Hawkes processes in finance

Emmanuel Bacry[1], Iacopo Mastromatteo[1], and Jean-François Muzy[1,2]

[1]Centre de Mathématiques Appliquées, CNRS, École Polytechnique,
UMR 7641, 91128 Palaiseau, France
[2]Laboratoire Sciences Pour l'Environnement, CNRS, Université de Corse,
UMR 6134, 20250 Corté, France

Abstract

In this paper we propose an overview of the recent academic literature devoted to the applications of Hawkes processes in finance. Hawkes processes constitute a particular class of multivariate point processes that has become very popular in empirical high frequency finance this last decade. After a reminder of the main definitions and properties that characterize Hawkes processes, we review their main empirical applications to address many different problems in high frequency finance. Because of their great flexibility and versatility, we show that they have been successfully involved in issues as diverse as estimating the volatility at the level of transaction data, estimating the market stability, accounting for systemic risk contagion, devising optimal execution strategies or capturing the dynamics of the full order book.

1 Introduction

The availability of high frequency financial data during the last decades allowed the empirical finance to devise and calibrate models of market microstructure, aiming at accounting for the intraday market dynamics in its finest details. Until recently, there were only few continuous time models for the high frequency price variations. Most approaches relied on discrete time models that consist either in aggregating the dynamics on intervals of a regular time grid or in considering the succession of discrete time events like trades (these models are generally referred to as "trading time" or "business time" models). Hasbrouck [36], Engle and Russel [28] in the nineties, were the first to advocate that the modeling of financial data at the transaction level could be advantageously done within the framework of continuous time "point processes". Since then, point processes applications to finance is an ongoing, very active topic in the econometric literature. We refer the reader to the recent review of Bowens and Hautsch [9] on this subject.

The first type of point processes proposed in the context of market microstructure is the ACD model introduced by Engle and Russel [28]. This model and its variants remains, by far, the most used model in high frequency econometrics [9]. In this class of models, the process is defined by the means of its "hazard function" that specifies

the conditional law of inter-event (or duration) intervals. However, point processes (or counting processes) can alternatively be represented by their "intensity function" that represents the condionnal probability density of the occurence of an event in the immediate future (see e.g. [22] for a comprehensive textbook on point process mathematical properties).

In a pioneering work, Bowsher [15], recognized the flexibility and the advantages of using the class of multivariate counting processes that can be specified by a conditional intensity vector. More specifically, he introduced a bivariate Hawkes processes in order to model the joint dynamics of trades and mid-price changes of the NYSE. Hawkes processes is a class of multivariate point processes that were introduced in the seventies by A.G Hawkes [37, 38] notably to model the occurrence seismic events. They involve an intensity vector that is a simple linear function of past events (see e.g. [10, 9] for other examples of dynamical intensity point processes). Hawkes models are becoming more and more popular in the domain of high frequency finance. This popularity can be explained above all by their great simplicity and flexibility, as anticipated by Bowsher [15]. These models can easily account for the interaction of various types of events, for the influence of some intensive factors (through marks) or for the existence of non-stationarities. They are amenable to statistical inference and closed-form formulae can be obtained in some particular situations. Moreover since their parameters have a straightforward interpretation (notably through the cluster representation), they lead to a quite simple interpretation of many aspects of the complex dynamics of modern electronic markets.

In this paper, we propose a survey of recent academic studies using Hawkes processes in the context of finance. As we already explained, there are many such studies. In order to present them in relation one to each other, we had to group them by "themes". Of course, the boundaries of these themes are unclear (e.g., some of the studies belong to several themes), so the choices we have made are unavoidably somewhat arbitrary. However we believe it helps capturing the "picture" of Hawkes models in finance.

This survey is organized as follows. Sec. 2 is devoted to the theory of Hawkes processes. It introduces the main definitions and the general properties that will be used all along the paper. The following Sections focus on the applications of Hawkes processes to finance. Sec. 3 starts with the main univariate models that can be found in the literature. That includes market activity or risk models (e.g., 1-dimensional market order flow models, extreme return models). Price models (mid-price or best limit price) are presented in Sec. 4 whereas Sec. 5 is devoted to impact models. In this Section, we do not only discuss the influence of market order flows on price moves but also the problems related to optimal execution. Models that involve more order flows are presented in Sec. 6. So-called level-I models (i.e., dealing "only" with the dynamics of the best limits) as well full order book models are discussed. Finally, various studies that did not clearly fit in any of the previous Sections are presented in Sec. 7 (e.g. systemic risk models, high-dimensional models or news models). More materials can be found in Appendices. In Appendix A, all the academic works that are discussed throughout our paper and which involves numerical experiments on financial data is listed in a single table. This table summarizes some essential characteristics of the models and data used in each work. Finally, two Appendices sum up the main results

2

about simulation (App. B) and estimation (App. C) of Hawkes processes.

2   The Hawkes process

As mentioned in the introduction, Hawkes processes are a class of multivariate point processes introduced by Hawkes in the early seventies [37, 38] that are characterized by a stochastic intensity vector. If N_t is a vector of counting processes[1] at time t, then its intensity vector λ_t is defined heuristically as (see e.g. [22] for a rigorous definition):

$$\lambda_t = \lim_{\Delta \to 0} \Delta^{-1} \mathbb{E} \left[N_{t+\Delta} - N_t \, | \mathcal{F}_t \right] \tag{1}$$

where the filtration $\mathcal{F}_t$ stands for the information available up to (but not including) time t. In the case of Hawkes processes, λ_t is simply a linear function of past jumps of the process as specified thereafter.

2.1   Definition

We consider a D-variate counting-process $N_t = \{N_t^i\}_{i=1}^D$, whose associated intensity vector is denoted as $\lambda_t = \{\lambda_t^i\}_{i=1}^D$.

Definition 1 (Hawkes process). *A Hawkes process is a counting-process N_t such that the intensity vector can be written as*

$$\lambda_t^i = \mu^i + \sum_{i=1}^D \int \mathrm{d}N_{t'}^j \phi^{ij}(t - t'), \tag{2}$$

where the quantity $\mu = \{\mu^i\}_{i=1}^D$ is a vector of exogenous intensities, and $\Phi(t) = \{\phi^{ij}(t)\}_{i,j=1}^D$ is a matrix-valued kernel such that:
- *It is component-wise positive, i.e., $\phi^{ij}(t) \geq 0$ for each $1 \leq i, j \leq D$;*
- *It is component-wise causal (if $t < 0$, $\phi^{ij}(t) = 0$ for each $1 \leq i, j \leq D$);*
- *Each component $\phi^{ij}(t)$ belongs to the space of L^1-integrable functions[2]*

Notation 1 (Convolution). *We can adopt a more compact notation so to rewrite Eq. (2) as*

$$\lambda_t = \mu + \Phi * \mathrm{d}N_t, \tag{3}$$

by defining the $*$ operation, corresponding to a matrix multiplication in which ordinary products are replaced by convolutions.

Notation 2 (Event times). *By introducing the couples $\{(t_m, k_m)\}_{m=1}^M$, where t_m denotes the time of event number m and $k_m \in [1, ..., D]$ indicates its component, Eq. (2) can also be rewritten as*

$$\lambda_t^i = \mu^i + \sum_{m=1}^M \phi^{i,k_m}(t - t_m), \tag{4}$$

Fig. 1 shows as an example a specific realization of a multivariate Hawkes process. Even though the process defined by Eq. (2) is well-defined by any choice of kernel $\Phi(t)$

---

[1] A counting process is a stochastic process $\{N_t\}_{t \geq 0}$, with values that are positive, integer, and increasing. By convention $N_0 = 0$.

[2] Strictly speaking, this definition only calls for $\phi^{ij}(t) \in L_{loc}^1$ but this is of no interest for this paper.
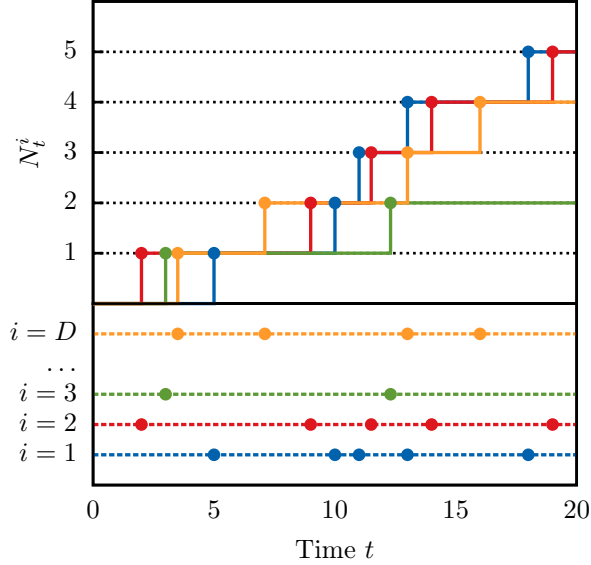
Figure 1: A realization of a multivariate Hawkes process. The dots represent individual events, while the different rows refer to different i coordinates.

satisfying the three conditions stated above, the stationary case characterized below is of particular relevance in most of the applications of Hawkes processes to finance (see Sec. 2.2 for some notable exceptions).

Proposition 1 (Stationarity)**.** *The process N_t has asymptotically stationary increments and λ_t is asymptotically stationary if the kernel satisfies the assumption, also refered to as the* **stability condition***:*
(H) *The matrix $||\Phi|| = \{||\phi^{ij}||\}_{i,j=1}^{D}$ has spectral radius smaller than 1.*

Notation 3 (Spectral radius)**.** *Here and in the following parts of the discussion, given a a scalar function $f(t)$ we denote with the symbol $||f||$ its L^1-norm, defined as $\int \mathrm{d}t\, |f(t)|$. For a matrix $F = \{f^{ij}\}$, the notation $||F||$ will denote its spectral radius. Finally, for a matrix of functions $F(t) = \{f^{ij}(t)\}_{i,j=1}^{D}$, we will write $||F|| = \{||f^{ij}||\}_{i,j=1}^{D}$*

From now on, we will always consider (unless specified) that assumption **(H)** holds and that the Hawkes processes are in the asymptotically stationary regime. In particular, the averages taken in the stationary state will be denoted with $\mathbb{E}[\dots]$, while the variances will be written as $\mathbb{V}[\dots]$. The consequences of the stationarity assumption **(H)** will be fully explored in the next subsection. Here, we will first provide a simple

implementation of a Hawkes process, whose prototypical version is the one in which the kernel functions $\phi^{ij}(t)$ are exponential functions.

Example 1 (Exponential kernel). *Consider a bivariate Hawkes process in which the kernel matrix has the form*[3]

$$\Phi(t) = \begin{pmatrix} \phi^{(s)}(t) & \phi^{(c)}(t) \\ \phi^{(c)}(t) & \phi^{(s)}(t) \end{pmatrix} \tag{5}$$

where the kernel components have the exponential form

$$\phi^{(s/c)}(t) = \alpha^{(s/c)}\beta^{(s/c)}e^{-\beta^{(s/c)}t}\mathbb{1}_{t\in\mathbb{R}^+}\,, \tag{6}$$

where $\mathbb{1}_x$ is an indicator function equal to 1 if x is true and zero otherwise. The above functions are L^1-integrable, so that the choice $\alpha^{(s/c)}, \beta^{(s/c)} > 0$ ensures that the associated Hawkes process is well-defined. For the process to be stable, one needs to additionally require that the spectral norm satisfies $||\Phi|| = ||\phi^{(s)}|| + ||\phi^{(c)}|| < 1$. Due to $\alpha^{(s/c)} = \int_0^\infty \phi^{(s/c)}(t)$, the stability condition becomes

$$\alpha^{(s)} + \alpha^{(c)} < 1\,. \tag{7}$$

The $\alpha^{(s/c)}$ parameters can be interpreted as the ones setting the overall strength of the interactions, while the $\beta^{(s/c)}$ control the relaxation time of the perturbations induced from past to future events.

The Hawkes process with exponential kernels has several advantageous properties, as it allows one to compute the expected value of arbitrary functions of N_t (see Sec. 2.3.4 and Ref. [29]), to be directly simulated (see App. B), or to compute efficiently its likelihood (see App. C.1). Most of these properties descend from a Markov property which in its simplest form is stated as follows:

Proposition 2 (Markov property for exponential kernels). *Consider a Hawkes process with exponential kernels $\phi^{ij}(t) = \alpha^{ij}\beta e^{\beta t}\mathbb{1}_{t\in\mathbb{R}^+}$. Then the couple (N_t, λ_t) is a Markov process. In particular, Eq. (2) for the intensity λ_t can be recast in Markovian form as*

$$\mathrm{d}\lambda_t = -\beta\lambda_t\,\mathrm{d}t + \alpha\beta\,\mathrm{d}N_t \tag{8}$$

This property can be extended to the case in which *(i)* the coefficients β^{ij} are non-constant across components and *(ii)* the kernel Φ contains a finite sum of exponentials. The price to pay in this more general setting is the introduction of an extra set of A auxiliary processes $\{\tilde{\lambda}_t^{(a)}\}_{a=1}^A$, suitably chosen so that the resulting $(A+1)$-uple $(N_t, \tilde{\lambda}_t^{(1)}, \ldots, \tilde{\lambda}_t^{(A)})$ is Markovian.

In the non-exponential case, the Hawkes process cannot be generally mapped to a Markovian process, implying that it is necessary to take track of all its past history in order to perform exact simulation and estimation. A particularly well-known example of a non-exponential kernel is the power-law one, proposed in Ref. [59] in order to describe temporal clusters of seismic activity.

---

[3]The upperscript (s) (resp. (c)) stands for the word *self* (resp. *cross*), since it describes the self- (resp. cross-) excitation of the two components.

Example 2 (Power-law kernel). *Let's now consider the case $D = 1$, and assume the kernel to be parameterized by the regularized power-law*

$$\Phi(t) = \frac{\alpha\beta}{(1 + \beta t)^{1+\gamma}} \mathbb{1}_{t \in \mathbb{R}^+} \tag{9}$$

Also in this case the process is well-defined for $\alpha, \beta > 0$. The stationarity condition in is met for

$$\alpha < \gamma\,, \tag{10}$$

indicating that the tail exponent γ of a power law-kernel should be positive for the increments of the process to be stationary.

2.2 Some extensions

Although the model defined in above section is the one originally introduced in [37, 38], and most widely used in the literature, several generalizations have been proposed since.

2.2.1 Marked Hawkes processes

Eq. (4) defining the Hawkes process can be enriched by endowing each event with a *mark* variable, thus obtaining a sequence of event times, components and marks $\{(t_m, k_m, \xi_m)\}_{m=1}^M$. One may further assumes events labeled with different marks to have different effects on the future intensities, leading to a dynamics for λ_t of the type

$$\lambda_t^i = \mu^i + \sum_{m=1}^M \phi^{i,k_m}(t - t_m, \xi_m)\,, \tag{11}$$

Finally, one needs to introduce a generating mechanism for the marks, which are typically assumed to be i.i.d. random variables drawn with each event and sampled from a common distribution $p(\xi)$. A typical choice for the interaction kernel is the one $\phi^{ij}(t, \xi) = \phi^{ij}(t)\chi^{ij}(\xi)$, in which one assumes a factorized form for the effect of the marks. This mechanism is used to describe events of different weights, and has been originally employed in order to model the occurrence of earthquakes of difference magnitudes [59]. In finance, marks can be used in order to model trades performed at times t_m with different volumes ξ_m (as e.g. in [30, 5], see Sec. 4 and 5) or a drawdown intensity (as e.g. in [27, 18], see Sec. 3.1). On a more general ground, note that multivariate Hawkes processes can also be seen as an example of Hawkes processes with interacting marks [51].

2.2.2 Exogenous non-stationarity

The exogenous intensity μ can be generalized to a deterministic function of time $\mu(t)$. This choice allows to model a non-stationary system in which the interaction kernel is indeed independent of time. In finance, this is the case when one wants to model intraday seasonalities (see the review article [9]) and/or spillover effects within successive days (as in [15]).

2.2.3 Endogenous non-stationarity

The scenarios $||\Phi|| > 1$ and $||\Phi|| = 1$ have also been considered in order to model non-stationarity induced by endogenous interactions. These two cases present indeed an important difference: while in the former one the average intensity grows exponentially in time, in the latter one the process may possess a finite average event rate. This second type of non-stationarity, which we will call *quasi-stationarity*, has raised a strong interest in the literature due to the fact that the condition $||\Phi|| \approx 1$ is often met when calibrating Hawkes processes to real financial data (see the detailed discussion in Sec. 3). The limiting behavior of a Hawkes process in the regime $||\Phi|| = 1$ has been analyzed by Brémaud and Massoulié in [17], where it is shown in particular that:

Proposition 3 (Degeneracy of critical, short-range Hawkes). *Let N_t be a univariate Hawkes process as in Eq. (2) such that $||\Phi|| = 1$ and $\mu = 0$. Then if*

$$\int_0^\infty \mathrm{d}t\, t\, \Phi(t) < \infty \tag{12}$$

the average of the conditional intensity is either 0 or $+\infty$.

