta'} \quad (\text{3-points AC, diagonal}) \\ \mathcal{D}(\tau v_1, \tau v_2) & \underset{\tau \rightarrow \infty}{\sim} \tilde{\mathcal{D}}(v_1, v_2) \tau^{-2\rho} \quad (\text{3-points AC, off-diagonal}) \end{array} \right. \quad (11)$$

where c_0, c_1, c_2 are constants and $\tilde{K}(v_1, v_2), \tilde{\mathcal{D}}(v_1, v_2)$ are bounded functions of (v_1, v_2) . The constraint $\epsilon > 0$ comes from the fact that $\|\phi\|_1$ must be finite, whereas $\delta > 1/2$ insures that the second and third moments are finite as well. We will furthermore assume (for simplicity) that $\epsilon < 1$, which is the interesting case in practice, and that the asymptotic behaviour of $K(\tau_1, \tau_2) \sim \tau_1^{-1-\epsilon}$ is restricted to a narrow channel around the diagonal $|\tau_1 - \tau_2| \ll \tau_1, \tau_2$, beyond which the off-diagonal power-law takes over.

The exponents β and ρ can then be related to δ and ϵ by plugging these ansatz into Eqs. (9) and (10) and carefully matching the asymptotic behaviours. One finds several possible phases for the auto-covariance structure:

1. In the *non critical* case $\text{Tr}(K) < 1$, we find:

$$\delta > (3 + \epsilon)/4 \Rightarrow \beta = 1 + \epsilon; \quad \beta' = 1 + \epsilon; \quad \rho = \delta, \quad (12)$$

$$(2 + \epsilon)/3 < \delta < (3 + \epsilon)/4 \Rightarrow \beta = 4\delta - 2; \quad \beta' = 1 + \epsilon; \quad \rho = \delta, \quad (13)$$

$$2 < \delta < (2 + \epsilon)/3 \Rightarrow \beta = 4\delta - 2; \quad \beta' = 3\delta - 1; \quad \rho = \delta, \quad (14)$$

The interpretation of these three phases is straightforward. In the first phase (12), the tail of the auto-correlation functions directly comes from the tail of the diagonal part of K : direct effects then dominate quadratic feedback effects. In the last two phases (13),(14) however, a more sophisticated phenomenon comes into play, as off-diagonal effects feedback in such a way that they generate correlations with slower decay than that of the diagonal part of the kernel itself. In these phases, there is a possibility that $\beta < 1$ (corresponds to a long memory process) provided $\frac{1}{2} < \delta < \frac{3}{4}$. This result is important as it means that QHawkes processes *need not be critical* (i.e. $\text{Tr}(K) = 1$) to generate long memory, unlike standard, linear Hawkes processes [10, 32, 20, 21].

2. In the *critical* case $\text{Tr}(K) \rightarrow 1, \lambda_\infty \rightarrow 0$, the situation is subtler, as in the standard Hawkes case where the relation between β and ϵ completely changes, and the condition $0 < \epsilon < 1/2$ must hold for the process to even exist [10]. In the present case, a similar mechanism operates and leads to:

$$\delta > 3/4 \Rightarrow \beta = 1 - 2\epsilon; \quad \beta' = 1 - \epsilon; \quad \rho = \delta, \quad (15)$$

$$2/3 < \delta < 3/4 \Rightarrow \beta = 4\delta - 2\epsilon - 2; \quad \beta' = 1 - \epsilon; \quad \rho = \delta. \quad (16)$$

$$(1 + \epsilon)/2 < \delta < 2/3 \Rightarrow \beta = 4\delta - 2\epsilon - 2; \quad \beta' = 3\delta - \epsilon - 1; \quad \rho = \delta. \quad (17)$$

provided $0 < \epsilon < 1/2$ and $\delta > (1 + \epsilon)/2$, otherwise the critical process does not exist or is trivial. So in this critical case, the process is always long-memory (i.e. $\beta < 1$), or ceases to exist, as for the linear Hawkes process.

3 The intra-day QHawkes model

3.1 QHawkes as a limit of QARCH

In this section we investigate the link between the QHawkes model given by (2) and the discrete QARCH model introduced by Sentana in [33], and revisited in depth in [14]. This will give us a way to calibrate the QHawkes on discretely sampled price time series. For a fixed time step $\Delta > 0$, we define for all $t \in \mathbb{R}$:

- the price (or log-price) increment between time t and time $t + \Delta$: $r_t^\Delta = P_{t+\Delta} - P_t$,
- the volatility at time t : $\sigma_t^\Delta = \sqrt{\mathbb{E} \left[r_t^{\Delta^2} | \mathcal{F}_t \right]}$.

The QHawkes model appears as the limit (in some sense) when $\Delta \rightarrow 0^+$ of the QARCH model

$$\sigma_t^{\Delta^2} = \sigma_\infty^{\Delta^2} + \sum_{\tau \geq 1} L^\Delta(\tau) r_{t-\tau\Delta}^\Delta + \sum_{\tau, \tau' \geq 1} K^\Delta(\tau, \tau') r_{t-\tau\Delta}^\Delta r_{t-\tau'\Delta}^\Delta, \quad (18)$$

where $\sigma_\infty^{\Delta^2} = \psi^2 \lambda_\infty \Delta$, $L^\Delta(\tau) = L(\tau\Delta) \Delta$ and $K^\Delta(\tau, \tau') = K(\tau\Delta, \tau'\Delta) \Delta$. Indeed, for $t \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E} \left[r_t^{\Delta^2} | \mathcal{F}_t \right] &= \psi^2 \mathbb{P} (P_{t+\Delta} - P_t \neq 0 | \mathcal{F}_t) + o(\Delta) \\ &= \psi^2 \lambda_t \Delta + o(\Delta), \end{aligned}$$

which implies the scaling:

$$\frac{\sigma_t^{\Delta^2}}{\Delta} \xrightarrow{\Delta \rightarrow 0^+} \psi^2 \lambda_t.$$

Thanks to this link between the two models, it is possible to calibrate a QARCH model on intra-day 5 minutes bin returns, as in [8, 14], and obtain some qualitative and quantitative insight on the structure of the underlying QHawkes model. Indeed, the direct calibration of the latter would be significantly harder – more noisy and computationally more demanding – and certainly beyond the scope of the present paper.

3.2 Intra-day calibration of a QARCH model

QARCH models have mainly been calibrated on daily data so far ([33], [14]). To give a starting point to our study of quadratic effects in high-frequency volatility, we calibrate a discrete QARCH on intra-day five-minute returns.

3.2.1 Dataset and notations

We consider the same dataset as in [1], which is composed of stock prices on intra-day five-minute bins. It includes 133 stocks of the New York Stock Exchange, that have been traded without interruption between 1 January 2000 and 31 December 2009. This yields 2499 trading days, with 78 five-minute bins per day. For each bin, the open, close, high and low prices ($O, C, H, L > 0$) are available. We consider the log-price process and define on each bin:

- The return $r = \ln(C/O)$.
- The Rogers-Satchell volatility $\sigma^{\text{RS}} = \sqrt{\ln(H/O) \times \ln(H/C) + \ln(L/O) \times \ln(L/C)}$.

3.2.2 Normalization procedure

To be able to consider that intra-day prices are (approximately) independent realizations of a stationary stochastic process, we need to normalize the data carefully. As a matter of fact, strong intra-day seasonalities may corrupt the calibration results. This can be avoided to some extent through a cross-sectional and historical normalization. We take advantage of our large dataset to compute a cross-sectional intra-day volatility pattern for each trading day and we normalize the

returns by this pattern, which dampens the effect of collective shocks on a given day. On the other hand, we use the intra-day/overnight model volatility model of [9] to factor out daily feedback effects and focus on pure intra-day dynamics. To fully explain our normalization protocol, we introduce the following notations:

- The 5-min bin index $1 \leq b \leq 78$, the day index $1 \leq t \leq 2499$ and the stock index $1 \leq u \leq 133$.
- The empirical averages: $\langle x_{u,t,\cdot} \rangle$ means conditional average of x over bins, for stock u and day t fixed; $\langle x_{u,\cdot,b} \rangle$ and $\langle x_{\cdot,t,b} \rangle$ are defined similarly as the conditional averages over days/stocks; $\langle x \rangle = \langle x_{\dots} \rangle$ means average of x over stocks, days and bins.

We compute the cross-sectional volatility pattern of day t , that we use to normalize the data of stock u , as:

$$b \in \{1, \dots, 78\} \mapsto v_{u,t}(b) \equiv \sqrt{\langle r_{u' \neq u,t,b}^2 \rangle}.$$

For stock u , the value $r_{u,t,b}^2$ is excluded from the average, so that the normalization protocol does not cap the large returns of stock u artificially. We also consider the open-to-close volatility $\sigma_{u,t}^D$ of day t for stock u , as computed by the intra-day/overnight model of [9] with the data of stock u over the days $\{1, \dots, t-1\}$. For $t=1$, we fix $\sigma_{u,1}^D = 1$.

The normalization protocol is as follows: $\forall u, t, b$,

- $r_{u,t,b} \leftarrow r_{u,t,b}/\sigma_{u,t}^D$, $\sigma_{u,t,b}^{\text{RS}} \leftarrow \sigma_{u,t,b}^{\text{RS}}/\sigma_{u,t}^D$, (normalization by open-to-close volatility)
- $r_{u,t,b} \leftarrow r_{u,t,b}/v_{u,t}(b)$, $\sigma_{u,t,b}^{\text{RS}} \leftarrow \sigma_{u,t,b}^{\text{RS}}/v_{u,t}(b)$. (cross-sectional intra-day normalization)

We further exclude trading days that involve at least one bin where the absolute return is greater than the average plus six standard deviations. This represents approximately 7% of trading days, i.e. one day every three weeks. Combined with the cross-sectional pattern normalization, this data treatment strongly dampens the impacts of exceptional news events, which would require a special treatment and that we do not aim to model here. Eventually, we set the mean of the squares to one and the average return to zero to make the stock universe more homogeneous: $\forall u, t, b$,

- $r_{u,t,b} \leftarrow r_{u,t,b}/\sqrt{\langle r_{u,\dots}^2 \rangle}$, so that $\langle r^2 \rangle = 1$,
- $\sigma_{u,t,b}^{\text{RS}} \leftarrow \sigma_{u,t,b}^{\text{RS}}/\sqrt{\langle \sigma_{u,\dots}^{\text{RS}2} \rangle}$, so that $\langle \sigma^{\text{RS}2} \rangle = 1$,
- $r_{u,t,b} \leftarrow r_{u,t,b} - \langle r_{u,\dots} \rangle$ so that $\langle r \rangle = 0$.

