UNIVERSITÀ di VERONA

Department of ECONOMICS

Working Paper Series
Department of Economics
University of Verona

Warnings about future jumps: properties of the exponential Hawkes model

Rachele Foschi, Francesca Lilla, Cecilia Mancini

WP Number: 13 June 2020

ISSN: 2036-2919 (paper), 2036-4679 (online)

Warnings about future jumps: properties of the exponential Hawkes model

Rachele Foschi, University of Pisa*
and
Francesca Lilla, Bank of Italy†
and
Cecilia Mancini, University of Verona‡

Abstract

Having observed a cluster of jumps produced by an exponential Hawkes process, we study and quantify the residual length of the cluster. We then formalize the stochastic increasingness property of the durations between two consecutive jumps, which strengthens their positive correlation. Finally we consider the case where the process is only observed discretely and provide bounds for the probability of observing a given number of consecutive jumps.

As an empirical exercise, we apply our results to a record of JPM's asset prices. First, we show that the identified jumps display dependence and clustering behavior. Second, we find that, under the exponential Hawkes model delivering the best QQ-plot, our formulas indicate a very high probability that an observed cluster of more than 1 jump did not exhaust yet. [1]

Keywords: clusters of jumps, exponential Hawkes process, residual length of a cluster, conditional probability of a configuration of jumps, financial assets returns, truncation

JEL classification codes: C02, C52, C58

1 Introduction

Motivation. When a financial trader observes a cluster of jumps in the price evolution of an asset, she could be interested in understanding whether the cause of such a cluster is going to produce further jumps or if instead its effects are already diminishing and the price will return to its "usual" fluctuations.

One possible explanation for the occurrence of such a cluster is offered by the asymmetry of the information circulating in the market: although price efficiency implies that asset prices incorporate

*Department of Management and Economics, via Ridolfi 10, 56124 Pisa (rachele.foschi@unipi.it)

†Directorate General for Economics, Statistics and Research, Via Nazionale 91, 00184 Roma (francesca.lilla@bancaditalia.it)

‡Corresponding author: Department of Economics, via Cantarane 24, 37129 (cecilia.mancini@univr.it)

[1]The views expressed in this paper are those of the authors and do not necessarily reflect those of their institutions.

Draft of June 29, 2020

all of the present information, the information is not immediately acknowledged by all traders (e.g., [14] and [13]). For instance, events such as the default or a temporary crisis of a firm are not isolated, sudden and completely unexpected facts: a firm starts to manifest financial difficulties well before the occurrence of the main event. The most informed traders have information about such difficulties and they tend to liquidate the assets issued by the firm, including the strictly correlated assets. This creates quick depreciations of the assets and can stimulate a herding behavior by the less informed traders, which impacts on the prices by exacerbating and amplifying the downward movements. This translates into a sequence of large negative variations in the asset price time series or, in other words, a cluster of jumps.

From this perspective, the occurrence of subsequent negative large variations of prices could be interpreted as warnings about difficult periods ahead and, for instance, as indications for portfolio management. Clearly, specular considerations can be made with respect to positive large variations of prices: good news and positive large variations can generate euphoria in the market participants and trigger further large variations.

In this work, we aim to capture the dependence mechanism of all of the jumps that are estimated in a price path. Consequently, we focus on the time of occurrence of price large variations, regardless of their sign. However, limiting the analysis to jumps with a specific sign is also possible, both in an univariate and in a multivariate framework.

The risk underlying an asset is split here into two categories: the risk of "continuous", "predictable and usually relatively small price adjustments; and the risk of abrupt, "abnormal" and possibly large fluctuations. The two types of risk are modeled by two radically different processes: a Brownian motion with stochastic volatility and a jump process. As e.g. in [4], we are here only interested in the jump part. Based on [17], with discrete observations a *jump* is detected when a return has size above a given threshold. It is also called *large variation*, or *jump over the threshold*. We focus on the point process counting the arrivals of the estimated jumps, and on its stochastic intensity.

From the mathematical perspective, a counting process whose jumps increase the intensity provides an effective way to capture the self-excitation mechanism of the jumps. In particular, inspired by [1] and [4], we assume for our point process the branching structure of a univariate Hawkes model. Among feasible kernels for an Hawkes process, both after data fitting tests and in view of its mathematical tractability and analytical properties, we focus on an *exponential Hawkes process*, which is characterized by a one-term exponential kernel. Under this specification, we obtain new theoretical properties which are useful for our purpose of quantifying the probability that the

phenomenon having generated a cluster of jumps did not exhaust its effects yet by the time at which we observe.