Hence in the $D = 1$ case, short-ranged kernels always lead to trivial processes. The next result shows instead that interactions of broader range allow instead a richer behavior.

Theorem 1 (Existence of critical, stationarity). *Let's now consider a Hawkes process with $||\Phi|| = 1$, $\mu = 0$ and*

$$\sup_{t \geq 0} t^{1+\gamma} \Phi(t) \quad \leq \quad R \tag{13}$$

$$\lim_{t \to \infty} t^{1+\gamma} \Phi(t) \quad = \quad r \tag{14}$$

where $r, R > 0$ and $\gamma \in\,]0, 1/2[$. Then the average intensity of such process is finite.

Summarizing, there exist non-trivial univariate Hawkes processes with $||\Phi|| = 1$ only for specific values of the tail exponent of $\Phi(t)$, which is required to lie in the interval $\gamma \in\,]0, 1/2[$.

Notice that, even though in $D = 1$ a quasi-stationary short-ranged Hawkes process is always degenerate, Ref. [42] describes a scaling regime for N_t in which it is possible to obtain to a non-degenerate process in the limit $||\Phi|| \to 1$ by appropriately choosing an observation timescale for the process. This behavior will be reviewed in Sec. 2.3.6.

In the multivariate setting, Ref. [51] shows that even in presence of kernels Φ satisfying the short-range condition Eq. (12) it is possible to define a non-trivial quasi-stationary Hawkes process in the large-dimensional limit. In particular Ref. [51] assumes a factorized form of the kernel, of the type

$$\phi^{ij}(t) = \alpha^{ij} f(t) \,, \tag{15}$$

with $||f|| = 1$, and consider the $D \to \infty$ limit of the process N_t. Such a limiting regime turns out to be well-defined also when $||\Phi|| = ||\alpha|| \xrightarrow[D \to \infty]{} 1$, provided that the matrix α has a sufficiently low density of eigenvalues in the vicinity of the critical point $||\alpha|| = 1$. Hence, a non-degenerate quasi-stationary limit for a Hawkes process can also be obtained as an effect of the interaction among a large number of components.

2.2.4 Non-linear Hawkes

Non-linear generalizations of the Hawkes process have been considered by several authors [16, 65, 62, 69]. The intensity function in the non-linear case is written as

$$\lambda_t^i = h\left(\mu^i + \sum_{i=1}^{D} \int \mathrm{d}N_{t'}^j \phi^{ij}(t-t')\right), \tag{16}$$

where $h(\cdot)$ is a non-linear function with support in $\mathbb{R}^+$. Typical choices for h include $h(x) = \mathbb{1}_{x \in \mathbb{R}^+}$ and $h(x) = e^x$. Note that the stability condition for various functions h was studied by Brémaud and Massoulié in [16]. The main advantage introduced by this extension is the possibility of modeling inhibition through negative valued kernels, although the price that has to be paid is the loss of mathematical tractability for most of the properties of the process. Yet, simulation and calibration of the model are possible even in this scenario [33, 8]. As we will see, negative valued kernels are found in the context of finance (see for instance Sec. 6.1).

Let us end this section by mentionning that we have only referred to the most common extensions of Hawkes original model but many further generalizations have been proposed like e.g., mixed diffusion-Hawkes models [29, 18], Hawkes models with shot noise exogenous events [23], Hawkes processes with generation dependent kernels [53].

2.3 Properties

The linear structure of the stochastic intensity λ_t of a Hawkes process allows us to characterize many of its properies in a completely analytical manner. Notably, its first- and second-order properties are particularly easy to compute, while a cluster representation of the process can be used in order to obtain a useful characterization of the Hawkes process. These properties are reviewed in the following Section.

2.3.1 First and second order properties

Assuming the hypothesis **(H)**, it is then possible to write explicitly the first- and the second-order properties of the model in term of the Laplace transform of the kernel Φ. In order to do this, it is necessary to introduce the function Ψ defined as follows:

Definition 2 (Kernel inversion)**.** *Consider a Hawkes process N_t with stationary increments. We define $\Psi(t)$ as the causal solution of the equation*

$$\Phi(t) + \Psi(t) * \Phi(t) = \Psi(t) \tag{17}$$

*As a consequence of **(H)**, $\Psi(t)$ exists and can be expressed as the infinite convolution*

$$\Psi(t) = \Phi(t) + \Phi(t) * \Phi(t) + \Phi(t) * \Phi(t) * \Phi(t) + \dots \tag{18}$$

The matrix function Ψ can be characterized analytically in term of the Laplace transform of the kernel Φ, which we define as follows:

8

Notation 4 (Laplace transform). *Given a scalar function $f(t) \in L^1(-\infty, +\infty)$, we denote its Laplace transform as*

$$\hat{f}(z) = \int_{-\infty}^{\infty} dt\, f(t)e^{zt} \tag{19}$$

Vector and matrix Laplace transforms are defined by applying the above transformation component-wise.

In the Laplace domain, Eq. (18) is then mapped to the algebraic relation

$$\hat{\Psi}(z) = (\mathbb{I} - \hat{\Phi}(z))^{-1} - \mathbb{I}, \tag{20}$$

where $\mathbb{I}$ denotes the identity matrix, allowing us to state the main result concerning the linear properties of a Hawkes process:

Proposition 4 (First- and second-order statistics). *For a Hawkes process N_t with stationary increments, the following propositions hold:*

1. The average intensity $\Lambda = \mathbb{E}[dN_t]/dt$ is equal to

$$\Lambda = (\mathbb{I} + \hat{\Psi}(0))\mu \tag{21}$$

2. The Laplace transform of the linear correlation matrix

$$c(t - t') = \frac{\mathbb{E}\left[dN_t dN_{t'}^T\right] - \mathbb{E}[dN_t]\mathbb{E}\left[dN_{t'}^T\right]}{dt dt'} \tag{22}$$

is equal to

$$\hat{c}(z) = (\mathbb{I} + \hat{\Psi}(-z))\, \Sigma\, (\mathbb{I} + \hat{\Psi}^T(z)), \tag{23}$$

where Σ is a diagonal matrix with non-zero elements equal to $\Sigma^{ii} = \Lambda^i$.

This useful characterization of the linear properties of a Hawkes process, first formulated in [37, 38] and then fully generalized in [8], allows to *(i)* obtain the linear predictions of a Hawkes model given μ and Φ as an input, *(ii)* calibrate non-parametrically the kernel Φ from empirical data by inverting relations (21) and (23) (see App. C.1).

Let us point out that, in Ref. [4], the authors have established that under general conditions, the empirical covariation of a multivariate Hawkes process converges towards its expected value, that can be easily expressed in terms of the Hawkes covariance matrix $c(t)$ as given by Eq. (23).

Example 3 (Exponential kernel). *Let's consider again the bivariate case described by Eq. (5) in the stationary case $1 > \alpha^{(s)} + \alpha^{(c)}$. In that case, the Laplace transform of the kernel $\hat{\Phi}(z)$ reads:*

$$\hat{\Phi}(z) = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} \hat{\phi}^{(s)}(z) + \hat{\phi}^{(c)}(z) & 0 \\ 0 & \hat{\phi}^{(s)}(z) - \hat{\phi}^{(c)}(z) \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \tag{24}$$

implying that the kernel matrix is diagonal in the symmetric and antisymmetric combinations $N_t^{\pm} = 2^{-1/2}(N_t^1 \pm N_t^2)$. The Laplace transform of the individual components of the kernel is

$$\hat{\Phi}(z) = \frac{\alpha^{(s/c)}}{1 - z/\beta^{(s/c)}}, \tag{25}$$

9

and due to the diagonal form of Eq. (5) the inverse kernel $\hat{\Psi}(z)$ can be computed straightforwardly. If one assumes $\mu = (\mu_0, \mu_0)$, then the relations above imply that $\Lambda = (\Lambda_0, \Lambda_0)$, with

$$\Lambda_0 = \frac{\mu_0}{1 - \alpha^{(s)} - \alpha^{(c)}} \,. \tag{26}$$

The Laplace transforms of the lagged cross-correlations of the symmetric and antisymmetric combinations $X_\pm(t)$, defined as

$$c^\pm(t - t') = \frac{\mathbb{E}\left[(\mathrm{d}N_t^1 \pm \mathrm{d}N_{t'}^2)(\mathrm{d}N_t^1 \pm \mathrm{d}N_{t'}^2)\right]}{2\mathrm{d}t\mathrm{d}t'} - \Lambda_0^2 \,. \tag{27}$$

result

$$\hat{c}^\pm(z) = \frac{\Lambda_0}{(1 - \hat{\phi}^{(s)}(-z) \mp \hat{\phi}^{(c)}(-z))(1 - \hat{\phi}^{(s)}(z) \mp \hat{\phi}^{(c)}(z))} \,. \tag{28}$$

Above expression can be inverted explicitly so to obtain the lagged cross-correlations in real space. In the simpler case $\beta^{(s)} = \beta^{(c)} = \beta_0$, one obtains for example

$$c^\pm(t) = \Lambda_0 \left(\delta(t) + \frac{\beta_0}{2} \frac{(\alpha^{(s)} \pm \alpha^{(c)})(2 - \alpha^{(s)} \mp \alpha^{(c)})}{(1 - \alpha^{(s)} \mp \alpha^{(c)})} e^{-(1 - \alpha^{(s)} \mp \alpha^{(c)})\beta_0 |t|} \right) \,. \tag{29}$$

Note that a singular component arises for $t = 0$ due to the assumption of unitary jumps $(\mathrm{d}N)^2 = \mathrm{d}N$. Above formula also shows that by moving the spectral norm $||\Phi|| = \alpha^{(s)} + \alpha^{(c)}$ close to the instability point $||\Phi|| = 1$, the decay of the symmetric mode of correlation function becomes slower and slower. Fig. 2 illustrates the result above for the autocorrelation function of the modes $N_t^\pm$.

Example 4 (Power-law kernel). *Let's now consider again the case of a power-law interaction kernel in dimension $D = 1$. For a kernel parameterized by Eq. (9), the Laplace transform reads*

$$\hat{\Phi}(z) = \alpha e^{-z/\beta}(-z/\beta)^\gamma \Gamma(-\gamma, -z/\beta) \,, \tag{30}$$

where $\Gamma(n, z)$ is the incomplete Gamma function. The inverse kernel $\hat{\Psi}(t)$ may be expressed in the Laplace domain as

$$\hat{\Psi}(z) = \frac{\alpha e^{-z/\beta}(-z/\beta)^\gamma \Gamma(-\gamma, -z/\beta)}{1 - \alpha e^{-z/\beta}(-z/\beta)^\gamma \Gamma(-\gamma, -z/\beta)} \,. \tag{31}$$

As shown in Sec. 2.1, in this case the spectral radius $||\Phi||$ is equal to $||\Phi|| = |\hat{\Phi}(0)| = \alpha/\gamma$, so that the model is stable for $\alpha < \gamma$. Under this assumption, the average intensity results

$$\Lambda = \mu \left(\frac{\gamma}{\gamma - \alpha} \right) \,. \tag{32}$$

For a fixed value of the exogenous intensity μ, this relation interpolates between a total intensity equal to the exogenous one (in the non-interacting case $\alpha = 0$) and an increasingly larger number of events as soon as α approaches the instability point

Auto-correlation   Normalized variance
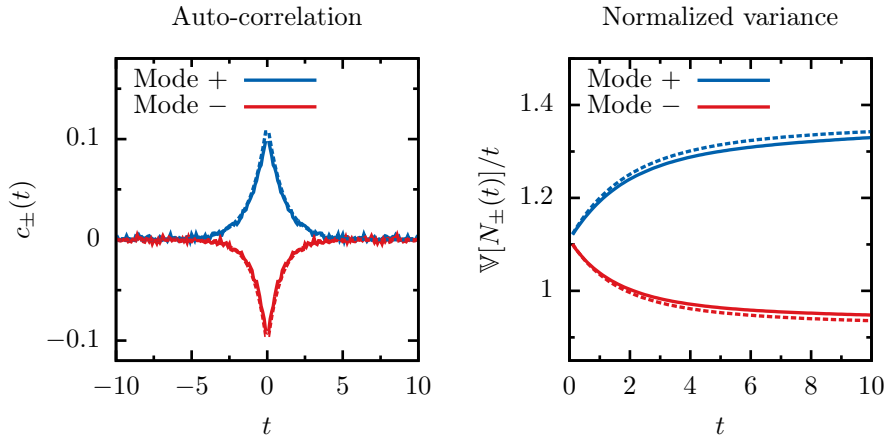
Figure 2: Autocorrelation function (left) and normalized variance (right) of the combinations $N_t^{\pm}$ diagonalizing the interaction kernel appearing in Eq. (5). We have used the parameter set $\alpha^{(s)} = 0$, $\alpha^{(c)} = 0.1$, $\mu_0 = \beta_0 = 1$ in order to simulate a single realization of the process of length $T = 10^5$. For such a value of T, the theoretical predictions (dashed lines) are almost exactly superimposed to the results of the simulations. The $\delta(t)$ component in the cross-covariance function has been omitted in the left panel for the sake of clarity.

$\alpha = \gamma$.
The Laplace transform of the lagged cross-correlation matrix results

$$\hat{c}(z) = \frac{\Lambda}{(1 - \hat{\Phi}(-z))(1 - \hat{\Phi}(z))} \ . \tag{33}$$

The above expression cannot be inverted analytically. One can indeed relate the tail behavior of the correlations to the small z behavior of $\hat{c}(z)$ thanks to Tauberian Theorems. In particular, for $\beta t \gg 1$ one has:

$$c(t) \sim \begin{cases} e^{-(\gamma-1)\beta t} & \text{for} \quad \gamma > 1 \\ (\beta t)^{-\gamma-1} & \text{for} \quad \gamma < 1 \end{cases} \tag{34}$$

This behavior is illustrated in Fig. 3, where we compare the autocorrelation function of several univiariate Hawkes processes with power-law kernel and different tail exponents. As a final note, we remark that for $\gamma < 1/2$, and close to the instability point $\alpha = \gamma$, the function $c(t)$ obeys an intermediate asymptotics $c(t) \sim t^{2\gamma-1}$, which holds as long as

$$\beta t \ll \left(\frac{\Gamma(1-\gamma)}{\gamma/\alpha - 1} \right)^{1/\gamma} \ . \tag{35}$$

In this particular regime, the Hawkes process develops an apparent Hurst exponent $H = 1/2 + \gamma$. This limiting behavior is a consequence of the quasi-stationarity condition $||\Phi|| = 1$, analyzed in Ref. [17] and reviewed in Sec. 2.2.3.
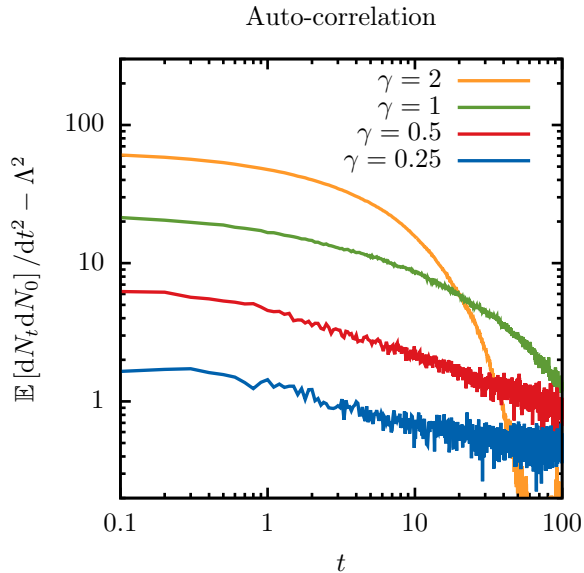
Figure 3: Comparison of the autocorrelation function for a set of univariate Hawkes processes with power-law kernels parametrized by Eq. (9). We used the values $\mu = \beta = 1$, $\alpha = 0.9\,\gamma$, and simulated processes of length $T = 5 \cdot 10^5$ (for $\gamma = 0.25$ and 0.5), $T = 10^6$ (for $\gamma = 1$) and $T = 5 \cdot 10^6$ (for $\gamma = 2$) in order to obtain the curves represented in the plot. The image illustrates the crossover from the exponential decay of correlation obtained for $\beta > 1$ to the power-law behavior detected for $\beta < 1$.

2.3.2 Characterization through second order statistics

This section justifies why the Hawkes process can be thought of as the simplest example of an interacting point process. In fact, it is entirely characterized by its first- and second-order properties: means and correlations uniquely determine a Hawkes process through the solution of a Wiener-Hopf system.