3.2.3 Calibration results

The calibration process is similar to [14] and [9]. A first estimate of the kernels is obtained with the Generalized Method of Moments, which uses a set of correlation functions that are empirically observable. Then, using this estimate as a starting point, we use Maximum Likelihood Estimation, assuming that the residuals are t-distributed (which accounts for fat tails that remain in the residuals). This second step significantly improves the precision of the calibration results, compared to a solo GMM estimation.

We find it worth to notice that as opposed to the daily calibration results of [14], a clear off-diagonal structure appears in the feedback matrix in the intra-day case (see Figure 1). Also, the intra-day leverage kernel is found to be close to zero, justifying the fact that we mainly consider $L \equiv 0$ throughout the paper. The spectral decomposition of quadratic kernel (see Figure 2) suggests that K is the superposition of a positive rank-one matrix and a diagonal one. Indeed, we obtain to a good approximation (see Figure 3)

$$K(\tau, \tau') \approx \phi(\tau)\delta_{\tau-\tau'} + k(\tau)k(\tau')$$

where

$$\phi(\tau) = g\tau^{-\alpha}, \quad k(\tau) = k_0 \exp(-\omega\tau),$$

with $g = 0.09$, $\alpha = 0.60$, $k_0 = 0.14$, $\omega = 0.15$. Note that $\omega = 0.15$ corresponds to a characteristic time of about thirty minutes (bin size $\times \omega^{-1}$) for the decay of the off-diagonal component. We

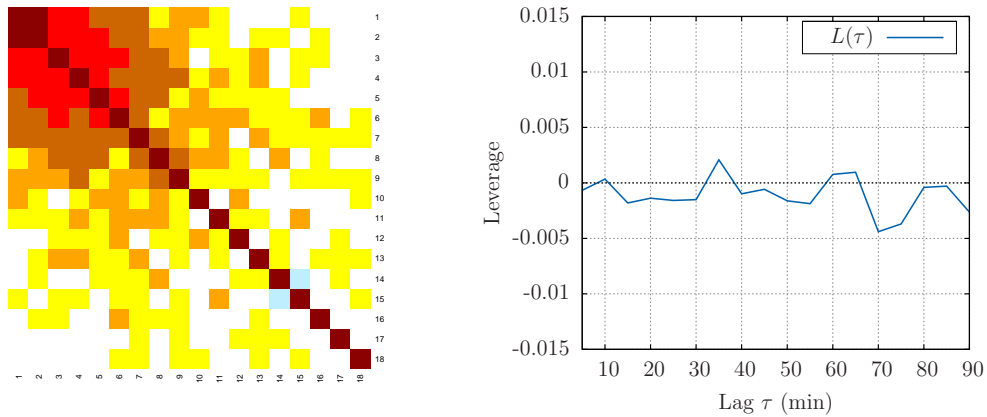


Figure 1: QARCH kernels calibrated on five-minute intra-day returns for US stocks. The maximum lag is 18 bins, i.e. one hour and a half of trading time. Left: heatmap of the quadratic kernel. White coefficients are close to zero, blue ones are negative and yellow/orange/red ones are positive, with darker shades as they increase in absolute value. We see that all the significant coefficients are positive, with a non-negligible off-diagonal component. Right: leverage kernel. It is hardly distinct from zero and can be considered as pure noise (as opposed to daily models where it is significantly negative).

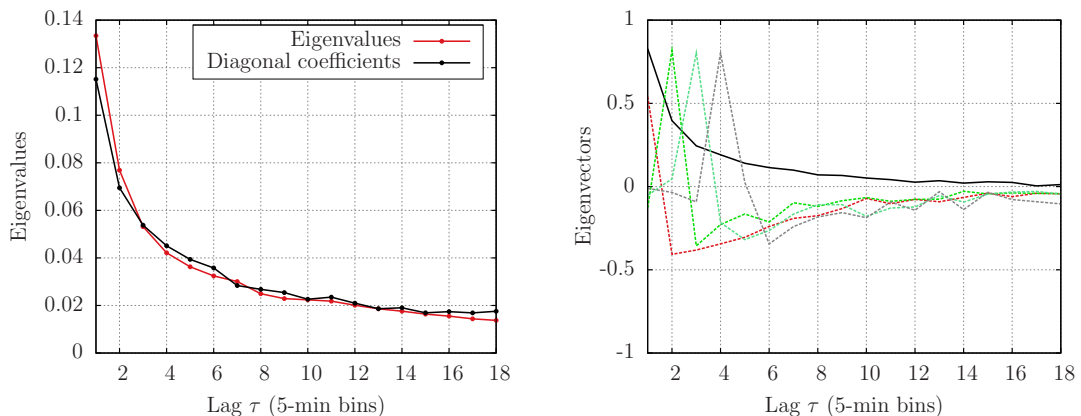


Figure 2: Spectral decomposition of the quadratic QARCH kernel. Left: ranked eigenvalues (plain dark line) and diagonal coefficients (dashed). One can see that the diagonal coefficients are very close to the eigenvalues, except for the first eigenvalue which is significantly larger than the maximum of the diagonal. Right: eigenvectors corresponding to the five largest eigenvalues. The first eigenvector (plain dark line) is a positive decaying kernel, the others are close to the canonical vectors $e_i(\tau) = \delta_{i-\tau}$.

then fix the off-diagonal part of the kernel K to its fitted value $k(\tau)k(\tau') = k_0^2 \exp(-\omega(\tau + \tau'))$, and we recalibrate the diagonal of K with a longer maximum lag of 60 bins (five hours of trading). We obtain

$$\phi_{\text{IR}}(\tau) = g' \tau^{-\alpha'}$$

with the new coefficients $g' = 0.09$, $\alpha' = 0.76$, not far from those obtained above on a shorter time span. The residuals ξ_t of the QARCH model, defined by

$$r_t = \sigma_t \xi_t,$$

where r_t is the five-minute return and σ_t is the QARCH volatility, are modeled with Student's t-distribution. The calibration of the model with $K(\tau, \tau') = \phi(\tau)\delta_{\tau-\tau'} + k(\tau)k(\tau')$ yields $\nu \approx 7.9$ degrees of freedom for the residuals, which gives a kurtosis $\kappa \approx 4.5$. This has to be compared with

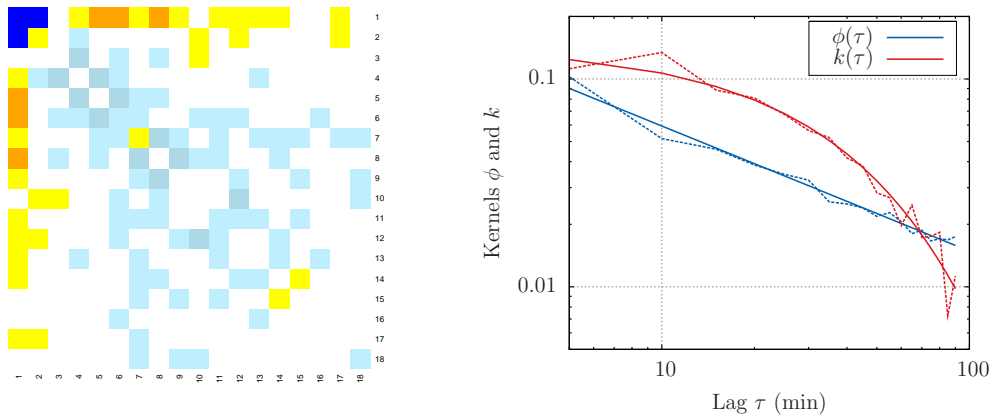


Figure 3: Fit of the kernel K by the sum of a power-law diagonal matrix and an exponential rank-one matrix. Left: heatmap of the difference between the fitted matrix and the original one. The coefficients are small (white or lightly-colored) except for the upper-left corner: the original matrix features a stronger short-term feedback. Right: kernels $\phi(\tau)$ and $k(\tau)$ that minimize the matrix distance $\sum [K(\tau, \tau') - \phi(\tau)\delta_{\tau-\tau'} - k(\tau)k(\tau')]^2$. The rank-one kernel k is plotted in red (and is larger for small τ 's), and the diagonal kernel ϕ is plotted in blue, both in log-log scale. The dashed lines are the power-law fit for $\phi(\tau)$ with exponent $\alpha = 0.6$, and the exponential fit for $k(\tau)$ with characteristic time about 30 min.