Originality of our results. First, we quantify the residual length of an observed cluster of jumps in terms of the model parameters (Sec.4.2). This is stronger than measuring the generic, and known, probability that a given jump is son of some previous observed jump ([27]).
Second, defined *interarrival time*, or simply *interarrival*, the duration between two consecutive jumps, for the exponential Hawkes we can state a strong positive dependency property, the *stochastic increasingness*, which describes the conditional probability of one interarrival time given the length of the previous one. Such a property implies the positive correlation but is not equivalent.
Third, by analyzing a discretized version of the exponential Hawkes process, we compute bounds for the probability that particular jumps configurations occur after the observation time.
Our analytic expressions can help a financial analyst or researcher to obtain important information about future unusual events, even when the exact time location of previous jumps is not completely known.

Finally we show an empirical application of our results to the 5 minutes prices of the JPM and S&P500 assets during periods that are about seven and half years long starting from 2006 and 2010, respectively. First, we find that the JPM asset estimated jumps display self-excitation and significant clustering behavior. This evidence is confirmed by the mismatch between the probabilities of having specific clusters of jumps for the Poisson best fitting process and the corresponding empirical probabilities. This means that the main features of the Hawkes process are present in the data and we cannot reject that such a process generated the estimated jump times. Moreover, we fit three types of Hawkes models to the JPM jumps, using maximum likelihood estimation, and we argue that the exponential model has best fit. We find that the best fitting QQ-plot is obtained with the parameters giving a local maximizer of the likelihood rather than with those of the global maximizer. While the empirical probability of the occurrence of a cluster of many jumps is in some cases enormously higher for the JPM prices than for the best fitting Poisson model, this probability is found to be quite compatible with the theoretical bounds obtained for an exponential Hawkes process. Furthermore, we find that under the best QQ-plot fitting model, the probability that an observed cluster of more than 1 jump did not exhaust yet is very high, mostly close to 1. This means that, for the specific JPM dataset, when we observed a cluster of jumps, we should have almost always expect at least a further jump.

In contrast, we find that with our data the S&P500 jumps do not display self-excitation. A

possible explanation for this can be found in the nature of the index: only concomitant jumps of the same sign in a relevant number of components result in a jump of the index. However, episodes generating this phenomenon seem to be sparse during the considered period, which excludes the credit crisis period of 2008. Thus, the S&P500 jumps do not allow us to perform the type of analysis that we have in mind.

Outline of the paper. Section 2 points out our contribution to the literature, Section 3 illustrates the theoretical framework. Section 4 describes the mentioned statistical and probabilistic properties of the exponential Hawkes process. Section 5 presents the bounds for the probabilities of the occurrence of particular configurations of jumps, accounting for the discrete nature of the data. Section 6 deals with the empirical application of our results to financial data and Section 7 concludes. A web appendix contains the proofs of the results and supplemental explanatory material.

2 Literature.

To the best of our knowledge, this paper represents the first attempt to study and to formalize a series of statistical and probabilistic properties of a univariate exponential Hawkes process. The formulas that we provide can be applied to a wide variety of problems in economics and finance, such as the problem of measuring contagion phenomena.

Choice of the kernel. The most theoretical strand of the literature about Hawkes processes deals with either the univariate or the multivariate case with general kernel, delivering general properties but not expliciting the probabilities of specific events of interest.

The exponential choice is widespread (e.g. [22], [19]), and is justified by the mathematical tractability. In fact in our case the choice is also motivated by the fact that it allows to obtain the best fit to our data among the most popular Hawkes processes.

Discretization. In view of the empirical application, in Sec.5 we take into account the fact that we apply a continuous-time model to a phenomenon that we only are able to observe it discretely. Our bounds for the probabilities of specific configurations of jumps consider that an estimated jump could occur at any time within a whole interval $[t_{i-1}, t_i)$. Instead in [23] the authors go in the opposite direction: when they estimate that a jump falls within an interval $[t_{i-1}, t_i)$ they associate to it a randomly chosen time of occurrence within $[t_{i-1}, t_i)$. In this way they recover the continuous time nature of the counting process. The advantage of our approach is that we can spot interesting

configurations of jumps and measure their probability, while in [23] the authors cannot do it.