Let us start by defining the *conditional intensity* matrix $g(t)$, that we define for $t > 0$ as

$$g^{ij}(t) = \frac{\mathbb{E}\left[dN_t^i \,|\, dN_0^j = 1\right]}{dt} - \Lambda^i, \quad \forall t > 0. \tag{36}$$

Then one can prove straightforwardly [8] from Eq. (23) that

$$c(t) = \Sigma g^T(t), \quad \forall t > 0 \tag{37}$$

12

which relates conditional averages and lagged cross-correlations. Hence by using Eq. (37) and (23), one can prove that [8]:

Theorem 2 (Wiener-Hopf equation). *Consider a Hawkes process defined by Eq. (1) satisfying the stationarity assumption* **(H)**. *Then the matrix function* $\chi(t) = \Phi(t)$ *is the unique solution of the Wiener-Hopf system*

$$g(t) = \chi(t) + \chi(t) * g(t) \quad \forall t > 0 \tag{38}$$

such that the components $\chi^{ij}(t)$ *are causal and* $\chi^{ij}(t) \in L^1 \; \forall i, j$.

This property implies that, when fixing an average intensity vector Λ and a conditional expectation $g(t)$, there exist at most one Hawkes process consistent with these observables. Indeed, such a process is not always guaranteed to exist, as a Hawkes process doesn't necessarily reproduce the linear properties for systems in which inhibition is relevant.

This result is the inverse one with respect to the one expressed by Eq. (23): while that equation expresses the fact that by fixing a kernel $\Phi(t)$ and an exogenous intensity μ the correlations are uniquely determined, the theorem above states that correlations and average intensities uniquely fix the interactions. While the direct result was first proved in [37], the converse was shown in [8] by using the Wiener-Hopf factorization technique.

Finally note that the Wiener-Hopf system (38) is very useful in applications to empirical data, as it allows us to estimate non-parametrically the interation kernel of a Hawkes process given a set of empirical observations (see App. C.2).

2.3.3 Auto-regressive projection

A Hawkes processes with stationary increments can always be linearly approximated by suitably defined auto-regressive processes. In particular one can show that [51]:

Proposition 5 (Auto-regressive projection). *Consider a Hawkes process* N_t *defined by Eq. (2) satisfying the stationarity assumption* **(H)**. *Then the convolution* $N_t^{(AR)}$ *defined by*

$$N_t^{(AR)} = (\mathbb{I}\delta(t) + \Psi(t)) * (\mu\, t + \Sigma^{1/2} W_t)\,, \tag{39}$$

where $\Psi(t)$ *is the infinite convolution of the Hawkes kernel given by Eq. (18), and* W_t *is a standard D-dimensional Brownian motion, satisfies*

$$\Lambda^{(AR)} = \Lambda \tag{40}$$
$$c^{(AR)}(t) = c(t)\,, \tag{41}$$

where Λ *and* c *denote respectively the average intensity and the lagged cross-correlation matrix of* N_t.

Hence, it is always possible to match the first and the second order properties of a stationary Hawkes process with interaction kernel Φ by using a convolution of Wiener processes. On the other hand, higher order moments cannot be matched in the same way.

2.3.4 Beyond second-order

The first- and second-order moments are not the only moments which can be computed analytically for a Hawkes process. In particular, Jovanović *et al.* [44] have recently developed a combinatorial procedure allowing to calculate cumulants (and consequently, moments) of arbitrary order of a Hawkes process. More precisely, given a set of components $S \in \{1, \ldots, D\}$ and one of times $t_S = \{t_1, \ldots t_{|S|}\}$, it is possible to define a *cumulant density* of a Hawkes process as

$$k\left(N^{(S)}\right) = \mathrm{d}t^{-|S|} \sum_\pi (|\pi| - 1!)(-1)^{|\pi|-1} \prod_{B \in \pi} \left\langle \prod_{i \in B} \mathrm{d}N^i_{t_i} \right\rangle, \qquad (42)$$

where the sum runs over all the partitions π of S, $|\pi|$ denotes their number of blocks and B labels individually the blocks of π. Above equation generalizes the definition of the average intensity, recovered for $|S| = 1$, and of the lagged cross-correlation function, obtained for $|S| = 2$. Moment densities can be obtained from above expression by writing

$$\left\langle \prod_{i \in S} \mathrm{d}N^i_{t_i} \right\rangle \mathrm{d}t^{-|S|} = \sum_\pi \prod_{B \in \pi} k\left(N^{(B)}\right). \qquad (43)$$

Ref. [44] shows how to express Eq. (42) as a sum of integral terms, which can be written explicitly in terms of μ and $\Psi(t)$. As each of such addends can be interpreted as a topologically distinct rooted tree with $|S|$ labeled leaves, the enumeration of all the contribution to Eq. (42) can be performed systematically.

In the special case of a Hawkes processes with exponential kernel, a useful result is obtained by Errais *et al.* [29] by exploiting Dynkin's formula for the couple (N, λ) in the marked framework described in Sec. 2.2.1. In particular, they are able to express the generating function for the couple (N, λ) in terms of the solution of an ordinary differential equation. While it may be necessary to solve numerically the equation for the generating function, closed-form expressions are available for specific moments of N_t (see also the work of Dassios and Zhao [23] for similar results on a slight generalization of Hawkes processes).

2.3.5 Martingale representation

The process defined by Eq. (2) admits a convenient martingale representation once one introduces suitably defined *compensators* $\int_0^t \mathrm{d}s\, \lambda^i_s$ [22].

Theorem 3 (Martingale representation). *Given a Hawkes process (2), the D stochastic processes*

$$Y_t = N_t - \int_0^t \mathrm{d}s\, \lambda_s \qquad (44)$$

are martingales with respect to the canonical filtration of the process N_t [22]. Additionally, under the stability condition (**H**), *the stochastic intensity λ_t admits the representation*

$$\lambda_t = \mu + \int_0^t \Psi(t - s)\mu\, \mathrm{d}s + \int_0^t \Psi(t - s)\, \mathrm{d}Y_s. \qquad (45)$$

14

While the above result is valid even in the non-stationary regime, Eq. (45) takes a particularly simple form in the asymptotic regime of large t, where it can be written as

$$\lambda_t \xrightarrow[t\to\infty]{} \Lambda + \int_0^t \Psi(t-s)\,\mathrm{d}Y_s \tag{46}$$

thanks to Eq. (21). The martingale property of the Hawkes process plays an important role in determining the first- and second-order properties discussed in above section, which are derived by means of Eq. (46) in [2, 8]. The representation Eq. (45) is also particularly useful in order perform predictions of the intensity of the Hawkes process given an historic filtration $\mathcal{F}_t$ of the process.

Example 5 (Prediction of the intensity). *Suppose that, given a Hawkes process which satisfies the stationary condtion* (**H**)*, one is interested in computing the predictor $\mathbb{E}\left[\lambda_t\,\middle|\,\mathcal{F}_s\right]$ for $t > s$. (see Ref. [40, 41] for direct applications in Finance). While, by naively applying the definition of the Hawkes processes, one can obtain an implicit equation of the type*

$$\mathbb{E}\left[\lambda_t\,\middle|\,\mathcal{F}_s\right] = \mu + \int_0^s \Phi(t-u)\mathrm{d}N_u + \int_s^t \mathrm{d}u\,\Phi(t-u)\mathbb{E}\left[\lambda_u\,\middle|\,\mathcal{F}_s\right] \tag{47}$$

the martingale representation Eq. (45) can be used in order to write the explicit expression

$$\mathbb{E}\left[\lambda_t\,\middle|\,\mathcal{F}_s\right] = \mu + \int_0^t \Psi(t-s)\mu\,\mathrm{d}s + \int_0^s \Psi(t-s)\,\mathrm{d}Y_s\,. \tag{48}$$

2.3.6   Scaling limit of the process

The structure of Hawkes processes is naturally adapted to describe systems in which the discrete nature of the jumps in the coordinates N_t is relevant, making this model especially suitable to the modelization of high-frequency data. Indeed, in many applications one additionally needs to control the limiting behavior the system in the opposite regime of low frequencies, where the granularity of events is disregarded, and the scaling in time of N_t needs to be known.

Diffusion towards a Brownian motion.   The following results, first proved in [4] allow us to establish that Hawkes processes, under appropriate hypotheses and after a suitable rescaling, behave at large times as linear combinations of Wiener processes.

Theorem 4 (Law of large numbers). *Consider a Hawkes process as in Eq. (2) satisfying the stationarity assumption* (**H**)*. Then*

$$\sup_{u\in[0,1]} ||T^{-1}N_{uT} - u\Lambda|| \xrightarrow[T\to\infty]{} 0 \tag{49}$$

almost surely and in L^2-norm.

Above result establishes a law of large numbers for the Hawkes process, valid for any stationary kernel. Indeed, under additional hypotheses it is also possible to formulate a corresponding functional central-limit theorem.

15

Theorem 5 (Central-limit theorem). *Suppose that for all $i, j \leq D$ the kernel $\Phi(t)$ satisfies*

$$\int_0^\infty \mathrm{d}t \, t^{1/2} \phi^{ij}(t) < \infty \,. \tag{50}$$

Then for $u \in [0, 1]$ one has the following convergence is in law for the Skorokhod topology:

$$T^{1/2} \left(T^{-1} N_{uT} - u\Lambda \right) \xrightarrow[T \to \infty]{} (\mathbb{I} + ||\Psi||)\Sigma^{1/2} W_u \,, \tag{51}$$

where W_t denotes a standard D-dimensional Brownian motion.

Example 6 (Exponential kernel). *Consider the bivariate Hawkes process analyzed in Sec. 2.1. Eqs. (5) and (51) above imply that*

$$\begin{aligned}
N_{uT} \quad &\xrightarrow[T \to \infty]{} \quad \Lambda_0 uT \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{52} \\
&+ \quad \frac{(\Lambda_0 T)^{1/2}}{(1 - \alpha^{(s)})^2 - (\alpha^{(c)})^2} \begin{pmatrix} 1 - \alpha^{(s)} & \alpha^{(c)} \\ \alpha^{(c)} & 1 - \alpha^{(s)} \end{pmatrix} \begin{pmatrix} W_u^1 \\ W_u^2 \end{pmatrix} \,,
\end{aligned}$$

which in term of the combinations $N_t^\pm = 2^{-1/2}(N_t^1 \pm N_t^2)$ reads

$$N_{uT}^\pm \xrightarrow[T \to \infty]{} (1 \pm 1)^{1/2} \Lambda_0 uT + \frac{(\Lambda_0 T)^{1/2}}{1 - \alpha^{(s)} \mp \alpha^{(c)}} W_u^\pm \,. \tag{53}$$

This behavior is summarized in Fig. 4, where we compare the rescaled processes $N_{uT}^\pm$ at different timescales T.

Other scaling limits   It is well known that many of the microscopic observables involved in the price formation process do not diffuse at large scales (e.g., at the daily, or even monthly, time-scale) toward Brownian motion dynamics. For instance, the trading activity (i.e., the market-order flow) seems to be long-range dependent [14], while the volatility displays clusters that can last for months. Since, as we will see in the next sections, Hawkes processes tend to mimic very well the high-frequency dynamics of financial time-series, it is natural to try to understand if diffusive limits other than the one described by Theorem 5 can be found. In a recent work by Jaisson and Rosenbaum [42], a sequence of rescaled univariate Hawkes processes

$$Z_u^{(T)} = \frac{1 - a_T}{T} N_{uT}^{(T)} \tag{54}$$

indexed by a time-scale parameter T and where $a_T \in [0, 1[$ is proved to converge, when $T \to +\infty$, towards an integrated Cox, Ingersoll, Ross process [20] under the main following conditions [4]

- the corresponding sequence of kernels $\phi^{(T)}(t)$ satisfies $\phi^{(T)}(t) = a_T \phi(t)$, where ϕ is differentiable and it is such that $||\phi|| = \hat{\phi}(0) = 1$, $\hat{\phi}'(0) < +\infty$, $||\phi'||_\infty < +\infty$ and $||\phi'|| < +\infty$,

---

[4]For precise formulation of the corresponding theorem, we refer the reader to [42].

$$Z_u^\pm = \frac{1-\alpha^{(s)}\mp\alpha^{(c)}}{(\Lambda_0 T)^{1/2}}\left(N_{uT}^\pm - \Lambda^\pm uT\right)$$
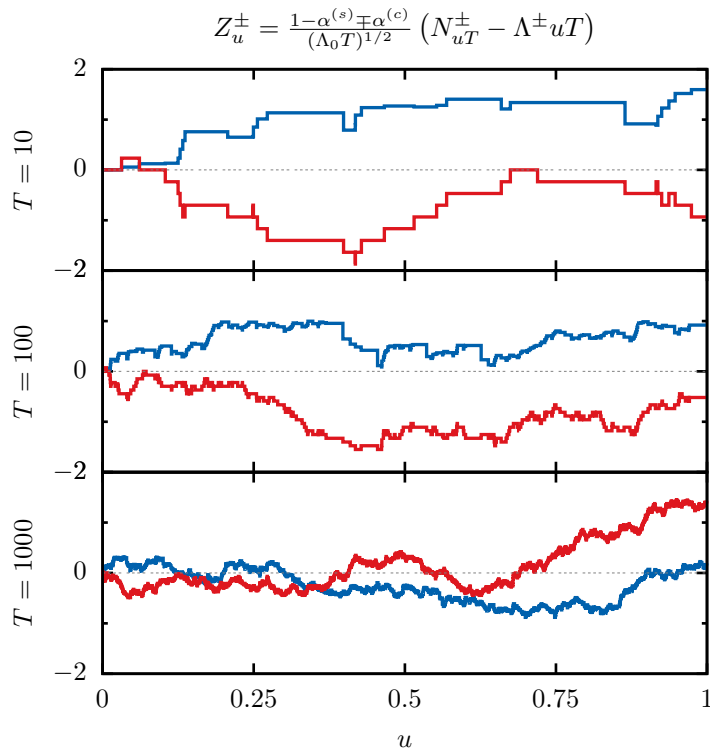
Figure 4: The figure illustrates the shape of the process $N_{uT}^\pm$ at different timescales $T = 10, 100, 1000$ for the same choice of paramters as in Fig. 2. The process has been rescaled according to Eq. (53) so to obtain a standard Wiener process W_u in the limit $T \to \infty$. The plot illustrates how the presence of discrete jumps, clearly visible at small times, becomes irrelevant in the scaling limit $T \to \infty$.

- the criticality condition **(H)** of Prop. 1 is met at a speed

$$\lim_{T \to +\infty} (1 - a_T)T = \kappa, \quad \kappa > 0. \tag{55}$$

Hence, even though a quasi-stationary short-ranged Hawkes process is always degenerate, one can detect a non-trivial behavior of the rescaled counting function by suitably choosing an observation timescale $T \sim (1 - ||\Phi^{(T)}||)^{-1}$. Let us point out that the above conditions do not allow for ϕ to have a power-law decay with an exponent strictly smaller than 2 (in the limit $t \to +\infty$). As reviewed in Sec. 3, several studies tend to show that for many financial time-series (e.g., market order flow time-series) the relevant kernel has power-law tails with an exponent above but rather close to 1. Thus, strictly speaking, the previous framework is inappropriate to describe such behavior. Using the same asymptotics ($T \to +\infty$), Jaisson [41] studied the asymptotic limit of the correlation function of a 1-dimensional Hawkes process with a power-law kernel which decreases with an exponent $1 + \gamma$ with $\gamma \in]0, 1/2[$. He proved that the auto-correlation function of the Hawkes process decreases asymptotically as a power-law with an exponent $1 - 2\gamma$, i.e., leading to a long-range dependence[5].

2.3.7 Clustering representation

Another useful property of the Hawkes process is the *clustering* property which emerges as a consequence of the linearity of Eq. (2). Such property allows one to *(i)* build an efficient *simulation* algorithm for the process (see App. B), *(ii)* introduce the notion of *parenthood* among different events (see Sec. 2.3.8 below), *(iii) infer* parenthood relations among successive events from empirical data (see the paragraph about the EM method in App. C.2).

Proposition 6 (Clustering representation). *Consider a positive integer D and a (non-necessarily finite) time interval $[0, T]$, in which we define a sequence of* events $\{(t_m, k_m)\}_{m=1}^M$ *according to the following procedure:*

- *For each $1 \leq i \leq D$, consider a set of* immigrant *events $\{(t_m^{(0)}, i)\}_{m=1}^{M_i^{(0)}}$ extracted with homogeneous Poissonian rate μ^i in the interval $[0, T]$.*
- *For each immigrant event of type j, labeled by $(t_{m'}^{(0)}, j)$, and for each $1 \leq i \leq N$, generate a sequence of* first-generation *events $\{(t_m^{(1)}, i)\}_{m=1}^{M_i^{(1)}}$ sampled with time-dependent Poissonian rate $\phi^{ij}(t - t_{m'})$ in the interval $[t_{m'}^{(0)}, T]$.*
- *Iterate above rule from generation $n - 1$ to generation n, so to obtain the event sequence $\{(t_m^{(n)}, k_m^{(n)})\}_{m=1}^{M^{(n)}}$, until no more events are generated in $[0, T]$.*

Then the union of all the events

$$\{(t_m, k_m)\}_{m=1}^M = \bigcup_{n=0}^{\infty} \{(t_m^{(n)}, k_m^{(n)})\}_{m=1}^{M^{(n)}} \tag{56}$$

corresponds to the one generated by the Hawkes process (2) in the time interval $[0, T]$.