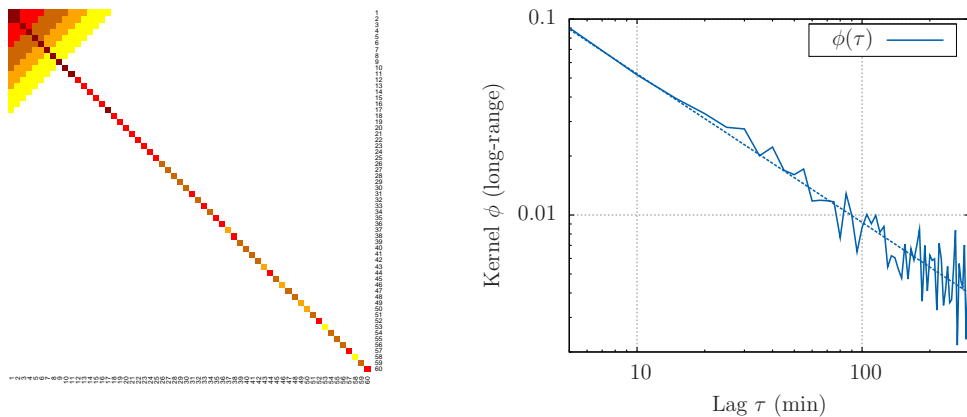


Figure 4: Long-range kernel K . Left: heatmap of the long-range kernel, with the off-diagonal fixed as its exponential rank-one fit, and with the diagonal calibrated with no constraints. Right: Hawkes kernel $\phi(\tau) = K(\tau, \tau) - k^2(\tau)$ fitted on 60 bins. The kernel $\phi(\tau)$ is plotted in log-log scale, with its power-law fit with exponent $\alpha' = 0.76$ (dashed).

the tail exponent ν_r of r_t itself, which is, as is well known, in the range $3 \rightarrow 4$, see also Fig. 5 below. Since ν is more than twice ν_r , the QARCH model with Gaussian residuals and this specific form of K accounts, to a good extent, for the fat tails of five-minute returns, that appear to be nearly entirely induced by the quadratic feedback mechanism. We will justify theoretically and numerically why this is the case in Sections 4 and 5.

In the QARCH model, the endogeneity ratio of the volatility (i.e. the proportion of the volatility that stems from feedback effects) is given by the trace $\text{Tr}(K)$ of the quadratic kernel. With our parametrization and a maximum lag of $q \geq 1$, one has

$$\text{Tr}(K) = \sum_{\tau=1}^q \phi(\tau) + \sum_{\tau=1}^q k^2(\tau).$$

We use the fits $k(\tau) = k_0 \exp(-\omega\tau)$ and $\phi_{\text{lr}}(\tau) = g'\tau^{-\alpha'}$ to compute $\text{Tr}(K)$ for $q = 78$, which is

the total number of five-minute bins in a trading day. We obtain

$$\sum_{\tau=1}^q \phi(\tau) \simeq 0.74, \quad \sum_{\tau=1}^q k^2(\tau) \simeq 0.06 \quad \Rightarrow \quad \text{Tr}(K) \simeq 0.80.$$

This endogeneity ratio implies that 80% of the intra-day volatility is due to endogenous feedback effects. Interestingly, it is close to the value obtained for QARCH and ARCH models at a daily time scale, see [14] and [9]. Note that although high, the endogeneity ratio is significantly below the critical limit $\text{Tr}(K) = 1$, which is the value found by calibrating a standard linear Hawkes process to activity data on much longer time horizons: Ref. [20, 21] report $n_H \approx 0.9$ on a time window of a few hours, and $n_H \approx 0.99$ when the kernel is extended to 40 days. We discuss this issue further in Section 4.2.

4 Volatility distribution in the ZHawkes model

4.1 SDE in the exponential case

If the kernels $\phi(\cdot)$ and $k(\cdot)$ of the ZHawkes model have an exponential form, the process is Markovian and one can write a stochastic differential equation to describe its evolution. Although this assumption is only justified for k , this case allows one to gain a good intuition on the model, so we investigate this limit in details. It also turns out that the Markovian case is actually extremely interesting mathematically.

For the sake of simplicity, let us assume that the price jumps are binary $\xi = \pm\psi$, and we set $\psi = 1$ without loss of generality. Besides, we note $k(t) = \sqrt{2n_Z\omega} \exp(-\omega t)$ and $\phi(t) = n_H\beta \exp(-\beta t)$, where n_H is the Hawkes norm and n_Z the Zumbach norm. We require:

$$\text{Tr}(K) = n_H + n_Z < 1.$$

Then the model can be written in this case: $\lambda_t = \lambda_\infty + H_t + Z_t^2$ where

$$\begin{cases} dH_t &= \beta [-H_t dt + n_H dN_t], \\ dZ_t &= -\omega Z_t dt + k_0 dP_t \end{cases} \quad (19)$$

The processes N and P jump simultaneously with intensity λ_t and amplitudes $\Delta N_\tau = 1$ and $\Delta P_\tau = \pm 1$ with equal probability. Although quite simple, this system of jump SDEs lacks tractability compared to a continuous diffusion. Thus, we turn to the low-frequency asymptotics that one obtains as the number of jumps in a given time window becomes large, while their amplitudes are scaled down accordingly. This is the object of the following section.

4.2 Low-frequency asymptotics

The low-frequency asymptotics of nearly critical Hawkes processes with short-ranged kernels have been investigated in details by Jaisson and Rosenbaum [25, 26]. They show that for suitable scaling and convergence to the critical point $n_H = 1$, the short memory Hawkes-based price process of Bacry et al. [6] converges towards a Heston process (since the Hawkes intensity converges towards a CIR volatility process). The same authors [26] show that when the kernel exhibits power-law behaviour $\phi(t) \sim t^{-1-\epsilon}$ with $1/2 < \epsilon < 1$, the limiting process for the intensity is a fractional Brownian motion with Hurst exponent $H = \epsilon - \frac{1}{2}$. When ϵ is close to $1/2$, as empirical data suggests [20], the roughness of the latter process is in agreement with the empirical results of [4, 19] who find a Hurst exponent H close to zero the *log*-volatility ($H = 0$ for the multifractal model of [4]). However, it is unclear how the Hawkes process intensity can be identified with the *log*-volatility. A fat-tailed behaviour cannot be reproduced by a simple, linear Hawkes process, as it is absent from Heston-CIR processes (see also below).

Here, we want to investigate the low-frequency asymptotics of the Markovian ZHawkes model, which, as we shall see, reveals very interesting new features, induced by quadratic feedback effects.

Choosing a time scale $T > 0$ that will eventually diverge, we define the processes $\bar{H}_t^T = H_{tT}$, $\bar{Z}_t^T = Z_{tT}$, $\bar{N}_t^T = N_{tT}$ and $\bar{P}_t^T = P_{tT}$, with the parameters β_T and ω_T that may depend on T , but with fixed endogeneity parameters n_H and n_Z : Equation (19) gives

$$\begin{cases} d\bar{H}_t^T &= -\beta_T [\bar{H}_t^T T dt + n_H d\bar{N}_t^T], \\ d\bar{Z}_t^T &= -\omega_T \bar{Z}_t^T T dt + \gamma_T d\bar{P}_t^T, \end{cases} \quad (20)$$

where $\gamma_T^2 := 2\omega_T n_Z$ and the common jump intensity of \bar{N}^T and \bar{P}^T is $T \times [\lambda_\infty + \bar{H}_t^T + (\bar{Z}_t^T)^2]$. Since the signs of the jumps of \bar{P}^T are assumed to be unpredictable and equal to ± 1 , the infinitesimal generator of the process is given by

$$\begin{aligned} \mathcal{A}^T f(h, z) &= -\beta_T h T \partial_h f(h, z) - \omega_T z T \partial_z f(h, z) \\ &+ T [\lambda_\infty + h + z^2] \left\{ \frac{1}{2} f(h + n_H \beta_T, z + \gamma_T) + \frac{1}{2} f(h + n_H \beta_T, z - \gamma_T) - f(h, z) \right\} \end{aligned} \quad (21)$$

for any functions f twice continuously differentiable on $(0, +\infty) \times \mathbb{R}$. We now consider the following scaling

$$\beta_T = \bar{\beta}/T, \quad \omega_T = \bar{\omega}/T, \quad (22)$$

with $\bar{\beta}, \bar{\omega} > 0$. Since we fixed the values of n_H and n_Z , our procedure can be called a ‘‘constant endogeneity rescaling’’, as opposed to the scaling used by Jaisson and Rosenbaum in [25] and [26], where the endogeneity ratio n_H of the process needs to converge to unity as T goes to infinity. Our choice is partly motivated by the calibration results of Section 3.2 for intra-day returns, that yield an endogeneity ratio in the range 0.7 – 0.9, close to what is obtained at the daily time scale in [14] and [9], and significantly away from the critical value $n_H = 1$. Equations (21) and (22) then combine as

$$\begin{aligned} \mathcal{A}^T f(h, z) &= -\bar{\beta} h \partial_h f(h, z) - \bar{\omega} z \partial_z f(h, z) \\ &+ T [\lambda_\infty + h + z^2] T \left\{ \frac{1}{2} f\left(h + n_H \frac{\bar{\beta}}{T}, z + \frac{\bar{\gamma}}{\sqrt{T}}\right) + \frac{1}{2} f\left(h + n_H \frac{\bar{\beta}}{T}, z - \frac{\bar{\gamma}}{\sqrt{T}}\right) - f(h, z) \right\}, \end{aligned}$$

where we introduced $\bar{\gamma} = \sqrt{2n_Z \bar{\omega}}$. We turn to the low-frequency asymptotics. As T goes to infinity, one has

$$\frac{1}{2} f\left(h + n_H \frac{\bar{\beta}}{T}, z + \frac{\bar{\gamma}}{\sqrt{T}}\right) + \frac{1}{2} f\left(h + n_H \frac{\bar{\beta}}{T}, z - \frac{\bar{\gamma}}{\sqrt{T}}\right) - f(h, z) = \frac{n_H \bar{\beta}}{T} \partial_h f(h, z) + \frac{\bar{\gamma}^2}{2T} \partial_{zz}^2 f(h, z) + o\left(\frac{1}{T}\right),$$

therefore $\mathcal{A}^T f(h, z)$ converges to

$$\mathcal{A}^\infty f(h, z) = -\bar{\beta} [(1 - n_H)h - n_H(\lambda_\infty + z^2)] \partial_h f(h, z) - \bar{\omega} z \partial_z f(h, z) + n_Z \bar{\omega} [\lambda_\infty + h + z^2] \partial_{zz}^2 f(h, z).$$