Choice of the returns to be analyzed. Our choice is to model the times of arrival of returns that, based on the theoretically justified threshold method, are considered to include jumps. Other authors (e.g. [22], [4]) analyze the self-exciting structure of returns above (or below) a fixed level. They do not separate the variations of the asset prices into the class of the moderate variations, compatible with the increments of a Brownian semimartingale with stochastic volatility, and the class of the abnormal, incompatible, variations.

At the aim of measuring the contagion effects of financial crises, in [1] the authors consider a multivariate stochastic volatility process with jumps to model the underlying assets. Also them keep the contributions of the Brownian part and of the jump process aggregated. To estimate the parameters the Generalized Method of Moments is applied, which forces the authors to specify a model also for the spot volatility. The univariate version of that model is included in our framework, since our volatility process is non-parametric.

In order to eliminate the daily periodicity of the large returns, [23] deseasonalize them. In our case we estimate the volatility process non-parametrically and our threshold is defined in terms of $\hat{\sigma}$. Thus we already account for seasonality in volatility. However both in [23] and here also the jumps show some kind of seasonality, which is proved by the steps on the QQ-plots.

3 Theoretical framework

Generalizing the univariate model proposed in [1], we consider the following dynamics for the log price of a financial asset:

$$dX_t = a_t dt + \sigma_t dW_t + J_t, \quad t \in [0, T], \tag{1}$$

with fixed $T > 0$, cádlág *drift* a and *volatility* σ coefficients, W a Wiener process, and J a pure jump semimartingale. We assume $\sigma_t(\omega) \geq \underline{\sigma}$ a.s., $\forall t \in [0, T]$, for some constant level $\underline{\sigma} > 0$.

We consider discrete high frequency and evenly spaced observations $X_{t_1}, X_{t_2}, ..., X_{t_m}$ of X on the finite time horizon $[0, T]$, with $t_i = i\delta$ and $\delta m = T$. By the *truncation method* we can estimate the time τ_ℓ of occurrence of a jump as the left extreme t_{i-1} of an interval $[t_{i-1}, t_i)$ where the return $\Delta_i X := X_{t_i} - X_{t_{i-1}}$ is larger in absolute value than the *threshold* $\sqrt{2\sigma_{t_{i-1}}\delta \log \frac{1}{\delta}}$, as explained in Section 6.2. These returns estimate the jump sizes of X as $\delta \to 0$ ([17]) because they are incompatible with a Brownian semimartingale, even if the spot volatility σ was high.

Here $\delta > 0$ is fixed and we define N the process counting the number of the estimated occurred

jumps. To verify the presence of clusters of jumps and to quantify the probability that such a cluster did not exhaust yet, we assume that J is independent on a, σ and W, and we model N using a univariate Hawkes process.[2] For the latter process, the jump intensity $\bar{\lambda}$ is *self exciting*; that is, for any $t \in \mathbb{R}_+$, $\bar{\lambda}_t$ is influenced, in a deterministic way, by the occurrence of the jumps prior to t.

A counting càdlàg process $N = \{N_t\}_{t\geq 0}$ with natural filtration $\mathcal{F} \doteq \sigma(N_s, 0 \leq s \leq t)$ is of Hawkes type when it is a *simple*[3] point process and has stochastic intensity $\bar{\lambda}$ given by

$$\bar{\lambda}_t = \lambda_0 + \int_0^t \Phi(t-s)dN_s = \lambda_0 + \sum_{\ell:0\leq T_\ell<t} \Phi(t-T_\ell), \tag{2}$$

where $\lambda_0 \in [0, +\infty)$ is constant, the *kernel* $\Phi \geq 0$ is a deterministic function defined on $\mathbb{R}_+$ and belonging to $L^1_{loc}(\nu)$, ν being the Lebesgue measure on $\mathbb{R}_+$ ([2] p.3), and T_ℓ are the random jump times of the very N. The last expression in (2) is obtained from the definition of integral with respect to dN_s ([8], Appendix A1.4). Note that, consistently with the literature dealing with Hawkes processes, we take $\int_0^t \Phi(t-s)dN_s \doteq \int_{[0,t)} \Phi(t-s)dN_s$.
We assume here $N_0 = 0$, thus the first jump of the intensity occurs after 0: at the first jump time of N the intensity passes from the starting level λ_0 to $\lambda_0 + \Phi(0)$. As soon as $\Phi(0) > 0$, at any jump time of N, the intensity $\bar{\lambda}$ undergoes an increase and, because $\bar{\lambda}_t$ allows to measure the probability of a jump occurrence just after t, an increase in $\bar{\lambda}$ implies a higher probability of having some other jumps immediately after. This generates a self-excitation mechanism, which in fact would be guaranteed by the conditions $\Phi \geq 0, \Phi \not\equiv 0$. When $\Phi \equiv 0$, the self excitation is absent and the Hawkes process reduces to a Poisson process.