---

[5]Let us notice that qualitative arguments for similar result were also given in [7] in a 2-dimensional framework.

Note that the construction above (depicted in Fig. 5) can be equivalently taken as a definition for the Hawkes process, once the information encoding the generation n is discarded by taking the union of all the events. Indeed, this richer definition
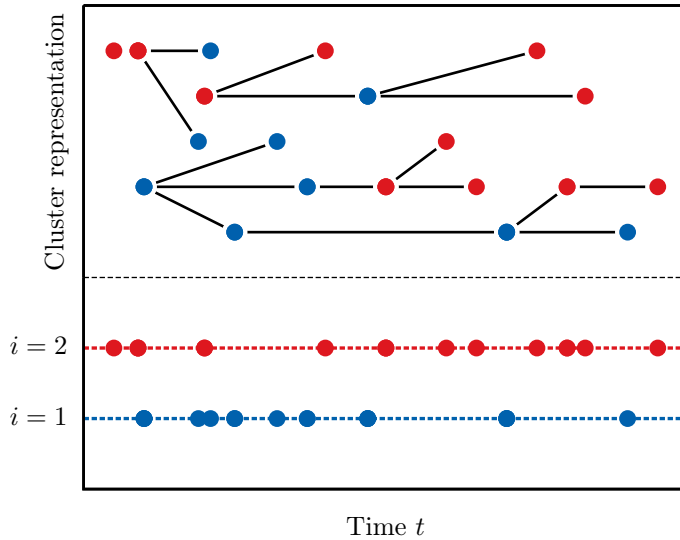
Figure 5: Cluster representation of a Hawkes process: while the upper panel represents the branching structure of a bivariate Hawkes process, the lower panel shows its projection obtained by disregarding the cluster structure. The different components $i \in \{1, 2\}$ are shown in different colors, while the connected structures in the upper panel denote three different clusters.

characterizes more transparently the stationarity condition **(H)**: by considering $T = \infty$ in above construction, it is possible to map the branching structure of the Hawkes process described above onto the one of a Galton-Watson tree with average offspring $||\Phi||$. The qualitative behavior of the model is in fact dictated by the average number of events generated by a parent event, equal to $\int_0^\infty \phi(t) = \hat{\phi}(0) = ||\Phi||$. The three phases of the Hawkes process (stationary $||\Phi|| < 1$, non-stationary $||\Phi|| > 1$ and quasi-stationary $||\Phi|| = 1$) correspond then to the three phases of a Galton-Watson branching process, more precisely:

1. For $||\Phi|| < 1$, we have a *sub-critical* phase in which each parent event generates on average less than one child event. This implies that the total progeny of each event is a.s. finite and the average number of generations before extinction is a.s. finite.

19

2. For $||\Phi|| > 1$, we have a *super-critical* phase in which more than one child event is generated by each parent event. In that case the total progeny of a parent event might be infinite with finite probability.

3. For $||\Phi|| = 1$ (the *critical* case), the total progeny is a.s. finite, but the total size of the progeny has large fluctuations leading to a divergence of the average number of generations before extinction.

2.3.8 Causality

The parenthood relation introduced by using the clustering representation of the Hawkes process allows us to discuss the problem of causality in this context. In particular, after constructing a Hawkes process according to the branching procedure described above, one can introduce the counting functions

$$\begin{aligned}
N_t^{i \leftarrow 0} &= \text{(Exogenously generated events of type } i) & (57)\\
N_t^{i \leftarrow j} &= \text{(Events of type } i \text{ with type } j \text{ direct ancestor)} & (58)\\
N_t^{i \leftarrow j*} &= \text{(Events of type } i \text{ with type } j \text{ oldest ancestor)} & (59)
\end{aligned}$$

so that $N_t^{i \leftarrow 0} + \sum_j N_t^{i \leftarrow j} = N_t^{i \leftarrow 0} + \sum_j N_t^{i \leftarrow j*} = N_t^i$. Hence, these quantities can be used in order to express the overall number of events of type i generated by an ancestor of a given type j. In particular, it is easy to prove that:

Proposition 7 (Causality). *For a stationary Hawkes process, the average increments of $N_t^{i \leftarrow 0}$, $N_t^{i \leftarrow j}$ and $N_t^{i \leftarrow j*}$ are expressed by*

$$\begin{aligned}
\mathbb{E}\left[\mathrm{d}N_t^{i \leftarrow 0}\right]/\mathrm{d}t &= \mu^i & (60)\\
\mathbb{E}\left[\mathrm{d}N_t^{i \leftarrow j}\right]/\mathrm{d}t &= \hat{\phi}^{ij}(0)\Lambda^j & (61)\\
\mathbb{E}\left[\mathrm{d}N_t^{i \leftarrow j*}\right]/\mathrm{d}t &= \hat{\psi}^{ij}(0)\mu^j\ . & (62)
\end{aligned}$$

The property above can be used in order to estimate the average fraction of events (directly or indirectly) caused by a specific component of a Hawkes process.

Alternative notions of causality for point-processes have also been investigated. In particular, the notion of Granger-causality has been extended to point-processes in [57, 25].

3 Univariate models

3.1 Models of market activity and risk

The first straightforward application of Hawkes processes in high frequency finance is probably to model the so-called volatility clustering phenomenon. Since volatility at the transaction level can be directly related to the number N_T of a given type of events (trades, mid-price changes,...) that occur in a given time interval of size T[6],

---

[6]One can have in mind a simple price model where the price is build as the sum of independent random shocks. In this case, the volatility at scale T is proportional to $(N_T)^{1/2}$. The empirical results of Ref. [68]

the self-exciting nature of Hawkes processes provides a very simple picture that can explain the correlated nature of volatility fluctuations. This idea was first proposed by Bowsher [15] who calibrated a univariate Hawkes model with mixture of exponential kernels using intraday equity data from NASDAQ and NYSE [7].

In Ref. [2], Bacry *et al.* recently introduced a non-parametric estimation method for multivariate symmetric Hawkes processes based on the spectral factorization of the covariance matrix by the means of the Hilbert transform. By calibrating a 1-dimensional Hawkes model to the occurrence of trades of the 10 years Euro-Bund future front contract over 75 trading days in 2009, they discovered two important empirical facts: (i) the model is very close to its stability threshold $||\Phi|| = 1$ and (ii) the empirical kernel $\Phi(t)$ is very well decribed, over a wide range of scales, by the power-law function Eq. (9)

$$\Phi(t) = \frac{\alpha\beta}{(1 + \beta t)^{1+\gamma}} \tag{63}$$

with $\gamma \simeq 0$. The first observation directly concerns the level of endogeneity and the one of stability of financial markets, a problem, as discussed in the next section, that has been addressed afterwards by Filimonov and Sornette or Hardiman *et al.* [31, 32, 34, 35]. The power-law nature of Hawkes kernels with an exponent $\gamma \simeq 0$ has been confirmed by studies that followed, notably by Hardiman *et al.* on mid-price changes of E-mini S&P500 futures [34] or by Bacry and Muzy on trades arrivals of EuroStoxx index futures [8]. The plots of Fig. 6 are directly extracted from these papers: they represent in log-log scales the estimated Hawkes kernel for the E-mini SP futures mid-price change events and the EuroStoxx market order occurences. One can see that the two estimated kernels, corresponding to different data, different markets and different estimation methods, are strikingly similar. This suggests some universality of both γ and $C = \alpha\beta^{-\gamma}$ parameters in the algebraic decay of Eq. (63).

The origin of this power-law behavior and, in particular, of the values of α and γ remains an open question. Previous empirical results motivated the work by Jaisson [41] (see also Ref. [7] and the section preceding Eq. (35)), who proved that, within a particular asymptotics (which includes the L^1 norm of the kernel converging to 1), a 1-dimensional kernel with a power-law decay with an exponent $1 + \gamma$ (with $\gamma \in]0, 0.5[$) leads to an auto-covariance function of the flow which is power-law with an exponent $1 - 2\gamma$, i.e., to a long-range dependence of the flow (see also the discussion in Sec. 2.3, Eq. (34)). The algebraic decay of Hawkes kernels can then be related to the volatility clustering properties. Notice that within trading time models [14], the long-range nature of offer and demand has been mostly explained in terms of order splitting dynamics, and is supported by empirical results such as [66], in which broker-resolved data is analyzed in order to distinguish among the herding and splitting components of the order flow. It is thus likely that the observed slowly decreasing nature of the kernel directly results from the splitting of the meta-orders. Such an hypothesis remains however to be supported by quantitative arguments and to be confirmed by

---

corroborate empirically this observation, allowing one to establish more precisely of the relation among impact per trade and volatility.

[7]More precisely Bowsher considered a generalization of Hawkes processes in order to account for the peculiar non-stationnarities observed at high frequency like intraday seasonalities and overnight gaps.
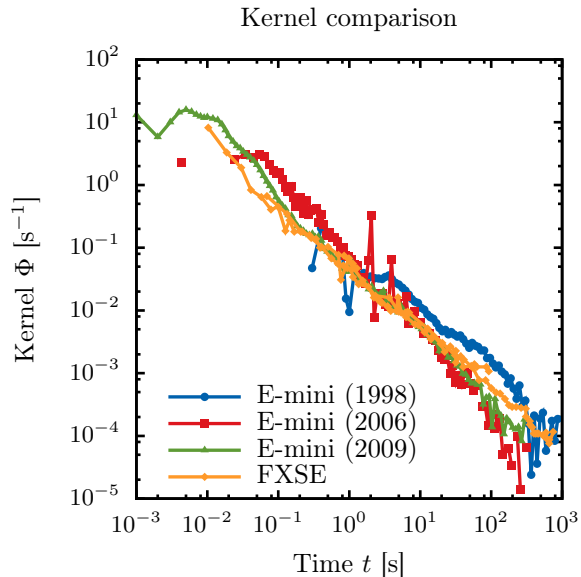
Figure 6: Inferred kernel $\Phi(t)$ of the univariate Hawkes process describing midpoint changes of the E-mini S&P500 and for the trade arrivals of the EuroStoxx in different years, reproduced respectively from Refs. [8] and [34]. Strikingly, considering the different markets, traded contract, type of events considered, and periods of time considered in these studies, the shape of $\Phi(t)$ is almost identical.

empirical observations. Beyond these fundamental questions, the power-law nature of Hawkes kernels remains a solid empirical fact which, at least, calls to question all the approaches based on exponential Hawkes models.

In the same lines of the previously cited studies, Da Fonseca and Zaatour [21], perform a parametric estimation (using a GMM approach, see Sec. C.1) of a 1-dimensional Hawkes process with exponential kernels of the form $\alpha\beta\exp(-\beta t)$ on (unsigned) market-order flow data. As expected, market-order clustering translates in a rather low value for $\beta \in [0.02, 0.1]$ (depending on the financial time-series) and the L^1 norm of the kernel is found to be very close to criticality, i.e., $\alpha \geq 0.9$ (and very often ≥ 0.95). One can also cite the work of Lallouache and Challet [45], who performed a maximum likelihood estimation on market orders using a sum of two exponential functions as the Hawkes kernel. They study forex EBS data which are throttled (market orders are gathered within slices of 0.1 seconds) which calls for a rather complex denoising preprocessing. Using goodness of fit tests, they conclude that, though the model performs well when applied to a 1 hour specific intraday time (averaged every day over 3 months), the intraday seasonality of (mainly) the exogeneous intensity μ does not allow to fit well a whole day.

Hawkes processes have also been used to model extreme price moves at a rather low frequency. In Ref. [27], Embrechts *et al.* study an equally weighted portfolio of 3 indices

(Dow Jones, Nasdaq and SP) on an hourly time-frame (on 14 years). Only extreme quantiles of returns are kept (the smallest and the largest 1% quantiles) leading to a 1-dimensional point process whose jumps correspond to an extreme return of any of the three indices. The jumps are then marked using a 3-dimensional marks coding the excess of each index respect to the corresponding 1% quantile. The so-obtained point process is modeled using a 3-dimensional Hawkes process (with an exponential kernel) marked by 3-dimensional Gamma-distributed i.i.d. random variables. A maximum likelihood estimation is performed and goodness of tests show that this model is well suited for modeling extreme price moves. Let us point out that, in the same work, a very similar experiment is performed on daily log-returns of an (home-made) index of stocks using this time a 2-dimensional Hawkes process (for coding positive or negative extreme jumps) with 1-dimensional marks. In the same spirit, in Ref. [18], Chavez-Demoulin and Mc Gill constructed a model for the excessses of an asset price above a given threshold. This model combines a Hawkes process for the exceedance occurences with Pareto distributed marks to the exceedance sizes. Within this approach, the author's goal was to describe the clustering of large drawdown events through the self-excited dynamics of the Hawkes process. By performing backtests on equities intraday data, they have shown that this model captures very well extreme intraday events, notably as compared to standard non-parametric methods based on extreme value theory.

3.2 Measuring the endogeneity of stock markets

In a recent series of papers [31, 32, 34, 35], some authors addressed, within the framework of Hawkes models, the important problem of the so-called *"volatility puzzle"*, namely the fact that the observed market volatility cannot be explained by classical economic theory. Indeed, it is well known that prices move too much compared to the flow of pertinent information that may impact the market. This observation naturally leads to the idea that price dynamics is highly endogenous, i.e. mainly driven by some internal feedback mechanisms. Filimonov and Sornette [31] were the first to propose a quantitative measure of the level of "market reflexivity". For that purpose, they model the high frequency mid-price variations of some stock index (namely the E-mini S&P500) as a 1-dimensional Hawkes process. As explained in Sec. 2.3.8, $||\Phi||$ can be interpreted as a branching ratio, i.e., the number of events generated by any parent event. Each exogenous event occurring at rate μ thus generates $||\Phi||/(1 - ||\Phi||)$ events and therefore the ratio of endogenous event rate to the overall rate Λ in one dimension is, according to Eq. (21),

$$\frac{\mu}{\Lambda} \left(\frac{||\Phi||}{1 - ||\Phi||} \right) = ||\Phi|| \ .$$

This means that $||\Phi||$ provides a direct measure of the fraction of endogenous events within the whole population of mid-price changes and thus a measure of the market reflexivity. By analyzing the E-mini S&P500 future contracts over the period 1998-2010, Filimonov and Sornette found that the degree of reflexivity has strikingly increased during the last decade. They suggested that this effect could be directly caused by the increasing amount of high frequency and algorithmic trading, raising the question of the impact of high frequency trading on the market stability. This analysis has

been revisited by Hardiman, Bercot and Bouchaud [34] who noticed that Filimonov and Sornette estimation relying on an exponential parametrization is biased because, as discussed previously, empirical evidences suggest that Hawkes kernels have a slow (power-law) decay. Accounting for this feature on their estimation of $||\Phi||$, Hardiman *et al.* found that the reflexivity of the E-mini S&P future hasn't been increasing during the last decade, but has remained constant at a value very close to the critical one $||\Phi|| = 1$. It is noteworthy that Hardiman *et al.* also provided, in their study, empirical evidences of the existence of a high frequency cut-off of the Hawkes kernel (namely the parameter β^{-1} in Eq. (63)) that decreases exponentially fast in time and that can be associated with the increase of the trading frequency. In a more recent paper, Filimonov and Sornette [32], have reviewed all the pitfalls associated with the estimation $||\Phi||$ in the case of a slowly decreasing kernel. They have shown that significant biases can be induced by the presence of outliers, edge effects or non stationary effects. Because some important issues are also related to the way one parametrizes the model (notably the choice of the high-frequency regularization), Hardiman and Bouchaud [35] proposed a simple non-parametric approximation of the branching ratio $||\Phi||$ that relies on Eq. (23). Indeed, by considering this equation in $z = 0$, it is possible to relate the integral of the correlation function, $\hat{c}(0)$ to $||\Phi||$. The number of events N_T in a window of size T, becomes, for T large enough, $\hat{c}(0) \simeq T^{-1}\mathbb{V}[N_T]$ and therefore one gets

$$||\Phi|| \simeq 1 - \left(\frac{\mathbb{V}[N_T]}{\mathbb{E}[N_T]}\right)^{1/2} . \tag{64}$$

This formula leads to a very intuitive intepretation of the degree of reflexivity: The occurrence of correlated events implies an increase of the variance of N_T with respect to its mean value (for a Poisson process, both quantities are equal so that one directly gets $||\Phi|| = 0$). Using this model free estimator, Hardiman and Bouchaud [35] have confirmed their former claims that the S&P 500 future appears to have, during the last ten years, a stable level of reflexivity, close to the criticality. Let us notice that this formula only holds for 1-dimensional Hawkes processes and has no simple extension in the multivariate situation.

Beyond the debate on the most suitable estimator for the reflexivity parameter and its genuine behavior, the pionneering work of Filimonov and Sornette provided a quantitative framework allowing to study the endogeneity of market fluctuations with Hawkes processes. They notably have shown that such approach can be used to study particular events such as the flash crashes of April and May 2010 [31]. Their results may be helpful to devise warning tools in order to anticipate extreme drawdowns which are of endogenous origin. The prospects and applications along this path are numerous. One important question concerns the extenstion of such studies by accounting other types of events like e.g. order book events (see Sec. 6).