The operator \mathcal{A}^∞ is the infinitesimal generator of the diffusion

$$\begin{cases} d\bar{H}_t^\infty &= \left[-(1 - n_H) \bar{H}_t^\infty + n_H (\lambda_\infty + (\bar{Z}_t^\infty)^2) \right] \bar{\beta} dt, \\ d\bar{Z}_t^\infty &= -\bar{\omega} \bar{Z}_t^\infty dt + \bar{\gamma} \sqrt{\lambda_\infty + \bar{H}_t^\infty + (\bar{Z}_t^\infty)^2} dW_t, \end{cases} \quad (23)$$

where W is a standard Brownian motion. A standard argument of Kallenberg [27] (Theorem 19.25) then gives the convergence of the process (\bar{H}^T, \bar{Z}^T) to $(\bar{H}^\infty, \bar{Z}^\infty)$ as T goes to infinity. Hence, one does *not* need that the norm of the process tends to 1 (i.e. that the process is nearly critical) for a non-degenerate limit process to be obtained. The above limiting process is the major result of this section. Although it was derived for a Markovian ZHawkes process, we believe that this is the limiting process for the whole class of non-critical ZHawkes processes with short memory, and is the analogue of the Heston-CIR limiting process for Hawkes, as in [25]. The limiting behaviour corresponding to long-memory/critical ZHawkes processes, in the spirit of [26], is left for future investigations. We now investigate some of the properties of the limiting process, Eq. (23), in particular the induced tail of the volatility distribution.

4.3 Tail of the volatility distribution

From now on we drop the superscript ∞ on \bar{H} and \bar{Z} ; the fact that we are studying the limiting process is implied. Let us note first that there is no Brownian part in the SDE for \bar{H} so that it can be solved explicitly as a deterministic function of $(\bar{Z}_s)_{s \leq t}$:

$$\bar{H}_t = \bar{H}_\infty + n_H \bar{\beta} \int_{-\infty}^t \exp(-(1 - n_H)\bar{\beta}(t - s)) \bar{Z}_s^2 ds; \quad \bar{H}_\infty := \frac{\lambda_\infty}{1 - n_H}$$

In the considered limit, \bar{H}_t can thus be written as the sum of a constant term and an exponential moving average of the square of \bar{Z}_s . We get the autonomous, but non-Markovian SDE for \bar{Z}_t :

$$d\bar{Z}_t = -\bar{\omega} \bar{Z}_t dt + \bar{\gamma} \sqrt{\bar{H}_\infty + \bar{Z}_t^2 + n_H \bar{\beta} \left[\int_{-\infty}^t \exp(-(1 - n_H)\bar{\beta}(t - s)) \bar{Z}_s^2 ds \right]} dW_t. \quad (24)$$

4.3.1 ZHawkes without Hawkes

We first consider the simpler case where the Hawkes feedback is zero, i.e. $n_H = 0$. This corresponds to the case where only the Zumbach term is present in the starting model, i.e. $\lambda_t = \lambda_\infty + Z_t^2$ in Equation (4). As we see in the sequel, this simpler model is still rich enough to reproduce some interesting empirical properties of the volatility process. One gets:

$$d\bar{Z}_t = -\bar{\omega} \bar{Z}_t dt + \bar{\gamma} \sqrt{\lambda_\infty + \bar{Z}_t^2} dW_t, \quad (25)$$

which is a particular case of Pearson diffusions, which are extensively described and classified by Forman and Sorensen [18]. The process $\bar{Z}/\sqrt{\lambda_\infty}$ fits in Case 3 of their classification (see [18] Section 2.1), with the dictionary $\mu \rightarrow 0, \theta \rightarrow \bar{\omega}$ and $a \rightarrow n_Z$. Therefore, \bar{Z}_t is ergodic and its stationary law is a Student t-distribution with $1 + 1/n_Z$ degrees of freedom and scale parameter $\sqrt{n_Z \lambda_\infty / (1 + n_Z)}$. This implies that stationary law of the square of \bar{Z}^∞ is a F-distribution with 1 and $1 + 1/n_Z$ degrees of freedom, and scale parameter $n_Z \lambda_\infty / (1 + n_Z)$. We will denote as

$$V_t = \psi^2 [\lambda_\infty + \bar{Z}_t^2]$$

the low-frequency squared volatility of the price (we reintroduced the jump size ψ for completeness). A straightforward change of variables yields the stationary density $q(v)$ of the process V as:

$$q(v) = \frac{\Gamma\left(1 + \frac{1}{2n_Z}\right)}{\Gamma\left(\frac{1}{2} + \frac{1}{2n_Z}\right) \sqrt{\pi v_\infty (v - v_\infty)}} \left(\frac{v_\infty}{v}\right)^{\left(1 + \frac{1}{2n_Z}\right)} \mathbb{1}_{\{v > v_\infty\}} \quad (26)$$

where $v_\infty = \lambda_\infty \psi^2$ is the baseline level of the squared volatility. For the tail exponent of the distribution of V_t , we get a power-law tail:

$$q(v) \underset{v \rightarrow +\infty}{\sim} C v^{-\left(\frac{3}{2} + \frac{1}{2n_Z}\right)} \quad (27)$$

with C an explicit constant. We find this result interesting for two reasons. First, one obtains a power-law behavior that emerges naturally from the fact that since the volatility behaves as $|\bar{Z}_t|$ for large values of \bar{Z}_t , the process describing its dynamics is simply a multiplicative Brownian motion with drift (see 25). This is at variance with the ‘‘diagonal’’ Hawkes counterpart of [25] where the coefficient in front of the Brownian noise is only the *square-root* of the volatility, which inevitably leads to a process that has a characteristic scale and thin tails. Second, the stationary distribution of V only depends on the Zumbach norm n_Z , that can be seen as the endogeneity of the process. This last result suggests that, similar to Hawkes processes where the asymptotic properties only depend on the norm n_H as soon as the kernel is short-ranged, the distribution (26) of the squared volatility should hold for any short-ranged kernel.

Another remark is that as soon as $n_Z \geq 1/3$, the variance σ_V^2 of the activity V explodes while its mean μ_V remains finite up to $n_Z \rightarrow 1^-$. Now, when fitting the time series generated by this

process using a simple Hawkes process, one finds $n_H \approx 1 - \sqrt{\mu_V(W)/\sigma_V^2(W)}$ for a suitable choice of window size W (see [21]). Therefore, the vanishing of the mean/variance ratio necessarily imposes that the fitted Hawkes process must be critical, i.e. $n_H = 1$! What we argue here is that this *apparent* criticality may in fact be induced by quadratic feedback effects, but does not necessarily imply that the true underlying process is critical.

Finally, note that in the diffusive limit where the price process satisfies the equation $d\bar{P}_t^\infty = \sqrt{V_t}dW_t$, the asymptotic stationary distribution for the returns is given by:

$$p(r) \underset{|r| \rightarrow \infty}{\sim} \frac{C'}{|r|^{1+\nu}}; \quad \nu \equiv 1 + \frac{1}{n_Z}.$$

The fat-tail volatility that is generated by our model naturally produces a fat-tail distribution of instantaneous returns, with exponent ν for the cumulative distribution equal to $1 + 1/n_Z \geq 2$. The more endogenous, the fatter the tails for the returns: this interpretation seems intuitive. For a critical process, $n_Z = 1$, the tail is such that the volatility of the returns diverges. A tail exponent for the cumulative distribution $\nu \approx 3$ (the so-called “inverse cubic law”, observed on a large universe of traded products) is obtained for $n_Z = 0.5$. Note however that the value of n_Z obtained above from calibrating the model is much smaller, $n_Z \approx 0.06$, leading to $\nu \approx 18$, far too large to explain the tail of financial returns. We will see now that, quite interestingly, the interaction with a non-critical Hawkes kernel can substantially reduce the value of ν .

4.3.2 ZHawkes with Hawkes

The case when $n_H > 0$ is more complicated but, remarkably, the tail exponent of the activity distribution $q(v)$ can still be analytically computed in some limits. The idea is to realize that when $\bar{Z} \rightarrow \infty$, the distribution of \bar{H} conditional to a certain large value of $\bar{Y} := \bar{Z}^2$ is of the form:

$$\Pi(\bar{H}|\bar{Y}) = \frac{1}{\bar{Y}} F\left(\frac{\bar{H}}{\bar{Y}}\right) + o(\bar{Y}); \quad (\bar{Y} \rightarrow \infty),$$

where $F(\cdot)$ is a certain scaling function which obeys a differential equation derived in Appendix B. Correspondingly, one can show that the far-tail of the distribution of $V_t = \psi^2 [\lambda_\infty + \bar{Z}_t^2]$ is still a power-law, given by:

$$q(v) \underset{v \rightarrow +\infty}{\sim} C'' v^{-\left(\frac{3}{2} + \frac{1}{2n_Z(1+a^*)}\right)}, \quad (28)$$

where C'' is another constant and a^* is defined as:

$$a^* = \int_0^\infty dx x F(x). \quad (29)$$

Introducing $\chi := \frac{2\bar{\omega}}{\beta}$ as the ratio of the correlation time scale of the Hawkes process to the one of the ZHawkes process, a full solution for F can be found in the two limits $\chi \rightarrow 0$ and $\chi \rightarrow \infty$, allowing one to fix the value of a^* . One finds (see Appendix B):

$$a^* \approx \frac{n_H}{1-n_H} \left[1 - \chi \frac{1-n_H-n_Z}{(1-n_H)^2} \right], \quad (\chi \rightarrow 0); \quad a^* \approx \frac{n_H}{\chi(1-n_Z)}, \quad (\chi, \chi n_Z \rightarrow \infty). \quad (30)$$

Two other limiting cases can be exactly solved: one is when $n_H \rightarrow 0$, one finds that $a^* \approx \frac{n_H}{\chi(1-n_Z)}$ still holds provided $a^* \ll 1$, and the other is $n_Z \rightarrow 0$, for which we find an explicit expression for a^* as the solution of a second degree equation (see Appendix B).