4   Price models

Describing the fluctuations of price at the finest time scales, with notably the goal of improving volatility and covariance estimations, is a central issue of financial econometrics. The notion of microstructure noise was considered by many authors as an

additional noise, superimposed to the standard diffusion, that accounts for the small scale behavior of the signature plot. Indeed, it is well know that the signature plot

$$C(\tau) = \frac{1}{T} \sum_{i=0}^{T/\tau} \left[P_{(i+1)\tau} - P_{i\tau} \right]^2$$

that corresponds to the quadratic variation of the mid-price P_t at scale τ, strongly increases when $\tau \to 0$ (see Fig. 7). Along the same line, the so-called Epps effect accounts for the vanishing covariation among the returns of pairs of assets when the return scale τ goes to zero. Bacry *et al.* in [3] proposed as an alternative to standard "latent price" models, to directly account for the discrete nature of price variations. They have been the first ones to describe the tick-by-tick variation of the mid-price, P_t, within the framework of Hawkes processes. In order to do so, these authors considered the two counting processes N_t^1 and N_t^2 associated with respectively the arrival times of upward and downward price changes and set:

$$P_t = P_0 + N_t^1 - N_t^2, \quad t > 0$$

where the couple (N_t^1, N_t^2) is a 2-dimensional Hawkes model. Since, in a first approximation, the dynamics of the upward moves and of the downward moves are expected to be the same, it is natural to consider a matrix kernel Φ as the one considered in Eq. (5) with equal diagonal terms $\phi^{(s)}(t)$ and anti-diagonal terms $\phi^{(c)}(t)$. It is well known that, at the microstructure level, the price is essentially mean reverting, thus in [3], the authors considered the "purely mean-reverting scenario", i.e., the case where $\phi^{(s)}(t) = 0$ and chose for $\phi^{(c)}(t)$ an exponential shape. Within this simple framework, Bacry *et al.* provided a closed-form expression for the signature plot. Calibrating the model using MLE or GMM estimations on Euro-Bund and Euro-Bobl future data, they have shown that the model is able to reproduce the scale behavior of the signature plot (see Fig. 7). Bacry *et al.* also considered a natural extension of the previous model to a 4-dimensional Hawkes model in order to describe the joint mid-price dynamics of a pair of assets and to reproduce the Epps effect [3]. Let us recall that, in Ref. [4], the authors established that under general conditions, the empirical covariation of a multivariate Hawkes process converges towards its expected value, that can be easily expressed in terms of the Hawkes covariance matrix $c(t)$ as given by Eq. (23). This result allows one, within the Hawkes price model of Bacry *et al.*, to provide analytical expressions for the signature plot, lead-lag behavior and the Epps effect in terms of Hawkes matrix kernel Φ [3, 4]. Let us also mention that, along the line the of previous model, in [21], the symmetric case was considered, i.e., the "purely trend following scenario": $\phi^{(s)}(t) = 0$ and $\phi^{(c)}(t)$ exponential shape. The authors compared this scenario with the "purely mean-reverting scenario" when used for daily volatility estimation using diffusive formula (53). They showed that the mean-reverting (resp. trend-following) scenario underestimates (resp. over-estimates) the volatility and concluded that a "full" model with both cross and self terms is more realistic and should lead to better volatility estimation. Let us notice that the existence of non-negligible diagonal and anti-diagonal terms has been confirmed by non-parametric estimations performed thereafter in [2, 7, 6].
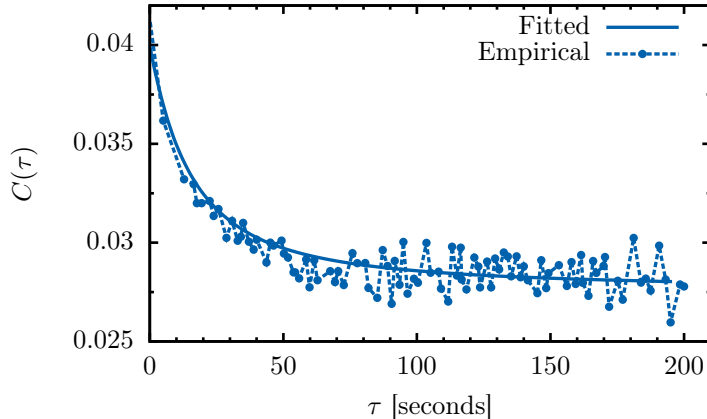
Figure 7: Reproducing the mid-price behevior at miscrostructure level using Bacry *et al.* 2-dimensional Hawkes model. The plots represent the Euro-Bund empirical signature plot as compared to its fit within the model of Bacry *et al.*. The figure has been reproduced from Ref. [3]

One also expects the involved kernels not to be exponential functions. Indeed, as shown in many works, (e.g., in Ref. [2]) and as discussed previously in Sec. 3.1, empirical self-exciting kernels are closer to a power-law than to an exponential. It is with that consideration in mind that Jaisson and Rosenbaum [42] developed the diffusive framework that we have already described in second part of Sec. 2.3.6. It is an alternative framework to the more classical Brownian motion diffusive framework developed in the first part of the same Section. In the second part of their paper, Jaisson and Rosenbaum used their framework to build a version of the 2-dimensional Hawkes model initially introduced by [3] that converges at large scales towards the Heston price model [39] which displays volatility clustering. This work can be seen as the very first step towards a "across scales" unified model that would fit both microstructure stylized facts of price (i.e., point process with strong mean reversion) and "diffusive" stylized facts (volatility clustering and multifractality). In that sense this is a very promising work.

We can also mention a very interesting generalization of the model introduced by Bacry *et al.* in [3]. In [69], Zheng *et al.* introduced a model for the coupled dynamics of the best bid and ask prices. It starts by coding each best price using the Bacry *et al.* model leading to a 4-dimensional price model. The main difficulty comes from the fact that one needs to encode in the model the fact that the ask price needs to lie strictly above the bid price. This is achieved through a spread point process whose dynamics is coupled with the dynamics of both best prices. It is used to measure the distance (in ticks) between the ask price and the bid price: the spread is increased (resp. decreased) by 1 each time either the ask (resp. bid) component jumps upward (resp. downward) or the bid (resp. ask) component jumps downward (resp. upward). Finally a non-linear term is introduced in the Hawkes model: the intensities of the downward

(resp. upward) jumps of the ask (resp. bid) price are set to 0 as soon as the spread process is equal to 1. The authors of [69] developed a whole new rigorous framework for this constrained, non-linear Hawkes model in which they were able to establish several properties (including a diffusive limits). They perform maximum likelihood estimation on real data (using exponential kernels) and show that they were able to reproduce rather well the signature plot.

Let us finally cite the work of Fauth and Tudor in [30] where the authors proposed to describe bid and ask prices of an asset (or a couple of assets) within the framework of marked multivariate Hawkes model. Motivated by the empirical observation (performed on high-frequency Euro/USD and Euro/GPB FX rates from 30-01-2012 to 10-03-2012) that, as the transaction volumes increase, the inter-trade durations decrease, the authors proposed to consider, in addition to an exponential Hawkes kernel, a multiplicative mark (the function χ in Sec. 2.2.1) that corresponds to a power-law function of the volumes: $\chi(v) = Cv^{\nu}$. Their model thus describes the events corresponding to an increase/decrease of the bid/ask as a four dimensional Hawkes process marked by transaction volumes. By adding suitable constraints in order to avoid infinite spread, they calibrated the model on FX rates data by a maximum likelihood approach. Fauth and Tudor have shown that their model is consistent with empirical data by reproducing the signature plots of the considered assets and the behavior of the high-frequency pair correlation function (Epps effect).

5 Impact models

5.1 Market impact modeling

Market impact modeling is a longstanding problem in market microstructure literature and is obviously of great interest both for theoreticians and practitioners (see e.g., [13] for a recent review). While for the former market impact reflects the mechanism enforcing the efficiency of markets, allowing prices to reflect fundamental information, for the latter it represents a cost which needs to be carefully minimized when executing an order. For a trader, market impact induces extra costs per transaction which needs to be added to the fees charged directly by the market, forcing him to split large orders and trade them incrementally in sequences of smaller child orders. Any of such sequences of orders is called a *meta-order*, and quantifying their effect on prices is at the heart of market-microstructure regulation discussions.

The theory of market price formation and the relationship between the order flow and price changes has made significant progress during the last decade thanks to the increasing availability of intraday data [14]. Many empirical studies have provided evidence that the price impact has, to many respects, some universal properties and is the main source of price variations. This corroborates the picture of an "endogenous" nature of price fluctuations that contrasts with the classical scenario according to which an "exogenous" flow of information drives the prices towards a fondamental value [14].

If a meta-order is placed at time $t = 0$ and executed until time $t = T$, an associated market impact curve can be defined by a proxy of the price variation it directly or

indirectly causes[8]. One generally distinguishes two phases: an increasing (concave) part during the execution of the meta-order (i.e., on the interval $[0, T]$), followed by a decaying (generally convex) resilient part. The existence of permanent impact, i.e., a non zero asymptotic value (for large time) of the market impact curve, is a central problem that remains under debate. The typical shape of a market impact curve is shown in Fig. 8, that was was obtained by Bacry *et al.* in [5] by averaging empirical impact curves over a large database of broker meta-orders. Let us point out that this type of measure of market impact cannot be obtained using anonymous market data (i.e, one cannot easily identify the meta-order of a given agent). This, together with the extremely slow intensity of the price signal with respect to its statistical fluctuations, is the reason why empirical results in this respect have been obtained only in relatively recent years.
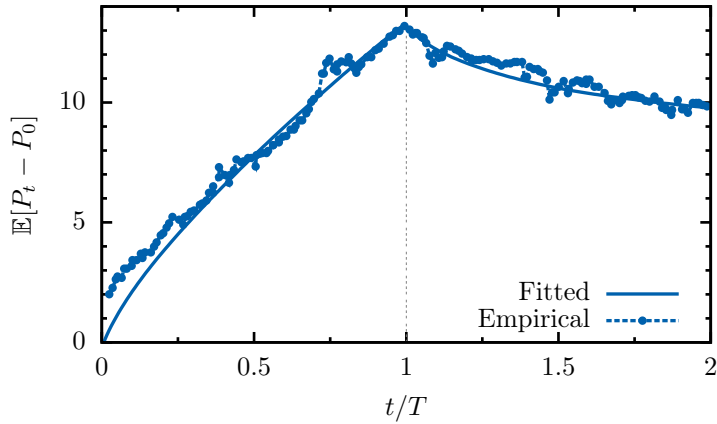
Figure 8: Averaged empirical market impact curves (normalized in time) over a large database of broker meta-orders. The fit is performed using the impulsive HIM model. This figure has been reproduced from [5].

Bacry and Muzy proposed recently a price impact model based on Hawkes processes [7]. They suggested to directly account for the joint dynamics of mid-price and the market order occurences. More precisely the four dimensions of their Hawkes model correspond to the mid-price upward and downward jumps and the buying and selling market order flow (they did not account for the volume of the orders nor the price jump sizes). The matrix kernel Φ can then be decomposed into four sub-blocks of size 2×2. The first one describes the self-excitement of the market order flow, the second one the self-excitement of the price, the third one the (market) impact of the trades on the price while the last one accounts for the feedback influence of price moves on market order flow intensity. By calibrating the model directly from anonymous high-frequency data (using the Wiener-Hopf non-parametric estimation technique described in App. C.2

---

[8]We focus here on the impact of meta-orders, rather than with the impact of individual orders, or with the one of trade imbalance, which have also received considerable interest in the literature.

on the most liquid maturity of EuroStoxx and Euro-Bund future contracts over 800 trading days from 2009 to 2012), the authors of [7] were able to disentangle the self and cross excitation dynamics of mid-price changes from the impact of market orders. In particular, they have shown that the market impact block is mainly diagonal: buying (resp. selling) orders are mostly triggering upward (resp. downward) price moves. Moreover the shape of the diagonal impact function is very localized around $t = 0$. This means that a market order no longer directly impacts the price after a very short delay (i.e. less that $0.1s$). They have also shown that the feedback sub-block is mostly anti-diagonal with slightly negative diagonal kernel functions (see Sec. 2.2.4 for a short discussion on negative kernels). That indicates that an upward (resp. downward) jump in the price tends to increase the intensity of the selling (resp. buying) market order flow and to decrease the intensity of the buying (resp. selling) market order flow. The shape of kernels involved in the self-excitation of the market order flow or the price jumps confirms former estimations performed within lower dimensional models: long-range correlation of the signs of the trades and mainly long-range mean-reversion of the price. All these results were confirmed (using a database with a much higher precision in time) in the work of [6] (see Sec. 6.1). Bacry and Muzy established that, within their framework, it is possible to determine the entire impact profile of some meta-order that is built by accounting for the "bare" direct localized marked impact and by a "dressed" impact that involves mainly the trades power-law self-excitation. They provided analytical expressions for this curve and notably established in both the increasing and the decreasing phase of the impact function, its relationship with the power-law behavior of the self-exciting kernel of market order events. Using the previously described empirical findings, they have shown that one can recover the typical shape depicted in Fig. 8.

In [5], an impact model based on the 2-dimensional Hawkes price model described in Example 1 of Sec. 2.1 (in which only mean-reversion influence has been kept, i.e., $\phi^{(s)}(t) = 0$) has been introduced. It takes into account the impact of an exogeneous (buying) meta-order strategy $r(t)$, where $r(t)$ corresponds to the trading rate of the (buying) strategy per unit of time (so $r(t)\mathrm{d}t$ corresponds to the number of shares bought between time t and $t + \mathrm{d}t$). More precisely, the so-obtained Hawkes Impact Model (HIM) writes

$$\lambda_t^1 = \mu + \phi^{(s)} \star \mathrm{d}N_t^2 + \phi^{(I)} \star f(r_t) \ \text{ and } \ \lambda_t^2 = \mu + \phi^{(s)} \star \mathrm{d}N_t^1 + \phi^{(x)} \star f(r_1), \quad (65)$$

where $\mathrm{d}N_t^1$ (resp. $\mathrm{d}N_t^1$) codes the upward (resp. downward) jumps of the price and $f(r(t))dt$ (with $f(0) = 0$) codes the infinitesimal impact of a buy order of volume $r(t)dt$. The function f corresponds to the *instantaneous* impact function and $\phi^{(I)}$ and $\phi^{(x)}$ correspond respectively to the impact kernel and the cross-impact kernel (this latter describes the impact of a buying order on downward jumps, of course, we expect that $||\phi^{(x)}|| << ||\phi^{(I)}||)$[9]. Following the empirical findings of [7], the impulsive-HIM model corresponds to the particular choice of an impulsive (very localized) impact kernel $\phi^{(I)}(t) = \delta(t)$, i.e., a Dirac distribution. Moreover, as for the choice of $\phi^{(x)}$, the impulsive-HIM model considers that the market reacts to the newly arrived order as if it triggered an upward jump: $\phi^{(x)}(t) = C\frac{\phi^{(s)}(t)}{||\phi^{(s)}||}$. The constant $C > 0$ is a very

---

[9]As for (s) and (c), (I) stands for *impact* and (x) for *cross-impact*.

intuitive parameter that quantifies the ratio of "contrarian" reaction (i.e. impact decay) and of the "herding" reaction (i.e. impact amplification). Analytical formula for the market impact curve were obtained and three cases of interest for C were distinguished in [5]: $C = 0$ corresponds to no contrarian reaction (strong permanent impact), $C = 1$ corresponds to a contrarian reaction as "strong" (in terms of the norm) as the herding one (no permanent impact) and finally $C \in]0, 1[$ corresponds to a contrarian reaction which is not zero but strictly smaller than the herding reaction. Fig. 8 shows a fit of the empirical market impact curve using this model with a power-law microstructure kernel ($\phi^{(s)}(t) \sim t^{-\gamma}$, when $t \to +\infty$), in which case, the market impact curve is proven to be decaying to the permanent market impact value with a power-law $t^{-(\gamma+1)}$.

Let us point out, that in a totally different framework, Jaisson [41], linked, in the asymptotic limit defined by Jaisson and Rosenbaum in [42] (see Sec. 2.3.6), the power-law exponent of the self-excitement kernel involved in the market order flow and the power-law exponent of the market impact decay. He derived his results, within a 2-dimensional Hawkes model (with only self-excitement kernels) for the market order flow, from (i) a price martingale hypothesis and (ii) a linear market impact hypothesis.

5.2 Optimal execution

A natural application of price impact models is to define optimal liquidation strategies. Hewlett [40] was the first to address this problem using Hawkes models. He proposed to model the occurrence of buy and sell market orders on FX markets using a bivariate exponential Hawkes process. He found that these events are mostly self-excited, the cross excitation intensity between buy and sell events being negligible. Using Eq. (47), Hewlett determined the expected future trade imbalance that, within a linear price impact model, allows one to determine the expected future price returns and the associated risk. He then showed that this approach allows one to devise a liquidation strategy that maximises a mean-variance utility function.

Application of Hawkes model for optimal execution has also been considered more recently by Alfonsi and Blanc [1] who modelled the price process using a linear impact of liquidity takers. More precisely, the price is decomposed a the sum of a "fundamental" price which variations are proportional to a fraction of the trade imbalance and a "transient" price that is moved by the remaining fraction of the trade imbalance but with a damping term that represents the market resiliency caused by the market makers behavior. Alfonsi and Blanc have provided explicit expressions of optimal liquidation strategy (i.e. the one with the minimum expected cost) when the flow of buy/sell market orders that impact the price is either a Poisson process or a 2-dimensional Hawkes process with a symmetric exponential kernel matrix. They notably show that price manipulation strategies (i.e. liquidation strategies with negative expected cost) always exist for a Poisson model while they can be excluded in the Hawkes model provided its parameters meet some specific conditions. According the these conditions, the self-excitation should exactly compensate the price resiliency so that resulting price is a martingale. The other condition leads to identify the fraction of endogenous orders within the Hawkes model with the proportion of market orders involved in the transient part of the price behavior.

30

6 Orderbook models

Modeling faithfully the occurrence of various type of orders in the order book with the aim of understanding the mechanisms at the origin of price formation, volatility and liquidity variations is probably the main challenge that the applications Hawkes processes in financial econometrics have to face. Although this goal is far from being reached, some authors have already tackled the problem and made significant progress on these issues.

6.1 Level-I book models

The Level-I book description concerns the events that exclusively occur at the best bid and best ask levels of the order book. Even if this approach discards most of the book information, since best bid or best ask values are directly related to the asset mid-price and market orders mostly impact the book at level-I, one can expect that this level of description is rich enough to capture most of the market features. The study of Biais *et al.* [11] confirms indeed that most of the activity of an order book takes place close to the best quotes, while Cont *et al.* [19] indicate that a substantial part of the dynamics of prices can be accounted for by the evolution best bid and best ask only.

One of the first applications of Hawkes models to order-book modeling at level-I was performed by Large [46] who formalized the concept of book resiliency, namely the ability of the order book to replenish after being depleted by a large trade. Large suggested to quantify resiliency by the way large trades alter future intensities of order occurences. For that purpose he introduced the response kernel $G^{ij}(t)$:

$$G^{ij}(t)\mathrm{d}t = \mathbb{E}\left[\mathrm{d}N_t^i \left| \mathcal{F}_0, \mathrm{d}N_0^j = 1\right.\right] - \mathbb{E}\left[\mathrm{d}N_t^i \left| \mathcal{F}_0\right.\right]. \tag{66}$$

$G^{ij}(t)$ simply describes the increase of the future conditional intensity of events of type i caused, directly or indirectly, by the the occurence, at time $t = 0$ of an event of type j. In the case of a Hawkes process of kernel matrix Φ, Large proved that G^{ij} satisfies the integral equation in Eq. (17) defining $\Psi(t)$. In other words, the matrix $\Psi(t)$ can be interpreted as the increase of the expected number of events after a lag t caused by the occurrence of some event at time 0. Large then considered order book data from LSE, modelled as a 10-dimensional Hawkes processes where the book events are classified according to whether they move or not the mid-price: market and limit orders that move the mid-prices (4 components if one distinguishes bid and ask), market and limit orders that leave the book unchanged (4 components) and cancel orders (2 components). The impact of "aggressive" orders on the rate of forthcoming events is then studied. The processed data consisted in LSE stock data (Barclays equity) timestamped at the resolution of $1s$ during the 22 trading days of January 2002. Large used MLE estimation within the class of exponential kernels in order to estimate Ψ. His results allowed him to provide a "causal" interpretation (Large prefers the term "precipitation" than "cause") of the main event occurrence in the book. He mainly found that aggressive limit orders are principally caused by aggressive market orders, measuring thereby the market resiliency in all his aspects, magnitude, trade direction and characteristic time. Consistently with former studies, Large estimated

that the studied stock value is resilient less than 40 % of cases and, when it is the case, the book replenishment occurs within a time frame of around $20s$. He also shown that market order dynamics is mostly self-excited and correlated over a large time. Agressive market orders are also triggered by aggressive limit orders as a consequence of the "race to liquidity".

A similar analysis of level-I order book data was recently conducted by Bacry *et al.* [6]. These authors made a slightly different categorization than Large and distinguished all book events (market, limit and cancel orders) that leave the mid-price unchanged (6 components accounting for bid and ask sides) from events that move the mid-price up or down (2 components). The dynamics of these event occurrence has then been modeled as a 8-dimensional Hawkes process. Unlike Large, Bacry *et al.* performed a non-parametric estimation of the matrix of kernels using the method described in Sec. C.2. They considered book data time-stamped at a time resolution of $10^{-6}s$ and the analysis has been performed up to lags of a few minutes. The event dynamics has thus been considered over a range of time scales close to eight decades. The main result reported by the authors is that the book event dynamics is mainly self-exciting, except for the mid-price changes for which cross-excitation effects are strongly dominating. This is illustrated in Fig. 9 where the values of the estimated kernel norms are reported using a color map: one can see that the resulting matrix is mainly diagonal except in the price sub-block, that is anti-diagonal. The observation
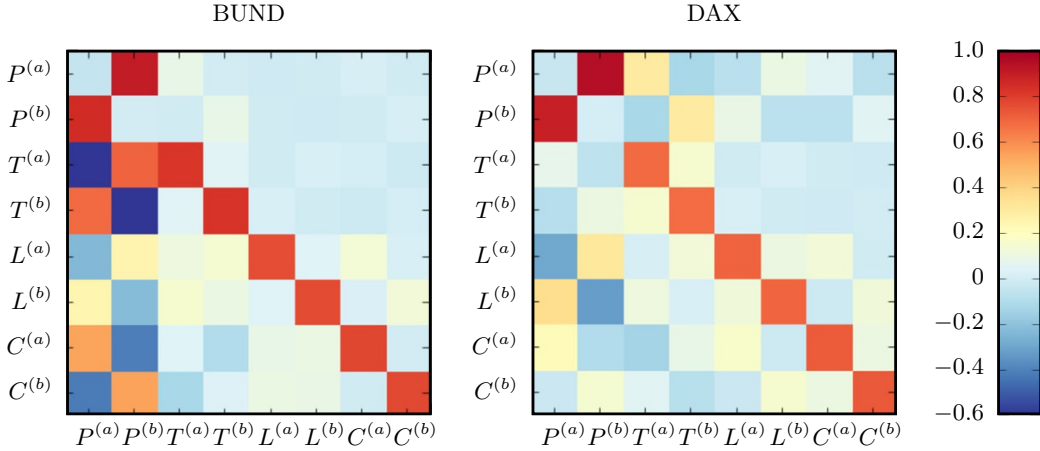
Figure 9: Empirical determination of the matrix of Hawkes kernel norms in the 8-dimensional model of level-I book events. P stands for mid-price change events, T for trade events, L for limit order events and C for cancel events. The superscripts (a) (b) indicates the direction, ask or bid of the events. This figure is reproduced from [6].

that price changes events are mainly triggered by price change events is in agreement of Large previously reported results. Moreover, Bacry *et al.* observations confirmed the previously reported empirical fact that this triggering effect is mostly anti-diagonal, i.e., present price changes impact future price changes in the opposite direction. Previous findings using low-dimensional model concerning the kernel shapes were also confirmed: the market is highly endogenous, whatever the type of event one considers and all the dominating kernels (the diagonal kernels for orders leaving the price unchanged and the anti-diagonal kernels for mid-price moves) are slowly decreasing, well described by a power-law behavior as in Eq. (63). The richness of the Hawkes model of [6] allowed the authors to account for a richer dynamical behavior than previous works and to describe and quantify the high-frequency influences between all types of events. They notably characterized the impact of price changes on the book event flow, a quantity that turns out to be very sensitive to the asset tick size (as estimated by the probability that the mid-price has to move). They also provided evidence of some inhibitory effects which result from negative values of some Hawkes kernels. For example it was observed that, for a large tick asset (like the Euro-Bund futures), an upward price move not only triggers forthcoming trades at bid but also inhibits trades on the ask side.

Restricting previous Large model to market order flows at best bid and best ask, Muni-Toke and Pomponio [56] studied the dynamics of trade-through orders. A market order is a trade-through if part of it is executed at the next best limit. Thus, for that purpose, they used the 2-dimensional Hawkes process (with exponential kernels) decribed by Eq. (53). Though very rare for some assets (e.g., Euro-Bund future), trade-through can be quite frequent on other contracts. For instance on the BNP stock [56] finds an average of 400 trade-throughs per day. Parametric estimation using MLE on exponential kernels have been performed on Euronext stocks restricting intraday-time to 9h30 to 11h30 am to avoid very strong intraday seasonality effects. Goodness of fit tests confirm that, as for the full market order flow, the main components of the 2-dimensional Hawkes kernel are the self-exciting kernel.

6.2   Full order book model

In [55], Muni Toke generalized the zero-intelligence model introduced in [64] to the Hawkes framework. While in the latter model the authors built a full model for the order book in which all the involved flows (i.e., limit and market orders at any level) are independent pure Poisson processes. It is clearly a very rough approximation, market orders are known to be long-range dependent (see Sec. 3.1) due to the splitting of large meta-orders. Moreover, one expect limit orders to be also highly auto-correlated as well as correlated with the market-order flow due to the fact market makers interact with market takers. In [55], a 2-agents model is introduced. It comprises the following agents:

- A *liquidity provider*:
  - Arrivals of limit orders are modeled by using either a pure homogeneous Poisson process or a 1-dimensional Hawkes process with an exponential kernel (cancellations are treated as in the zero-intelligence model, i.e., each order has a life-time which is an i.i.d. exponential random variable using whose parameter is fixed a priori)

– The price-levels of new limit orders are randomly chosen by first choosing the (bid or ask) side with probability 1/2 and then by sampling their values from a Student distribution
- A *liquidity taker*:
  – Modeled using either a pure homogeneous Poisson process or a 1-dimensional Hawkes process with an exponential kernel.

All the volumes of both limit and market orders are i.i.d. and exponentially distributed variables. Hawkes processes use exponential kernels which are estimated parametrically using Maximum Likelihood Estimation.

Not surprisingly, inter-trade times of market orders are shown to be much better modeled by a 1-dimensional Hawkes process (referred to as MM since the only Hawkes kernel involved is one which deals with the influence of Market orders on themselves) than by a pure Homogeneous Poisson process (HP). In the same way, inter-time of limit orders are shown to be much better modeled by a 1-dimensional Hawkes process (LL) than by HP. Along the same lines, the left panel of Fig. 10 shows that inter-time between a given market order and the first arrival time of a limit order after that market order is much better reproduced by a 2-dimensional Hawkes (referred to as MM+LL+LM) model with self-exciting kernels for both market (MM) and limit orders (LL) and a single cross-exciting kernel which corresponds to the influence of past market orders on future limit orders (LM). The figure shows that a model with two independent Hawkes models (MM+LL) with no cross-exciting kernels (LM) does not perform well. In [55], Muni-Toke claims that the other cross-exciting kernel (the one describing the influence of past limit orders on future market orders) is negligible. Thus the order book dynamics seems to be driven mainly by the liquidity taker agent rather than the liquidity provider, i.e., in a first approximation, the market makers strategy consists basically in reacting to liquidity takers, whereas liquidity takers take decisions independently from market makers.

Finally the right panel of Fig. 10 shows the so-obtained distribution of the bid-ask spread for empirical data, the pure Homogenous Poisson (HP) model, the MM+LL Hawkes model and the MM+LL+LM model. Again, this latter model is the one which best fits the empirical data.

Let us mention a theoretical work by Jedidi and Abergel [43] in which a Hawkes-based markovian framework for the whole orderbook dynamics is studied.

7 Other models

7.1 Systemic risk models

Hawkes models have been also be used at coarser time scales where asset prices are mainly diffusive. They can be used in mixed models in order, for instance, to account for jumps that occur over the diffusion process. This is the spirit of the model developed by Aït-Sahalia *et al.* [63] that proposed to describe the contagion of a crisis across all the world markets by superimposing a multivariate self-excited Hawkes process to a standard multivariate continuous diffusion model. According to this model, called be
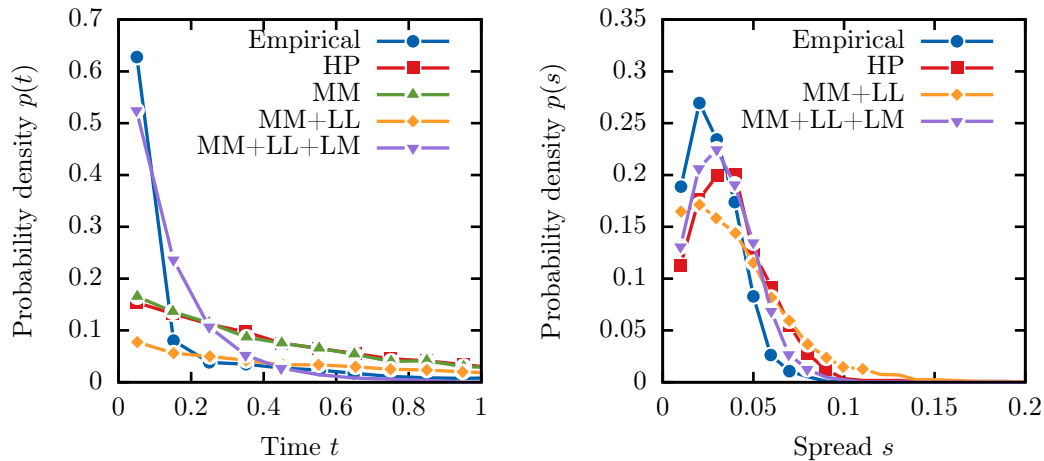
Figure 10: (Left panel) Empirical density function of the distribution of the duration between a given market order and the first arrival of a limit order after that market order. The density function is displayed for empirical BNPP time-series, and for three different models that were fitted (using maximum likelihood) on these data, namely: a purely Homogenous Poisson model (HP) for both market and limit orders, a model with a pure poisson process for limit orders and a 1-dimensional Hawkes process for the market orders (MM), a model with two independent 1-dimensional Hawkes processes for limit and market orders (LL+MM) and finally this last model with a cross-exciting term which characterize the influence of past market orders on future limit orders (LM). (Right panel) Empirical density function of the distribution of the bid-ask spread. The density function is displayed for empirical BNPP time-series, and for three different models that were fitted (using maximum likelihood) on these data, namely: a purely Homogenous Poisson model (HP) for both market and limit orders, a model with a pure poisson process for limit orders and a 1-dimensional Hawkes process for the market orders (MM), a model with two independent 1-dimensional Hawkes processes for limit and market orders (LL+MM) and finally this last model with a cross-exciting term which characterize the influence of past market orders on future limit orders (LM). These figures are reproduced from Ref. [55].

the authors, "Mutually exciting Jump-Diffusion", the log-price vector satisfies:

$$\mathrm{d}X_t = \mu_t + \sigma_t \mathrm{d}W_t + Z_t \mathrm{d}N_t \tag{67}$$

where W_t is a D-dimensional Brownian motion, σ_t is a stochastic volatility and N_t is a D-variate Hawkes process that accounts for the self-excited nature of price jumps occurrence (Z_t is a random variable accounting for the direction and the intensity of the jump). A GMM estimation procedure is proposed in [63] based on a closed-form expressions of some moments associated with the returns variations in the monovariate and bivariate cases with exponential Hawkes kernels. This estimation has been applied to five international equity indices data associated with respectively US, Europe, Asian, Pacian and Latin America zones. The authors found that the jumps terms have significant self-excitation components and as far as the "contagion" effect is concerned, it seems that the US equities are the ones with the greatest influence on other markets.

In Ref. [29], Errais *et al.* proposed to model credit default events in a portfolio of securities as correlated point processes. More specifically they considered that the dynamics of such events is described by a marked Hawkes process with exponential kernels. A stated by Prop. 2, in this case, the couple of processes (λ, N) is a Markov process. Thanks to the Dynkin formula, the authors provided explicit expressions for the conditional distribution of both the marked and counting processes. This result was used to price portfolios if credit derivatives such as index and tranch swaps. Their model was then calibrated from index and swap data during September 2008 that witnessed several credit default notably Lehman-Brothers default. The authors have shown that by, capturing the dependence of default events, their model provided good fits of market data, unlike standard approaches that failed during this period. Errais *et al.* emphasized that marked Hawkes models with exponential kernels belong to the more general class of affine point processes introduced by Duffie *et al.* [26]. An affine point process involves a stochastic intensity vector which is a Markov process with drift, diffusion and jump terms. Errais *et al.* [29] have shown that their approach remains tractable within the class of affine point processes that provides a richer jump interaction structure.