The corresponding exponent for the asymptotic tail of the cumulative distribution of returns is now given by:

$$\nu = 1 + \frac{1}{n_Z(1+a^*)}, \quad (31)$$

with:

- for $n_H = 0$ (ZHawkes without Hawkes), one recovers the previous case where $a^* = 0$ and $\nu = 1 + 1/n_Z$.

- for $0 < n_H < 1$ and $\chi \rightarrow 0$ (Hawkes much “faster” than ZHawkes), the exponent ν is *decreased* to $\nu = 1 + (1 - n_H)/n_Z + O(\chi)$.
- for $0 < n_H < 1$ and $\chi \rightarrow \infty$ (Hawkes much “slower” than ZHawkes), the exponent ν is again *decreased* from $\nu = 1 + 1/n_Z$ by an amount $\sim 1/\chi$.
- In the case $n_Z \rightarrow 0$, one finds $\nu = 1 + \frac{b}{n_Z}$, where b can be computed in terms of n_H and χ , see Appendix B.

The results of this section are, we believe, quite interesting. First, the two-dimensional limit process defined by Eqs. (23) leads to power-law tails for the volatility that can be exactly characterized in some limits. From a theoretical point of view, the possibility of computing exactly the tail exponent in this model is potentially important if our ZHawkes process turned out to be a central ingredient to model the dynamics of financial markets. Second, we have found that although the Hawkes kernel per-se does not lead to power-law tails (i.e., $\nu \rightarrow \infty$ when $n_Z \rightarrow 0$), the Hawkes kernel actually “cooperates” with the ZHawkes kernel to make the tails of the distribution fatter. The case of empirical interest is $n_Z = 0.06$, $n_H \approx 0.8$ leads to $\nu = 1 + (1 - n_H)/n_Z \approx 4$ for $\chi \rightarrow 0$, which indeed remains in the experimental range for a non-Markovian ZHawkes process with parameters calibrated on intraday data, as will be shown by numerical simulations in the next section.

We find this phenomenon quite remarkable: whereas the Hawkes feedback alone is not able to explain fat-tails, only a relatively small amount of quadratic (Zumbach) feedback generates power-law tails in the correct range (remember that $n_Z = 0.06 \ll n_H$). Note however that this ZHawkes family of models leads a continuously varying exponent (as a function of the parameters) rather than a fixed, universal exponent like in many physical situations. This begs the question: is there any mechanism that would explain why the feedback parameters n_Z, n_H, χ lie in a rather restricted interval, such as to explain the apparent universality of the tail exponent of (mature) financial markets?

5 Numerical simulation results

5.1 Empirical tails of the volatility process

In this section, we compare numerically the volatility process generated by the ZHawkes model, with a standard Hawkes-based price model and with the financial data studied in Section 3.2.1. We simulate a ZHawkes model with an exponential Zumbach part and a power-law Hawkes part, with parameters inspired by the QARCH calibration of Section 3.2: for t expressed in minutes,

$$\phi(t) = 0.0016 \times (1 + 0.01 \times t)^{-1.2}, \quad k(t) = 0.003 \times \exp(-0.03 \times t),$$

so that $n_H = 0.8$, $n_Z = 0.1$ and $\text{Tr}(K) = 0.9$. Note that to simulate a stationary ZHawkes model, we choose a decay exponent above 1 for ϕ , although the QARCH calibration suggests a slower decay for t corresponding to intraday time scales. Although not fully satisfactory, this is the simplest way to enforce stationarity without having to introduce a more complicated functional form for $\phi(t)$ that would model overnight effects and daily time scales. As a benchmark, we also simulate a standard Hawkes-based price process ($n_Z \equiv 0$) with $\phi = (1 + 0.01 \times t)^{-1.3}$, $n_H = 0.99$, which is close to the calibration results of [20].

It is important to note that to simulate the ZHawkes and the Hawkes model, we choose constant price jumps $\Delta P_\tau = \pm\psi$. Therefore, our numerical results for the distribution of the volatility can by no means be attributed to the kurtosis of individual price jumps.

For both simulated and real data, we consider the Rogers-Satchell volatility times series for five-minute bins. We use the Hill exponent [24] as an estimator of the empirical tail exponent of the volatility

$$\nu_{\text{hill}} = 1 + \frac{1}{\frac{1}{n} \sum_{i=1}^n \log(\sigma_i / \sigma_{\min})}$$

where σ_{\min} is some cutoff and $\sigma_i \geq \sigma_{\min}$ are the volatilities in the far tail region of the distribution. One obtains $\nu_{\text{hill}} = 4.50$ for the (normalized) five minutes returns of US stocks (in agreement with

many previous determinations of this exponent), $\nu_{\text{hill}} = 5.07$ for the ZHawkes model and $\nu_{\text{hill}} = 12.4$ for the standard Hawkes-based model without ZHawkes feedback. Even with a norm close to one and a slowly-decaying kernel, the standard Hawkes model cannot reproduce the tails observed on US stock data. Instead, the ZHawkes model, with a norm strictly below unity and a short-lived Zumbach effect, naturally produces fat tails very similar to those observed empirically, even with a rather small n_Z . These observations are illustrated by Figures 5 and 6.

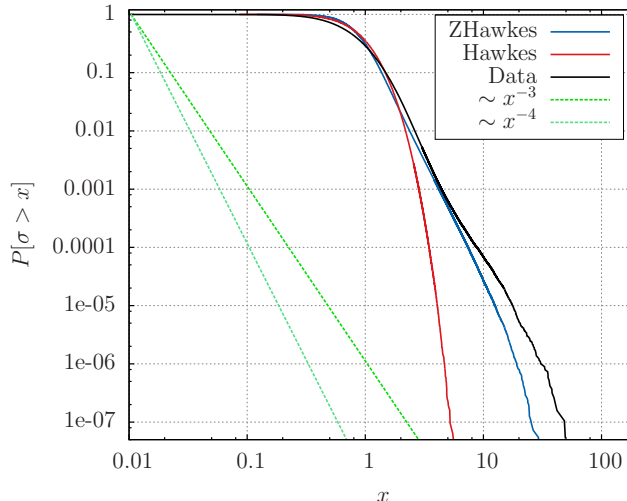


Figure 5: Cumulative density function of the Rogers-Satchell volatility for US stock data (plain line), simulated Hawkes data (red dashed line), and simulated ZHawkes data (blue dot-dashed line). Notice how well the empirical distribution function is reproduced by the ZHawkes model, calibrated as in Section 3.2.3.

5.2 Time-reversal asymmetry of ZHawkes processes

Another salient feature of financial markets is, as discussed in the introduction, the time-reversal asymmetry (TRA) of the price time series. The authors of [14] study this feature for stock data on the one hand, and for a simulated FIGARCH volatility process on the other. The chosen observable is the cross-correlation of present Rogers-Satchell volatilities σ_t^2 with past squared returns $r_{t-\tau}^2$, to that of present squared returns with past volatilities, which is found to be such that $\langle r_{t-\tau}^2 \sigma_t^2 \rangle_t > \langle r_t^2 \sigma_{t-\tau}^2 \rangle_t$ for $\tau > 0$, both for real data and FIGARCH processes.

This observation is one of the main motivations for the model introduced in the present paper, since standard models that use Brownian SDEs are TRS by construction and cannot reproduce this asymmetry. In this section, we measure the amount of TRA for the simulated ZHawkes process and for the Hawkes benchmark described in the previous section, and for the financial dataset studied in Section 3.2.1.

As in Sections 3.2, we consider the returns and the Rogers-Satchell volatilities defined for intra-day five-minute bins. Here, the maximum lag q is fixed to 36 (36 bins of 5 minutes = 3 hours of trading) and the lag index τ varies between 1 and q . We introduce

- The cross-correlation function of the Rogers-Satchell volatility and absolute returns

$$C(\tau) = \frac{\langle \sigma_t^{\text{RS}} \times |r_{t-\tau}| \rangle - \langle \sigma^{\text{RS}} \rangle \langle |r| \rangle}{\sqrt{\langle \sigma^{\text{RS}2} \rangle - \langle \sigma^{\text{RS}} \rangle^2} \sqrt{\langle r^2 \rangle - \langle |r| \rangle^2}}.$$

- The time asymmetry ratio

$$\Delta(\tau) = \frac{\sum_{\tau'=1}^{\tau} [C(\tau') - C(-\tau')]}{2 \sum_{\tau'=1}^q \max(|C(\tau')|, |C(-\tau')|)} \in [-1, 1].$$