Let us also quote the recent similar work of Dassios and Zhao [23] who addressed the question of default risk modelling and contagion propagation within the framework of the so-called "Dynamic Contagion Process" that is an exponential marked Hawkes model but where the Poisson exogenous events (the "immigrants") are replaced by a shot-noise (constructed as the first generation of the same exponential Hawkes model with a different mark probability density). The authors have established the theoretical distributional properties of this new process (that remains a Markov process) and provided analytic expressions for the its probability generating function. As in Errais *et al.* [29], they showed that their approach is particularly suitable for modelling the dependence structure of arriving events with dynamic contagion and proposed an application to credit risk,

7.2   Accounting for news

In [61], Rambaldi *et al.* used Hawkes processes for modeling the impact of news on the EBS (foreign exchange market) quotation time-series, consisting in a list of quotation

timestamps. Let us note that, since EBS data are throttled (aggregated below a window of $0.1s$), a randomization procedure has been used for homogenization. The model for the quotation arrival times is a 1-dimensional Hawkes model involving either a double exponential or a power-law (expressed as a sum of 15 exponentials so to have a convenient Maximum Likelihood Estimation procedure). The impact of the news (on the quotation time-series) can be seen as particular instances of localized non stationarities. Thus, Rambaldi *et al.* introduced an exogenous kernel $\phi^{(news)}$ (an exponential function) that accounts for the impact of a particular instance of a news, leading to the model [10]:

$$\lambda_t = \mu + \phi \star dN_t + \phi^{(n)}(t - t_0), \tag{68}$$

where t_0 is the time of occurence of a particular news. Let us point out that $\phi^{(n)}$ (where (n) stands for *news*) is allowed to have a non-causal component (i.e., $\phi^{(n)}(t)$ is non zero for $t < 0$) in order to account for anticipation effects. Estimation is performed (using regular Maximum Likelihood Estiamtion procedure) on a 3 hour period around the considered news. Though the results are very noisy, the authors showed that the model captures nicely both the amplitude and the time scale of the news effect. The distribution of the L^1 norm $||\phi^{(n)}||$ for the different news has broad distribution (that goes beyond 1) that clearly reflects the diverse effects of news on the market. Moreover, using some proxies for quantifying how unexpected a particular news is, Rambaldi *et al.* clearly showed that the norm $||\phi^{(n)}||$ is not only related to the news impact but also to its degree of surprise.

7.3 High-dimensional models

Modeling co-jumps. In [12], Bormetti *et al.* modeled the complex dynamics of the extreme returns in a basket of stocks. They started by elaborating a rather elaborate procedure for identifying extreme returns (referred to as "jumps" in the paper, corresponding to anomalous values) from 1-minute returns time-series. On a basket of $N = 20$ Italian stocks they found up to 280 jumps on a single stock on the overall period (of 88 days) and up to a total of 505 jumps per day on all the stocks. They clearly established that some jumps arrive at the same time, i.e., within the same time-window. This prevents from modeling the overall process as an N-dimensional Hawkes process: Within such a framework, co-jumps cannot occur without introducing an impulsive component inside the kernel Φ. Bormetti *et al.* finally suggested the following model

- A 1-dimensional Hawkes process is used for modeling the co-jumps arrival times. The kernel of the Hawkes process is chosen to be a sum of exponential functions.
- Each time t a co-jump occurs, each stock i, independently one form each other, has a probability p_i to jump. These probabilities are estimated empirically from real data.
- For each stock i, a 1-dimensional Hawkes process is used to model the idiosyncratic jumps of that stock, i.e., the jumps that are do not correspond to co-jumps.

---

[10] Formally, this model can be seen as a 1-dimensional version of the 2-dimensional price impact model of Eq. (65) in which $f(r_t)$ is replaced by the Dirac distribution centered at time t_0, $\delta(t - t_0)$.

Again, the kernels of the Hawkes process is chosen to be a sum of exponential functions.

In their paper, Bormetti *et al.* developed a precise procedure to perform estimation of this model which is shown to be quite robust. Moreover they showed that it is able to capture simultaneously the time clustering of jumps and the high synchronization of jumps across assets.

Clustering with graph models. In [49], Linderman and Adams developed a probabilistic model that combines Hawkes processes with random graphs models, that they applied on S&P100 data. Each component of the Hawkes kernel codes the changes (of more than 0.1%) in the last traded price of a given asset during a whole week. Thus a 100-dimensional point process with 182.037 events is obtained. The kernels are chosen as

$$\phi^{ij}(t) = A^{ij} W^{ij} h_{\theta_{ij}}(t), \tag{69}$$

where A is a random binary (0 or 1) valued matrix, W a random matrix with positive entries and $h_\theta(t)$ a parametric kernel (of parameters θ) such that $\int h_\theta = 1$ (a logistic normal density with two parameters is chosen). Intraday seasonality is modeled using an exogenous intensity of the form $\mu = m + \nu e^{y(t)}$ where ν is a constant matrix and $y(t)$ is univariate Gaussian process. The random graph model is used to reflect the probability of the different network structures through the prior distributions of the matrices A and W. They basically depend on a latent distance which is chosen in $\mathbb{R}^2$ imposing an overall sparsity (20%) and a characteristic distance scale. Thus each stock k corresponds to a latent coordinate x^k in $\mathbb{R}^2$ that is estimated through a fully-Bayesian, parallel inference algorithm. A figure is displayed where each stock is represented as a dot in the latent 2-dimensional space with a color coding corresponding to its corresponding sector (among six sectors). Linderman and Adams showed that some sectors, notably energy and financials, tend to cluster together, indicating an increased probability of interaction between stocks in the same sector. Other sectors, such as consumer goods, are broadly distributed, suggesting that these stocks are less influenced by others in their sector. One can think that this approach will pave the way, with the framework of multivariate Hawkes processes, to very promizing applications where one will process large amounts of data associated with a great number of interacting components/agents in order to shed new light on the complexity of financial markets.

Finally, let us note that Mastromatteo and Marsili [52] tried to overcome the problem of the estimation of a high-dimensional Hawkes process by mapping it onto a graphical (Ising) model, so to describe the clustering of trading times in a set stocks. They reconstructed the interaction network of the $D = 100$ most traded stocks of the NYSE during the year 2003 by using machine learning techniques, and reported an overall scaling of the weight matrix $W \propto D^{-1}$. Their study evidences the presence of a large collective market mode of the matrix A, driving the overall level of market activity close to the critical point.

Appendices

A   Table of financial applications found within each discussed paper

Each of the academic works discussed throughout our paper involving numerical experiments on financial data is listed in the following table. Apart from the comlumn names that are explicit: D stands for dimension of the Hawkes model, T for the length of the historical data used to calibrate the model, $\mathrm{d}t$ for the time resolution of the data, $\Phi(t)$ for the shape of the Hawkes kernel. N-Exp stands for a sum of N exponential, PL for "power-law " and NP for "non-parametric".

				Fitted Hawkes				
		D	N_t	Contracts	T	dt	$\Phi(t)$	Comments
Section 2								
[15]	Bowsher	1-2	Trade	Stocks	2 months	1s	2-Exp	+ seasonality
[18]	Chavez *et al.*	1	Extreme ΔP	Stocks	1 year	1ms	2-Exp	Marked
[42]	Jaisson *et al.*	8	Level-I	Futures	1 year	1μs	NP	
Section 3								
[2]	Bacry *et al.*	1	Trade	Futures	3.5 months	1ms	NP	
		2	Price	Futures	3.5 months	1ms	NP	
[8]	Bacry *et al.*	1	Trade	Futures	4.5 years	1ms	NP	
[21]	Da Fonseca *et al.*	1	Trade	Stocks Futures	2 years	1ms	Exp	
		2	Mid-Price	Stocks Futures	2 years	1ms	Exp	
[27]	Embrechts *et al.*	2	Extreme ΔP	Indices	17 years	1day	Exp	Marked
		1	Extreme ΔP	Indices	14 years	1h	Exp	3d-Marked
[31]	Filimonov *et al.*	1	Mid-Price	Futures	13 years	1s	Exp	
[32]	Filimonov *et al.*	1	Mid-Price	Futures	15 years	1s	Exp	
		1	Mid-Price	Futures	15 years	1s	PL	
[34]	Hardiman *et al.*	1	Mid-Price	Futures	14 years	1ms	N-Exp	
		1	Mid-Price	Futures	1 year	1ms	NP	
[35]	Hardiman *et al.*	1	Mid-Price	Futures	16 years	1s	Exp	
		1	Mid-Price	Futures	16 years	1ms	NP	
[45]	Lallouache *et al.*	1	Trade	FX	3 months	0.1s	2-Exp	
Section 4								
[3]	Bacry *et al.*	2	Price	Futures	2 hours	1ms	Exp	
[30]	Fauth *et al.*	4	Trades	FX-Rates	2 months	1ms	Exp	Marked
[69]	Zheng *et al.*	2	Trade	FX	10 days	0.1s	Exp	Constrained
Section 5								
[7]	Bacry *et al.*	4	Price+Trade	Futures	3.75 years	1ms	NP	
[5]	Bacry *et al.*	2	Price	Stocks	1 year	1ms	PL	
[40]	Hewlett	2	Trades	FX	2 months	1s	Exp	
Section 6								
[46]	Large	10	Level-I	Stocks	22 days	1s	Exp	
[56]	Muni Toke *et al.*	2	Trade-through	Stocks	5 months	1ms	Exp	
[55]	Muni Toke	2	Orderbook	Stocks	15 days	1ms	Exp	
Section 7								
[12]	Bormetti *et al.*	1×20	Price	Stocks	88 days	1 mn	Exp	Co-jump
[29]	Errais *et al.*	1	Default	Credit	21 days	1day	Exp	
[49]	Linderman *et al.*	100	Extreme ΔP	Stocks	1 week	1s	Exp	Fixed decay Random Graph
[52]	Mastromatteo *et al.*	100	Trades	Stocks	1 year	1s	Exp	
[61]	Rambaldi *et al.*	1	Quotation	FX	1 year	0.1s	N-Exp	News
[63]	Sahalia *et al.*	1-2	Price	Indices	[22, 33] years	1day	Exp	Contagion

B Simulation

There are two main frameworks for simulating Hawkes processes: an intensity-based framework and a cluster-based framework, depending on whether it makes direct use of the intensity regression Eq. (2) or of the cluster representation of the Hawkes process (see Sec. 2.3.7). The most popular method being certainly the intensity-based thinning method initially introduced by Lewis [48] in the general context of non homogenous Poisson processes, and modified later by Ogata [58].

Thinning. The thinning algorithm is an incremental procedure in which the successive jumping times are generated sequentially. It can be easily adapted to the multi-dimensional context and can be generalized to the marked case. In its simplest version (valid for decreasing kernels), given a starting time t it basically consists in sampling a candidate jumping time $t + \Delta t$ using an exponentially distributed random variable Δt with parameter $\lambda_t^{(tot)} = \sum_{k=1}^{D} \lambda_t^k$. Then a random variable U is uniformly drawn in the interval $[0, \lambda_t^{(tot)}]$. If $U < \lambda_t^{(tot)} - \lambda_{t+\Delta t}^{(tot)}$, the jump is rejected. If not rejected, the jump is assigned to the component i, where i is the largest index satisfying $U \geq \lambda_t - \sum_{k=1}^{i-1} \lambda_{t+\Delta t}^k$. Let us point out that the complexity of the algorithm can be substantially reduced in the case of exponential kernels.

Time-change Another common intensity-based approach for simulating a non homogenous Poisson process N_t uses the following well known fact: if one defines the cumulative intensity function $F_t = \int_0^t \lambda_u du$, then $N_{F_t^{-1}}$ is an homogenous Poisson process of intensity 1. Thus in order to perform simulation one needs to know how to simulate $F_{t_{m+1}}^{-1} - F_{t_m}^{-1}$ where $t_{m+1} - t_m$ is exponentially distributed. This algorithm has been applied for instance in the case of exponential kernels in [24]. In fact, in this simpler case it is possible to invert analytically the function F.

Cluster algorithm The cluster approach consists basically in simulating the branching structure described in Sec. 2.3.7. Hence, the algorithm is sequential in the generations index n rather then in real time t. One can see for instance [54] which describes how one can perform simulation of marked Hawkes processes using the branching structure. A short survey and a more comprehensive list of references can be found in the recent work [24].

C Statistical inference

C.1 Parametric estimation

Maximum Likelihood Estimation. The most commonly used technique for parametric estimation of Hawkes processes is the Maximum Likelihood Estimator

(MLE), which has been first introduced by Ogata in [60]. The log-likelihood of a non-homogeneous, multi-dimensional Poisson defined as in Eq. (2) reads

$$\log \mathcal{L}(\mu, \Phi) = \sum_{i=1}^{D} \int_0^T \ln(\lambda_t^i) \mathrm{d}N_t^i - \sum_{m=1}^{M} \ln \lambda_{t_m}^{k_m}, \tag{70}$$

where the couples $\{(t_m, k_m)\}_{m=1}^{M}$ denote respectively event times and components. Given a Hawkes kernel Φ_θ depending upon a set of parameters θ, it is possible to estimate it from Eq. (70) by solving the problem

$$(\mu^*, \theta^*) = \operatorname*{argmax}_{(\mu, \theta)} \log \mathcal{L}(\mu, \Phi_\theta). \tag{71}$$

In the case of a general Hawkes process, the computation of the likelihood (or its gradient) is of the order of $O(M^2 D)$ where M is the total number of jumps. The fact that it reduces to $O(MD)$ when the kernels are exponential functions is one of the major reasons why exponential kernels (or sum of exponential kernels) are so commonly used in a parametric framework [27, 55, 56, 12, 61, 69].

EM based Estimation. In [67], the authors use an Expectation Maximization (EM) based technique, an iterative procedure comprising the alternation of two steps of *expectation* (E) and *maximization* (M), in order to exploit the cluster representation of the Hawkes process described in Sec. 2.3.7. Given a paramterization of the Hawkes kernel Φ_θ, and a set of parameters (μ, θ), each iteration consists in:

- **E-step** $(\mu, \theta) \to \{p_{mm'}\}$: Estimating, for all ordered pair of jumps $(t_m, t_{m'})$ (with $t_m < t_{m'}$), the probability $p_{mm'}$ that, within the cluster framework described in Sec. 2.3.7, t_m is an ancestor of $t_{m'}$.
- **M-step** $\{p_{mm'}\} \to (\mu, \theta)$: Computing the model parameters (μ, θ) from the probabilities $\{p_{mm'}\}$ estimated in the E-step.

This algorithm converges very quickly when the product of the average intensity Λ and the characteristic timescale τ of the support of Φ is small, i.e., $\Lambda\tau \ll 1$, meaning that few jumps are detected in intervals of size τ.

Let us point out that General Method of Moments (GMM) can be also used [21], based, for instance, on auto-covariance analytical formula such as (29).

High-dimensional estimation. All the previous techniques cannot be applied in a large dimensional context (e.g., $D \geq 100$) without substantial modifications. For instance, even in the case of "simple" exponential kernels, the number of parameters is of the order of D^2, so that applications to contexts in which D is large become unfeasable due to overfitting and/or computational issues. In order to address the issue of parametric estimation in large dimensions, one needs to use algorithms which involve regularization. Within the last year, Hawkes processes in large dimension have been

the subject of numerous academical papers (see for instance [70]). Most of them, adapt more or less "classical" convex optimization techniques in the framework of Hawkes processes with exponential kernels of the form $\alpha^{ij}\beta^{ij}e^{-\beta^{ij}t}$. Let us point out that, in all these papers, in order to get a convex log-likelihood (70), the parameters β^{ij} are a priori fixed (i.e., they are not estimated), and generally not chosen to depend on i or j. Thus, in that case, the kernel matrix Φ can be written in matrix form as $\Phi = \alpha\beta e^{-\beta t}$, where α is the so-called (weighted) adjacency matrix $\alpha = \{\alpha^{ij}\}_{1\leq i,j\leq D}$. α describes the "connections" between the different components of the Hawkes process. In order to solve the so-obtained convex-optimization of the log-likelihood, penalizations terms on α are customarily introduced. They are generally of two sorts: an L^1 penalization that ensures sparsity and a trace-norm penalization that ensures low-rank. To the best of our knowledge, these types of algorithms have only been used once in the context of finance (see Sec. 7.3 about [49]). Clearly, one can easily forecast that there will be, in the near future, a huge number of new algorithms for parametric estimation in large dimensions, some of them, improving the state of the art.

C.2   Non-parametric estimation

Very few non-parametric estimators of the kernel matrix of a Hawkes process have appeared in the academic literature so far. We review them in the next section.

EM based estimation. Historically, the first one [47], corresponds to a non-parametric version of the EM estimation algorithm described in the previous section. It is based on regularization (via L^2 penalization) of the method initially introduced by [50] in the framework of ETAS model for seismology. It has been developed for 1-dimensional Hawkes processes. The maximum likelihood estimator is computed using the same two steps as in the parametric case. The E-step corresponds to estimating all the $p_{mm'}$ probabilities and the M-step corresponds to estimating $\phi(t)$ and μ from these probabilities. In [47], some numerical experiments on particular cases are performed successfully even when the exogenous intensity μ depends slowly on time: the whole function $\mu(t)$ is estimated along with the kernel $\phi(t)$ with a very good approximation. However, as explained in [8], this method has two main drawbacks:
- the convergence speed of the EM algorithm drastically decreases when the decay speed of the kernel $\phi(t)$ is low (e.g., power-law decaying kernel),
- the probabilistic interpretation of the kernel values involved in the EM method prevents the kernels to have negative values (see Sec. 2.2.4).

Contrast function based estimation. In a recent series of papers [62, 33], some authors proposed, within a rigourous statistical framework, a second approach for non-parametric estimation. It relies on the minimization of the so-called L^2 *contrast function*. Given a realization of a Hawkes process $\tilde{N}_t$ on an interval $[0, T]$, associated

with the parameters $(\tilde{\mu}, \tilde{\Phi}(t))$, the estimation is based on minimizing the *contrast function* $C(\mu, \Phi)$:

$$(\mu^*, \Phi^*) = \underset{(\mu, \Phi)}{\operatorname{argmin}}\, C(\mu, \Phi), \tag{72}$$

where

$$C(\mu, \Phi) = \left(\sum_{i=1}^{D} \int_0^T \lambda_t^i(\mu, \Phi)^2 \mathrm{d}t - 2 \int_0^T \lambda_t^i(\mu, \Phi) \mathrm{d}\tilde{N}_t^i \right) \tag{73}$$

and

$$\lambda_t^i(\mu, \Phi) = \mu^i + \sum_{j=1}^{D} \int_{-\infty}^{t} \Phi^{ij}(t-s) \mathrm{d}\tilde{N}_s^j \ . \tag{74}$$

Let us point out that, minimizing the expectancy of the contrast function is equivalent to minimizing the L^2 error on the intensity process. Indeed, if $\mathcal{F}_t$ is the information available up to time t^-, since $\mathbb{E}\left[\mathrm{d}\tilde{N}_t^i \,|\, \mathcal{F}_t \right] = \tilde{\lambda}_t^i \mathrm{d}t$, one has

$$\underset{(\mu, \Phi)}{\operatorname{argmin}} \mathbb{E}\left[C(\mu, \Phi) \right] = \underset{(\mu, \Phi)}{\operatorname{argmin}} \sum_{i=1}^{D} \left(\mathbb{E}\left[\lambda^i(\mu, \Phi)^2 \right] - \mathbb{E}\left[\lambda^i(\mu, \Phi)\tilde{\lambda}^i \right] \right) \tag{75}$$

$$= \underset{(\mu, \Phi)}{\operatorname{argmin}} \sum_{i=1}^{D} \mathbb{E}\left[(\lambda^i(\mu, \Phi) - \tilde{\lambda}^i)^2 \right]. \tag{76}$$

The minimum value (zero) is of course uniquely reached for $\Phi = \tilde{\Phi}$ and $\mu = \tilde{\mu}$.

In [33], the authors chose to decompose Φ on a finite dimensional-space (in practice, the space of the constant piece-wise functions) and to solve directly the minimization problem (72) in that space. For that purpose, in order to regularize the solution (they are essentially working with some applications in mind for which only a small amount of data is available, and for which the kernels are known to be well localized), they chose to penalize the minimization with a Lasso term (which is well known to induce sparsity in the kernels), i.e., the L^1 norm of the components of Φ. Let us point out, that minimizing the contrast function and minimizing the expectancy of the contrast function are two different stories. The contrast function is stochastic, and nothing guarantees that the associated linear equation is not ill-conditioned. In [33], the authors prove that, under certain conditions on Φ, the linear equation is invertible, i.e., the associated random Gram matrix is almost surely positive definite. In practice (they study real signals from neurobiology), they choose the components of Φ to be piece-wise constant.

The Wiener-Hopf approach. Let us point out that, since λ_t^i is expressed linearly in terms of μ^i and of the $\{\phi^{ij}(t)\}_{ij=1}^{D}$, minimizing the L^2 error (75) is equivalent to solving a linear equation, which is nothing but the Wiener Hopf Eq. (38). Consequently, minimizing the expectancy of the contrast function is equivalent to solving the Wiener-Hopf equation. In [7, 8], the authors chose to perform non-parametric

estimation by directly inverting this system. They have proved that, as long as $g(t)$ corresponds to a conditional intensity of a Hawkes process, (38) has a unique solution in $\chi(t)$ which is the kernel matrix $\Phi(t)$. The algorithm uses quadrature technique in order to discretize the system. It can be summarized in the following way:

- Estimation of the vector Λ (simply using as an estimator of Λ^i the number of jumps of the realization of N^i divided by the overall time realization)
- Non-parametric estimation of the matrix $g(t)$ defined by (36) (using kernel density estimation techniques)
- Fix the number of quadrature points (e.g., gaussian quadrature) to be used as well as the support for all the kernel functions.
- Disretize the Wiener-Hopf system on the quadrature points and inverse it. This leads to an estimation of the kernel functions on the quadrature points. Estimation on a finer grid as well as L^1 norm can be obtained through simple quadrature formula.
- Estimation of the exogenous intensity μ using (21).

We refer the reader to [8] where the full algorithm (including the methodology for choosing the bandwdith value for density kernel estimation as well as the number of quadrature points) is described precisely along with the comparisons with the other methods presented above. It has actually been improved in [6] to take particular care of slowly decreasing kernels.

As compared to the approach developed in [33], this algorithm is better adapted to the case large amount of data is available. As compared to the EM approach, we shall advocate its use over the EM algorithm mainly in two cases (which are often statisfied when dealing with finacial data)

- either the kernel functions are not localized (e.g., power-law). Indeed, in that case the EM algorithm is known to be very slow to converge (see [47] and [8]),
- or some of the kernel functions have negative values (see second point on the EM algorithm above). We refer the reader to [33] for justification why Wiener-Hopf approach allows to deal with negative kernels.

Finally, let us point out that, in the particular case the kernel matrix is known to be symmetric (which is always true if the dimension $D = 1$), the method developed in [2] uses a spectral method for inverting (23) and deduces an estimation from the second-order statistics. It can be seen as a particularly elegant way of solving the Wiener-Hopf equation.

Acknowledgements

References

[1] A. Alfonsi and P. Blanc. Dynamic optimal execution in a mixed-market-impact hawkes price model. *arXiv preprint arXiv:1404.0648*, 2014.

[2] E. Bacry, K. Dayri, and J.F. Muzy. Non-parametric kernel estimation for symmetric hawkes processes. application to high frequency financial data. *Eur. Phys. J. B*, 85:157, 2012.

[3] E. Bacry, S. Delattre, M. Hoffmann, and J.F. Muzy. Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, 13:65–77, 2013.

[4] E Bacry, S Delattre, M. Hoffmann, and J.F. Muzy. Some limit theorems for hawkes processes and application to financial statistics. *Stochastic Processes and their Applications*, 123(7):2475–2499, 2013.

[5] E. Bacry, A. Iuga, M. Lasnier, and C-A. Lehalle. Market impacts and the life cycle of investors orders. *arXiv preprint arxiv:1412.0217v2*, 2014.

[6] E. Bacry, T. Jaisson, and J.F. Muzy. Estimation of slowly decreasing hawkes kernels: Application to high frequency order book modelling. *arXiv preprint arXiv:1412.7096*, 2015.

[7] E. Bacry and J.F. M. Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance*, 14(7):1–20, 2014.

[8] E. Bacry and J.F. Muzy. Second order statistics characterization of hawkes processes and non-parametric estimation. *arXiv preprint arXiv:1401.0903*, 2014.

[9] L. Bauwens and N. Hautsch. *Modelling financial high frequency data using point processes*. Springer, 2009.

[10] L. Bauwens and N. Hautsch. Stochastic conditional intensities processes. *Journal of Financial Econometrics*, 4:450–493, 2009.

[11] B. Biais, P. Hillion, and C. Spatt. An empirical analysis of the limit order book and the order flow in the paris bourse. *the Journal of Finance*, 50(5):1655–1689, 1995.

[12] G. Bormetti, L.M. Calcagnile, M. Treccani, F. Corsi, S. Marmi, and F. Lillo. Modelling systemic price cojumps with hawkes factor models. *arXiv preprint arXiv:1301.6141*, 2013.

[13] J.P. Bouchaud. Price impact. In R. Cont, editor, *Encyclopedia of Quantitative Finance*. John Wiley & Sons Ltd., 2010.

[14] J.P. Bouchaud, J. D. Farmer, and F. Lillo. How markets slowly digest changes in supply and demand. In Thorsten Hens and Klaus Schenk-Hoppe, editors, *Handbook of Financial Markets: Dynamics and Evolution*, pages 57–156. Elsevier: Academic Press, 2008.

[15] C.G. Bowsher. Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141(2):876–912, 2007.

[16] P. Brémaud and L. Massoulié. Stability of nonlinear hawkes processes. *The Annals of Probability*, pages 1563–1588, 1996.

[17] P. Brémaud, L. Massoulié, et al. Hawkes branching point processes without ancestors. *Journal of applied probability*, 38(1):122–135, 2001.

[18] V. Chavez-Demoulin and JA McGill. High-frequency financial data modeling using hawkes processes. *Journal of Banking & Finance*, 36(12):3415–3426, 2012.

[19] R. Cont, A. Kukanov, and S. Stoikov. The price impact of order book events. *Journal of financial econometrics*, 12(1):47–88, 2013.

[20] J.C. Cox, Jr Ingersoll, and S.A. Ross. A theory of the term structure of interest rates. *Econometrica*, 5:385–407, 1985.

[21] J. Da Fonseca and R. Zaatour. Hawkes process: Fast calibration, application to trade clustering, and diffusive limit. *Journal of Futures Markets*, 34(6):548–579, 2014.

[22] D.J. Daley and D. Vere-Jones. *An introduction to the theory of point processes*, volume 2. Springer, 1988.

[23] A. Dassios and H. Zhao. A dynamic contagion process. *Advances in applied probability*, 43(3):814–846, 2011.

[24] A. Dassios and H. Zhao. Exact simulation of hawkes process with exponentially decaying intensity. *Electronic Communications in Probabilities*, 18(62), 2013.

[25] M. Dhamala, G. Rangarajan, and M. Ding. Estimating granger causality from fourier and wavelet transforms of time series data. *Phys. Rev. Lett.*, 100(1):18701, 2008.

[26] D. Duffie, D. Filipovic, and W. Schachermayer. Affine processes and applications in finance. *Annals of Applied Probability*, 13:984–1053, 2003.

[27] P. Embrechts, T. Liniger, L. Lin, et al. Multivariate hawkes processes: an application to financial data. *Journal of Applied Probability*, 48:367–378, 2011.

[28] R.F. Engle and J.R. Russel. Autoregressive conditional duration: A new model for irregulary spaced transaction data. *Econometrica*, 66:1127–1162, 1998.

[29] E. Errais, K. Giesecke, and L.R. Goldberg. Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics*, 1(1):642–665, 2010.

[30] A. Fauth and C. Tudor. Modeling first line of an order book with multivariate marked point processes. *arXiv preprint arXiv:1211.4157v1*, 2012.

[31] V. Filimonov and D. Sornette. Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review E*, 85(5):056108, 2012.

[32] V. Filimonov and D. Sornette. Apparent criticality and calibration issues in the hawkes self-excited point process model: application to high-frequency financial data. *arXiv preprint arXiv:1308.6756*, 2013.

[33] N.R Hansen, P. Reynaud-Bouret, and V. Rivoirard. Lasso and probabilistic inequalities for multivariate point-processes. *arXiv preprint arXiv:12080570, To appear in Bernoulli*.

[34] S.J. Hardiman, N. Bercot, and J.P. Bouchaud. Critical reflexivity in financial markets: a hawkes process analysis. *arXiv preprint arXiv:1302.1405*, 2013.

[35] S.J. Hardiman and J.P. Bouchaud. Branching ratio approximation for the self-exciting hawkes process. *arXiv preprint arXiv:1403.5227*, 2014.

[36] J. Hasbrouck. Measuring the information content of stock trades. *Journal of Finance*, 46:179–207, 1991.

[37] A.G. Hawkes. Point spectra of some mutually exciting point processes. *J. R. Statist. Soc. B*, 33:438–443, 1971.

[38] A.G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[39] S.L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, pages 385–407, 1993.

[40] P. Hewlett. Clustering of order arrivals, price impact and trade path optimisation. In *Workshop on Financial Modeling with Jump processes, Ecole Polytechnique*, pages 6–8, 2006.

[41] T. Jaisson. Market impact as anticipation of the order flow imbalance. *arXiv preprint arXiv:1402.1288*, 2014.

[42] T. Jaisson and M. Rosenbaum. Limit theorems for nearly unstable hawkes processes. *To appear in The Annals of Applied Probability.*, 2014.

[43] A. Jedidi and F. Abergel. Stability and price scaling limit of a Hawkes-process based order book model. 2013.

[44] Stojan Jovanivić, John Hertz, and Stefan Rotter. Cumulants of hawkes point processes. *arXiv preprint arXiv:1409.5353*, 2014.

[45] M. Lallouache and D. Challet. Statistically significant fits of hawkes processes to financial data. *Available at SSRN 2450101*, 2014.

[46] J. Large. Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, 10(1):1–25, 2007.

[47] E. Lewis and G. Mohler. A nonparametric em algorithm for multiscale hawkes processes. *preprint*, 2011.

[48] P.A. Lewis and G.S. Shedler. Simulation of nonhomogeneous poisson processes by thinning. *Naval Research Logistics Quarterly*, 3:403–413, 1979.

[49] S.L. Linderman and R.P. Adams. Discovering latent network structure in point process data. *arXiv preprint arXiv:1402.0914v1*, 2014.

[50] D.. Marsan and O. Lengliné. Extending earthquakes' reach through cascading. *Science*, 319:1076, 2008.

[51] I. Mastromatteo, E. Bacry, and J.F. Muzy. Linear processes in high-dimension: phase space and critical properties. *arXiv preprint arXiv:1412.6998*, 2014.

[52] I. Mastromatteo and M. Marsili. On the criticality of inferred models. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(10):P10012, 2011.

[53] B. Mehrdad and L. Zhu. On the hawkes process with different exciting functions. *arXiv preprint arXiv:1403.0994*, 2014.

[54] J. Moller and J.G. Rasmussen. Perfect simulation of hawkes processes. *Adv. in Appl. Probab.*, 37(3):629–646, 2005.

[55] I. Muni Toke. Market making" in an order book model and its impact on the bid-ask spread. In *Econophysics of Order-Driven Markets, New Economic Windows.* Springer, 2010.

[56] I. Muni Toke and F. Pomponio. Modelling trades-through in a limit order book using hawkes processes. *Economics: The Open-Access, Open-Assessment E-Journal*, 6(2012-22), 2012.

[57] A.G. Nedungadi, G. Rangarajan, N. Jain, and M. Ding. Analyzing multiple spike trains with nonparametric granger causality. *Journal of computational neuroscience*, 27(1):55–64, 2009.

[58] Y. Ogata. On lewis' simulation method for point processes. *Ieee Transactions On Information Theory*, 27:23–31, January 1981.

[59] Y. Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, pages 9–27, 1988.

[60] Y. Ogata and H. Akaike. On linear intensity models for mixed doubly stochastic poisson and self- exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(1):102–107, January 1982. ArticleType: research-article / Full publication date: 1982 / Copyright 1982 Royal Statistical Society.

[61] M. Rambaldi, P. Pennesi, and F. Lillo. Modeling fx market activity around macroeconomic news: a hawkes process approach. *arXiv preprint arXiv:1405.6047*, 2014.

[62] P. Reynaud-Bouret and S. Schbath. Adaptive estimation for hawkes processes; application to genome analysis. *Ann. Statist*, 38:2781–2822, 2010.

[63] Y.A. Sahalia, J. Cacho-Diaz, and R.J.A. Laeven. Modeling financial contagion using mutually exciting jump processes. (15850), 2014.

[64] E. Smith, J.D. Farmer, L. Gillemot, and S. Krishnamurthy. Statistical theory of continuous double auction. *Quantitative Finance*, 3:481–514, 2003.

[65] D. Sornette and G. Ouillon. Multifractal scaling of thermally activated rupture processes. *Phys. Rev. Lett.*, 94:038501, Jan 2005.

[66] Bence Tth, Imon Palit, Fabrizio Lillo, and J. Doyne Farmer. Why is equity order flow so persistent? *Journal of Economic Dynamics and Control*, 51(0):218 – 239, 2015.

[67] A. Veen and F.P. Schoenberg. Estimation of space-time branching process models in seismology using an em-type algorithm. *J. Amer. Statist. Assoc.*, 103:614–624, 2008.

[68] M. Wyart, J.P. Bouchaud, J. Kockelkoren, M. Potters, and M. Vettorazzo. Relation between bid–ask spread, impact and volatility in order-driven markets. *Quantitative Finance*, 8(1):41–57, 2008.

[69] B. Zheng, Roueff F., and F. Abergel. Ergodicity and scaling limit of a constrained multivariate hawkes process. *SIAM J. Financial Math*, 5, 2014.

[70] K. Zhou, H. Zha, and L. Song. Learning triggering kernels for multi-dimensional hawkes processes. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pages 1301–1309, 2013.