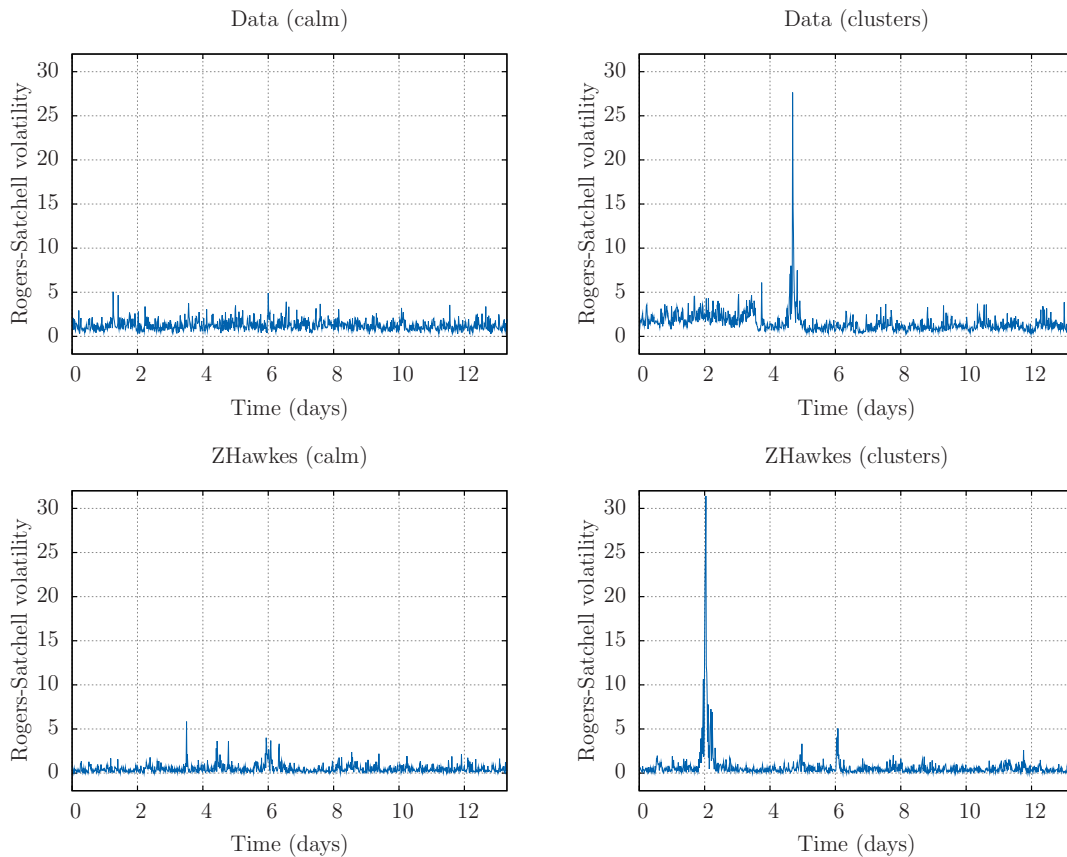


Figure 6: Time series of Rogers-Satchell volatility. Above: real data ; below: simulated ZHawkes data ; left: period of calm ; right: cluster of intense activity.

Note that we choose to compute the cross-correlation function using the *absolute* returns instead of the *squared* returns, since it yields results that are less noisy and more robust to tail events (and thus less sensitive to the normalization method).

We compare the time asymmetry ratios $\Delta(\tau)$ for real stock returns, returns simulated with the ZHawkes model and returns simulated with a standard Hawkes-based price model. The results are illustrated by Figure 7. The standard Hawkes model, perhaps surprisingly, does not generate any detectable TRA: $|\Delta(\tau)| < 10^{-3}$ for all τ . Thus it is clear that the Hawkes model with no off-diagonal quadratic feedback cannot reproduce the time asymmetry observed in intra-day volatility, for which $\Delta(\tau)$ is one hundred times larger. On the other hand, the ZHawkes model with parameters in line with the QARCH calibration of Section 3 features some time asymmetry, which is not only of the correct sign but also reproduces the right order of magnitude, without any further parameter adjustment. However, the function $\tau \mapsto \Delta(\tau)$ is found to be concave for the ZHawkes model (as expected on general grounds) and, strangely, convex for stock data. Even with a thorough normalization protocol, intra-day returns are not rigorously stationary, and we believe that the convexity of $\tau \mapsto \Delta(\tau)$ observed on real data is spurious, as it should saturate to a value less than 1 beyond some time scale. Such convexity would probably be hard to reproduce with a simple model, unless some non-stationary is added by hand.

6 Conclusion

The central message of our study is that the standard Hawkes feedback, where past activity increase the intensity of the current activity, fails at accounting for two essential features of the dynamics of markets: a) the fat-tails in the activity/volatility cannot be reproduced and b) the time-reversal

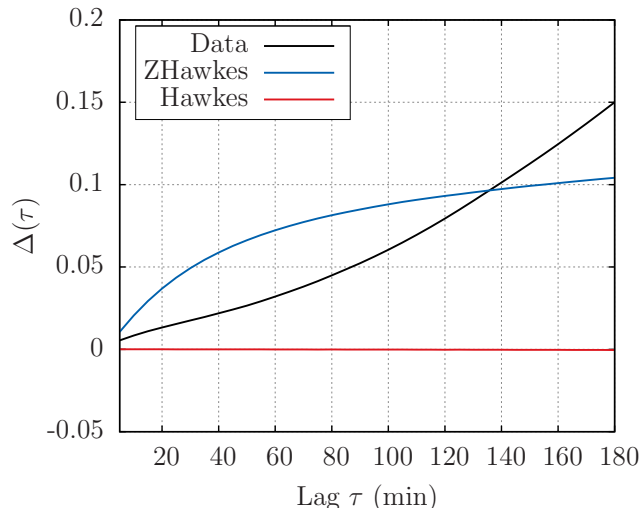


Figure 7: Time asymmetry ratio $\Delta(\tau)$ for US stock data (plain line), simulated Hawkes model (red line), and simulated ZHawkes model (blue dot-dashed line). Note that the Hawkes process does not generate any detectable TRA.

asymmetry between past daily volatilities and future intraday volatilities or vice-versa is completely absent within the Hawkes framework. This was not a priori obvious, since Hawkes processes are constructed on the idea of a feedback from the past. We have thus proposed QHawkes processes as simple, intuitive generalisations of the Hawkes process which posit that the feedback is in fact not only on the past activity, but on past price returns themselves.

A QHawkes model can be seen as a consistent definition of a Quadratic ARCH (QARCH) model as a continuous-time point-processes. This in fact allowed us to calibrate a QHawkes model on the intraday returns of 133 NYSE stocks. We find that the matrix kernel of the QHawkes has a diagonal part (corresponding to the standard Hawkes component) and a off-diagonal, rank-one part that we call “ZHawkes”. It corresponds to Zumbach’s insight that local trends in the price, both up or down, generate more future activity. ZHawkes processes have some interesting properties that standard Hawkes processes lack, namely: (i) the quadratic feedback naturally produces a multiplicative dynamics for the volatility, generating power-law tails for the volatility and the returns, (ii) it can generate long memory without necessarily be at its critical point (iii) it reproduces a level of time-reversal asymmetry (TRA) that is fully compatible with what is measured on actual financial data. The continuous limit SDE corresponding to exponential kernels is found to be a tractable two-dimensional generalization of Pearson diffusions. In particular the tail exponent of the volatility can be exactly computed in several cases and, quite remarkably, fall within the empirical range even when the ZHawkes kernel is of small amplitude. These mathematically tractable diffusions are reminiscent of the log-normal volatility processes considered in [34, 4, 7] and more recently [19], and provide a natural “microscopic” mechanism for a multiplicative process for the volatility itself, which up to now has remained quite a mysterious hypothesis [26].

We hope our paper motivates more developments on the family of QHawkes models. We have indeed only touched upon the mathematical properties and the empirical relevance of such models but we believe that deeper work on the subject would be valuable, in particular concerning the precise calibration of the model itself. A completely open question at this stage is the treatment of overnights and the generalisation of the model to describe longer time scales (our calibration was restricted to intraday data), generalizing the QARCH description proposed by two of us in [9]. In particular, we know that time-reversal asymmetry can still be detected on time scales of days or weeks [37, 14] and this can certainly not be reproduced with a ZHawkes kernel decaying over 30 minutes, as found here. Similarly, multiplicative log-normal models for the volatility have commonly been considered for daily returns. How much is the fat-tailed, long memory of the volatility, recently described within the context of standard Hawkes process, should in fact be traced to the QHawkes mechanism proposed here is, in our opinion, a very interesting question for

future research.

To conclude, we believe that a comprehensive understanding of the volatility process, from the scale of the event up to macroscopic scales, would seem very valuable in several respects, in particular that of market design. One would perhaps understand how a change in market microstructural rules (e.g. the tick size) may affect its macroscopic properties (e.g. volatility). Finding a solid, behavioural microscopic foundations to the volatility process seems crucial: when fully understood, simple constraints on the agents might then change the overall, macroscopic market behaviour. We hope that our generalized Hawkes process could provide some clues on this issue.

Acknowledgments

We want to thank R. Chicheportiche, J. Gatheral, S. Hardiman, Th. Jaisson, I. Mastromatteo and M. Rosenbaum for many insightful discussions on these issues.

References

- [1] R. Allez and J.-P. Bouchaud. Individual and collective stock dynamics: intra-day seasonalities. *New Journal of Physics*, 13(2):025010, 2011.
- [2] E. Bacry, K. Dayri, and J.-F. Muzy. Non-parametric kernel estimation for symmetric hawkes processes. application to high frequency financial data. *The European Physical Journal B*, 85(5):1–12, 2012.
- [3] E. Bacry, S. Delattre, M. Hoffmann, and J.-F. Muzy. Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, 13(1):65–77, 2013.
- [4] E. Bacry, J. Delour, and J. Muzy. Modelling financial time series using multifractal random walks. *Physica A: Statistical Mechanics and its Applications*, 299(1):84–92, 2001.
- [5] E. Bacry, I. Mastromatteo, and J.-F. Muzy. Hawkes processes in finance. *arXiv preprint arXiv:1502.04592*, 2015.
- [6] E. Bacry and J.-F. Muzy. Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance*, 14(7):1147–1166, 2014.
- [7] L. Bergomi. Smile dynamics ii. *Available at SSRN 1493302*, 2005.
- [8] P. Blanc. Modélisation de la volatilité des marchés financiers par une structure arch multifréquence. *Masters thesis, Université de Paris VI Pierre et Marie Curie*.
- [9] P. Blanc, R. Chicheportiche, and J.-P. Bouchaud. The fine structure of volatility feedback ii: overnight and intra-day effects. *Physica A: Statistical Mechanics and its Applications*, 402:58–75, 2014.
- [10] P. Brémaud, L. Massoulié, et al. Hawkes branching point processes without ancestors. *Journal of applied probability*, 38(1):122–135, 2001.
- [11] J. Buescu. Positive integral operators in unbounded domains. *Journal of Mathematical Analysis and Applications*, 296(1):244–255, 2004.
- [12] D. Challet, M. Marsili, Y.-C. Zhang, et al. Minority games: interacting agents in financial markets. *OUP Catalogue*, 2013.
- [13] S. Chatterji. *Cours d’analyse Tome 3: Équations différentielles ordinaires et aux dérivées partielles*. Number vol. 1 in Cours d’analyse. Presses polytechniques et universitaires romandes, 1998.
- [14] R. Chicheportiche and J.-P. Bouchaud. The fine-structure of volatility feedback i: Multi-scale self-reflexivity. *Physica A: Statistical Mechanics and its Applications*, 410:174–195, 2014.

- [15] J. C. Cox, J. E. Ingersoll Jr, and S. A. Ross. An intertemporal general equilibrium model of asset prices. *Econometrica: Journal of the Econometric Society*, pages 363–384, 1985.
- [16] M. Cristelli, L. Pietronero, and A. Zaccaria. Critical overview of agent-based models for economics. *Proceedings of the International School of Physics “Enrico Fermi” Course CLXXVI, Complex Materials in Physics and Biology*, edited by F. Mallamace and H.E. Stanley, 2011.
- [17] V. Filimonov and D. Sornette. Apparent criticality and calibration issues in the hawkes self-excited point process model: application to high-frequency financial data. *Quantitative Finance*, (ahead-of-print):1–22, 2015.
- [18] J. L. Forman and M. Sørensen. The pearson diffusions: A class of statistically tractable diffusion processes. *Scandinavian Journal of Statistics*, 35(3):438–465, 2008.
- [19] J. Gatheral, T. Jaisson, and M. Rosenbaum. Volatility is rough. *Available at SSRN 2509457*, 2014.
- [20] S. J. Hardiman, N. Bercot, and J.-P. Bouchaud. Critical reflexivity in financial markets: a hawkes process analysis. *The European Physical Journal B*, 86(10):1–9, 2013.
- [21] S. J. Hardiman and J.-P. Bouchaud. Branching-ratio approximation for the self-exciting hawkes process. *Physical Review E*, 90(6):062807, 2014.
- [22] A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.
- [23] S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, 6(2):327–343, 1993.
- [24] B. M. Hill et al. A simple general approach to inference about the tail of a distribution. *The annals of statistics*, 3(5):1163–1174, 1975.
- [25] T. Jaisson and M. Rosenbaum. Limit theorems for nearly unstable hawkes processes. *arXiv preprint arXiv:1310.2033*, 2013.
- [26] T. Jaisson and M. Rosenbaum. Rough fractional diffusions as scaling limit of nearly unstable heavy tailed hawkes processes. *to appear*, 2015.
- [27] O. Kallenberg. *Foundations of modern probability*. Springer Science & Business Media, 2002.
- [28] P. J. Laub, T. Taimre, and P. K. Pollett. Hawkes processes. *arXiv preprint arXiv:1507.02822*, 2015.
- [29] Y. Pomeau. Symétrie des fluctuations dans le renversement du temps. *Journal de Physique*, 43(6):859–867, 1982.
- [30] J. B. Ramsey and P. Rothman. *Characterization of the time irreversibility of economic time series: Estimators and test statistics*. CV Starr Center for Applied Economics, New York University, Faculty of Arts and Science, Department of Economics, 1988.
- [31] J. B. Ramsey and P. Rothman. Time irreversibility and business cycle asymmetry. *Journal of Money, Credit and Banking*, pages 1–21, 1996.
- [32] A. Saichev and D. Sornette. Generation-by-generation dissection of the response function in long memory epidemic processes. *The European Physical Journal B*, 75(3):343–355, 2010.
- [33] E. Sentana. Quadratic arch models. *The Review of Economic Studies*, 62(4):639–661, 1995.
- [34] E. M. Stein and J. C. Stein. Stock price distributions with stochastic volatility: an analytic approach. *Review of financial Studies*, 4(4):727–752, 1991.
- [35] A. C. Zaanen. Linear analysis. 1956.
- [36] G. Zumbach. Time reversal invariance in finance. *Quantitative Finance*, 9(5):505–515, 2009.
- [37] G. Zumbach and P. Lynch. Heterogeneous volatility cascade in financial markets. *Physica A: Statistical Mechanics and its Applications*, 298(3-4):521–529, 2001.

A Exact equations relating the kernel and the auto-correlation functions

To simplify notations, we write (in this appendix only) $\varphi(t) = K(t, t)$.

For $s < t$, one has $\mathcal{C}(t-s) = \lambda_\infty \bar{\lambda} - \bar{\lambda}^2 + \mathbb{E} \left[A_t \frac{dN_s}{ds} \right] + 2\mathbb{E} \left[M_t \frac{dN_s}{ds} \right]$.

$$\mathbb{E} \left[A_t \frac{dN_s}{ds} \right] = \int_{-\infty}^t \varphi(t-u) \mathbb{E} \left[\frac{dN_u}{du} \frac{dN_s}{ds} \right] du.$$

For $u \neq s$, $\mathbb{E} \left[\frac{dN_u}{du} \frac{dN_s}{ds} \right] du = [\mathcal{C}(u-s) + \bar{\lambda}^2] du$, and for $u = s$, $\mathbb{E} \left[\left(\frac{dN_u}{du} \right)^2 \right] du = \kappa \mathbb{E} \left[\frac{dN_u}{(du)^2} \right] du = \kappa \bar{\lambda}$, where κ is the kurtosis of the law μ of the jumps of P ($\kappa = 1$ if $\Delta P_\tau = \pm \psi$). Thus,

$$\mathbb{E} \left[A_t \frac{dN_s}{ds} \right] = \text{Tr}(K) \bar{\lambda}^2 + \kappa \bar{\lambda} \varphi(t-s) + \int_{-\infty}^t \varphi(t-u) \mathcal{C}(u-s) du.$$

On the other hand,

$$\begin{aligned} \mathbb{E} \left[M_t \frac{dN_s}{ds} \right] &= \frac{1}{\psi^2} \int_{-\infty}^t \mathbb{E} \left[\Theta_{t,u} \frac{dP_u}{du} \frac{dN_s}{ds} \right] du \\ &= \frac{1}{\psi^2} \int_{-\infty}^t \int_{-\infty}^{u-} K(t-u, t-r) \mathbb{E} \left[\frac{dN_s}{ds} \frac{dP_u}{du} \frac{dP_r}{dr} \right] dr du \\ &= \int_{-\infty}^{s-} \int_{-\infty}^{u-} K(t-u, t-r) \mathcal{D}(s-u, s-r) dr du, \end{aligned}$$

since ΔP_r and $(\Delta P_r)^3$ are centered, which implies that $\mathbb{E} \left[\frac{dN_s}{ds} \frac{dP_u}{du} \frac{dP_r}{dr} \right] = 0$ for $u \geq s$. Taking $t = \tau > 0$ and $s = 0$, we obtain

$$\mathcal{C}(\tau) = \kappa \bar{\lambda} \varphi(\tau) + \int_{-\infty}^{\tau} \varphi(\tau-u) \mathcal{C}(u) du + 2 \int_{0+}^{\infty} \int_{u+}^{\infty} K(\tau+u, \tau+r) \mathcal{D}(u, r) dr du.$$

For $t > t_1 > t_2$, one has $\mathcal{D}(t-t_1, t-t_2) = \frac{1}{\psi^2} \mathbb{E} \left[A_t \frac{dP_{t_1}}{dt_1} \frac{dP_{t_2}}{dt_2} \right] + \frac{2}{\psi^2} \mathbb{E} \left[M_t \frac{dP_{t_1}}{dt_1} \frac{dP_{t_2}}{dt_2} \right]$. The first term gives

$$\begin{aligned} \frac{1}{\psi^2} \mathbb{E} \left[A_t \frac{dP_{t_1}}{dt_1} \frac{dP_{t_2}}{dt_2} \right] &= \frac{1}{\psi^2} \int_{-\infty}^t \varphi(t-u) \mathbb{E} \left[\frac{dN_u}{du} \frac{dP_{t_1}}{dt_1} \frac{dP_{t_2}}{dt_2} \right] du \\ &= \int_{t_1+}^t \varphi(t-u) \mathcal{D}(u-t_1, u-t_2) du. \end{aligned}$$

The second term is given by

$$\frac{1}{\psi^2} \mathbb{E} \left[M_t \frac{dP_{t_1}}{dt_1} \frac{dP_{t_2}}{dt_2} \right] = \frac{1}{\psi^4} \int_{-\infty}^t \int_{-\infty}^{u-} K(t-u, t-r) \mathbb{E} \left[\frac{dP_{t_1}}{dt_1} \frac{dP_{t_2}}{dt_2} \frac{dP_u}{du} \frac{dP_r}{dr} \right] dr du$$

Since $r < u$ in the integral and $t_2 < t_1$, the expected value is zero if $u \neq t_1$. For $u = t_1$, we have $\mathbb{E} \left[\left(\frac{dP_u}{du} \right)^2 \frac{dP_{t_2}}{dt_2} \frac{dP_r}{dr} \right] du = \psi^2 \mathbb{E} \left[\frac{dN_u}{(du)^2} \frac{dP_{t_2}}{dt_2} \frac{dP_r}{dr} \right] du = \psi^2 \mathbb{E} \left[\frac{dN_{t_1}}{dt_1} \frac{dP_{t_2}}{dt_2} \frac{dP_r}{dr} \right]$. Thus,

$$\mathbb{E} \left[M_t \frac{dP_{t_1}}{dt_1} \frac{dP_{t_2}}{dt_2} \right] = \frac{1}{\psi^2} \int_{-\infty}^{t_1-} K(t-t_1, t-r) \mathbb{E} \left[\frac{dN_{t_1}}{dt_1} \frac{dP_{t_2}}{dt_2} \frac{dP_r}{dr} \right] dr.$$

For $r \neq t_2$, one has $\frac{1}{\psi^2} \mathbb{E} \left[\frac{dN_{t_1}}{dt_1} \frac{dP_{t_2}}{dt_2} \frac{dP_r}{dr} \right] dr = \mathcal{D}(t_1-t_2, t_1-r) dr$. On the other hand $r = t_2$ yields $\mathbb{E} \left[\frac{dN_{t_1}}{dt_1} \frac{dN_{t_2}}{(dr)^2} \right] dr = \mathbb{E} \left[\frac{dN_{t_1}}{dt_1} \frac{dN_{t_2}}{dt_2} \right] = \mathcal{C}(t_1-t_2) + \bar{\lambda}^2$. We obtain

$$\mathbb{E} \left[M_t \frac{dP_{t_1}}{dt_1} \frac{dP_{t_2}}{dt_2} \right] = K(t-t_1, t-t_2) [\mathcal{C}(t_1-t_2) + \bar{\lambda}^2] + \int_{-\infty}^{t_1-} K(t-t_1, t-r) \mathcal{D}(t_1-t_2, t_1-r) dr.$$

We eventually obtain by taking $\tau_2 = t > \tau_1 = t - t_1, t_2 = 0$,

$$\begin{aligned} \mathcal{D}(\tau_1, \tau_2) &= 2K(\tau_1, \tau_2)[\mathcal{C}(\tau_2 - \tau_1) + \bar{\lambda}^2] + \int_{(\tau_2 - \tau_1)^+}^{\tau_2} \varphi(\tau_2 - u) \mathcal{D}(u - \tau_2 + \tau_1, u) du \\ &\quad + 2 \int_{-\infty}^{(\tau_2 - \tau_1)^-} K(\tau_1, \tau_2 - u) \mathcal{D}(\tau_2 - \tau_1, \tau_2 - \tau_1 - u) du. \end{aligned}$$

B Asymptotic analysis of the Hawkes + ZHawkes process

In order to analyze the coupled Hawkes + ZHawkes processes, we first write the Fokker-Planck equation for the joint probability $\Pi(h, y)$ of $h \equiv \bar{H}$ and $y \equiv \bar{Y} = \bar{Z}^2$. Setting $t \leftarrow \beta t$ as the new time, we find:

$$\begin{aligned} \frac{\partial \Pi}{\partial t} &= -\frac{\partial}{\partial h} \{[-(1 - n_H)h + n_H(\lambda_\infty + y)] \Pi\} \\ &\quad - \chi \frac{\partial}{\partial y} \{[(n_Z - 1)y + n_Z(\lambda_\infty + h)] \Pi\} + 2\chi n_Z \frac{\partial^2}{\partial y^2} \{[y(\lambda_\infty + h + y)] \Pi\} \end{aligned} \quad (32)$$

We will study the stationary distribution of the process, such that the left-hand side of the above equation is zero. We introduce the conditional distribution of h for a given y , $\Pi(h|y)$, and the marginal distribution of y , $\pi(y)$, as:

$$\pi(y) := \int_0^\infty dh \Pi(h, y); \quad \Pi(h|y) = \frac{\Pi(h, y)}{\pi(y)}, \quad (33)$$

and the generating function of $\Pi(h|y)$, as:

$$Z(z|y) = \int_0^\infty dh e^{-zh} \Pi(h|y), \quad (34)$$

such that $Z(0|y) = 1$ and $Z'(0|y) := -a^*$ is the conditional average of h for a given y .

Now we assume, and self consistently check, that for large y , $\Pi(h|y)$ is of the form $1/y F(h/y)$, which means that h is a random variable of order y . This implies:

$$Z(z|y) = G(x = zy); \quad G(x) := \int_0^\infty du e^{-zu} F(u). \quad (35)$$

Multiplying Eq.(32) by e^{-zh} and integrating over h then leads, in the stationary state, to:

$$\begin{aligned} -x\pi(y) [(1 - n_H)G'(x) + n_H G(x)] - \chi \frac{\partial}{\partial y} \{[(n_Z - 1)G(x) - n_Z G'(x)] y \pi(y)\} \\ + 2\chi n_Z \frac{\partial^2}{\partial y^2} \{[G(x) - G'(x)] y^2 \pi(y)\} = 0, \end{aligned} \quad (36)$$

where we have assumed $y \gg \lambda_\infty$. In the asymptotic limit, $\pi(y)$ behaves as a power law: $\pi(y) \propto A/y^{1+\mu}$. Indeed, injecting this ansatz into the last equation leads to a non-trivial equation for $G(x)$ where y and A have completely disappeared:

$$\begin{aligned} x[(1 - n_H)G'(x) + n_H G(x)] &= \chi [\mu n_Z H(x) - n_Z x H'(x) - \mu G(x) + x G'(x)] \\ &\quad + 2\chi n_Z [x^2 H''(x) + 2(1 - \mu)x H'(x) - \mu(1 - \mu)H(x)], \end{aligned} \quad (37)$$

where we have introduced the shorthand $H(x) = G(x) - G'(x)$. Let us first analyze this equation for $x = 0$; without any further assumptions one has, with $G(0) = 1$ and $G'(0) = -a^*$:

$$\mu n_Z (1 + a^*) - \mu - 2n_Z \mu (1 - \mu) (1 + a^*) = 0 \Rightarrow \mu = \frac{1}{2} + \frac{1}{2n_Z(1 + a^*)}, \quad (38)$$

where the unphysical solution $\mu = 0$ was discarded. We thus need to solve Eq. (37) for $G(x)$ and determine a^* from the value of $-G'(0)$. An easy case is $\chi = 0$. One immediately finds:

$$(1 - n_H)G'(x) + n_H G(x) = 0 \Rightarrow G_0(x) = e^{-n_H x / (1 - n_H)}, \quad (39)$$

leading to $a_0^* = n_H / (1 - n_H)$. The small χ expansion is also conveniently performed by setting $G(x) = G_0(x) + \chi g_1(x) + \chi^2 g_2(x) + \dots$. To first order in χ , the equation for g_1 reads:

$$(1 - n_H)g_1'(x) + n_H g_1(x) = G_0(x) \left[\frac{n_H(1 - n_H - n_Z)}{(1 - n_H)^2} + \frac{n_H^2}{(1 - n_H)^2} x \right], \quad (40)$$

and thus, with the right boundary condition for $g_1(x)$,

$$g_1(x) = \left[\frac{n_H(1 - n_H - n_Z)}{(1 - n_H)^3} x + \frac{n_H^2}{2(1 - n_H)^3} x^2 \right] e^{-a_0^* x}. \quad (41)$$

To first order, one thus finds:

$$a^* = \frac{n_H}{(1 - n_H)} \left[1 - \chi \frac{(1 - n_H - n_Z)}{(1 - n_H)^2} + O(\chi^2) \right]. \quad (42)$$

In the opposite limit $\chi \rightarrow \infty$, one finds that $G(x) = 1$ solves the equation, as expected since in this limit h cannot follow the dynamics of y , and therefore one expects that in the limit $y \rightarrow \infty$, $F(u) \approx \delta(u)$ and thus $G(x) = 1$. When χ is large but not infinite, one can expect that $F(u)$ has a width of order χ^{-1} , and thus that $G(x)$ is a function of x/χ . This means that each derivative of G brings an extra factor χ^{-1} . Setting $a^* = a/\chi$ and matching the terms in Eq. (37), we find:

$$\chi(\mu + x(1 - n_Z))G'(x) + (a\mu + n_H x)G(x), \quad (43)$$

or:

$$\ln G(x) = -\frac{1}{\chi(1 - n_Z)} [n_H x + \mu(a - n_H/(1 - n_Z)) \ln(\mu + (1 - n_Z)x)], \quad (44)$$

which shows that our assumption that $G(x)$ is a function of x/χ singles out $a = n_H/(1 - n_Z)$ as the only possibility, in which case:

$$G(x) =_{\chi \rightarrow \infty} e^{-\frac{n_H x}{\chi(1 - n_Z)}}. \quad (45)$$

This means that in this limit, $\Pi(h|y) \approx \delta(h - \frac{n_H}{\chi(1 - n_Z)} y)$.

Finally, let us consider the limit $n_Z \rightarrow 0$ for a finite χ . The idea now is to postulate that for small n_Z , $\Pi(h|y)$ is strongly peaked around $a^* y$, with a width that goes to zero as $\sqrt{n_Z}$. This translates into the following ansatz for $G(x)$

$$G(x) = e^{-a^* x} \mathcal{G}(\sqrt{n_Z} x). \quad (46)$$

We can now analyze Eq. (37) in the regime $n_Z \rightarrow 0$ with fixed $z = \sqrt{n_Z} x$. The leading order terms are of order $1/n_Z$, and lead to an equation that is identically satisfied. The next two orders, $O(1/\sqrt{n_Z})$ and $O(1)$ allow us to fix both the function $\mathcal{G}(z)$ and the value of a^* . We find in particular:

$$\mathcal{G}(z) = \exp \left[\frac{(1 + a^*)^2 [(1 - n_H + \chi)a^* - n_H] z^2}{\chi} \right], \quad (47)$$

which shows that the distribution $\Pi(h|y)$ is in fact gaussian in that limit. We also find that a^* obeys the following equation:

$$(\gamma a^* - n_H)(\gamma + (\gamma + 2\chi)a^*) = a^{*2} \chi^2 \quad \gamma = 1 - n_H + \chi. \quad (48)$$

The solution takes a simple form in the limits $\chi \rightarrow 0$ and $\chi \rightarrow \infty$, where we recover the results obtained above.

In the general case, Eq. (37) is a third order, linear ODE for $G(x)$; imposing the correct boundary condition $G(x \rightarrow \infty)$ selects special values of a^* for any triplet (n_H, n_Z, χ) . The largest admissible value of a^* corresponding to the smallest value of the tail exponent will be the physical solution. Unfortunately, we have not been able to make progress yet on this general case.