Under a *branching processes* perspective, N is the superposition of a Poisson process with parameter λ_0 and the process counting the remaining jumps, which are triggered by the fact that after any previous jump $\bar{\lambda}$ increased. The jumps generated by the Poisson process are called *fathers* and, in our model, they represent sporadic and totally independent jumps having non-persistent

[2] As in many papers in the literature (e.g. [9], Sec. 4, besides [1]) we assume that N is independent on the other components of the model for X. Another strand of literature instead assumes that λ can be an affine function of spot σ. Thus in the empirical section we also considered for our estimated jump times $\hat{\tau}_i$ a point process having the alternative $\check{\lambda}_t = a + b\sigma_t$. Using the same $\hat{\sigma}_t$ non parametrically obtained following the procedure described in Sec.6.2, we found positive and significant MLE $\hat{a}, \hat{b}$. However the QQ-plot of the values $\int_{\tau_{i-1}}^{\tau_i} \check{\lambda}_t dt$ against values of independent E(1) rvs turned out to be worse than the one of the values $\int_{\tau_{i-1}}^{\tau_i} \lambda_t dt$, being λ_t the conditional jump intensity of our best fit Hawkes model (the latter QQ-plot is in Fig.3, left-hand.)

[3] Having, a.s., at each t at most one jump.

causes. The others are called *descendants* and represent jumps generated as consequences of more persistent phenomena.

Given $\mathcal{F}_{t-}$, we exactly know when and where the jump times before t occurred. To indicate that they are no longer random, we define them τ_n, rather than T_n. If we know the parameters of the model and we are able to observe the process N continuously, then we exactly know the path realized by the intensity up to $t-$ and we call λ_t the intensity process conditional to $\mathcal{F}_{t-}$. In particular, λ_t is defined by $\lambda_t \equiv \lim_{dt\to 0} \frac{1}{dt} P(N_{t+dt} - N_t = 1|\mathcal{F}_{t-})$, so that

$$P(N_{t+dt} - N_t = 1|\mathcal{F}_{t-}) = \lambda_t dt + o(dt). \tag{3}$$

We also have

$$\lambda_t = E[\bar{\lambda}_t|\mathcal{F}_{t-}] = \lambda_0 + \sum_{\ell:0\le\tau_\ell<t} \Phi(t-\tau_\ell),$$

so that λ (other than $\bar{\lambda}$) turns out to be a cáglád process. The existence and uniqueness of a point process satisfying (3) with $\bar{\lambda}$ as in (2) is guaranteed, for example, in [3] (Thm 3.1). The process N_t has asymptotically stationary increments and $\bar{\lambda}_t$ is asymptotically stationary, as $t \to +\infty$, if the kernel function Φ satisfies the *stability condition* $\int_0^{+\infty} \Phi(x)dx < 1$ ([2] Proposition 1, [6], Proposition 4.4).

In the empirical application of this paper, we specify Φ either in exponential form, which is expressed as a finite sum of exponential terms, or in power form. In the exponential case, we take $\Phi(x) = \sum_{p=1}^{P} \alpha_p e^{-\beta_p x}$, with constants $\alpha_p \ge 0$ and $\beta_p > 0$ such that $\beta_p \ne \beta_q$ for $q \ne p$. If $\alpha_p = 0$ for any p, then $\Phi \equiv 0$ and N is a Poisson process. If $P = 1$ and $\alpha_1 \doteq \alpha > 0$, then the effect of any single jump of N is to increase $\bar{\lambda}$ instantaneously by α when a jump occurs, but then such an effect decays exponentially with speed $\beta_1 \doteq \beta$: a higher β indicates a weaker persistence of the effect of the occurred jumps; that is, a quicker return of $\bar{\lambda}$, after its jump, towards the level λ_0. A model with $P > 1$ and $\alpha_p > 0$ for all p is more flexible, because the impact on $\bar{\lambda}$ of each single jump is more complicated: the instantaneous increase $\sum_p \alpha_p$ of $\bar{\lambda}$ splits into P quantities α_p smoothing at the different speeds β_p. If we allowed for $\beta_p = \beta_q$ then the description of $\alpha_p e^{-\beta_p x} + \alpha_q e^{-\beta_q x}$ could be simplified in $(\alpha_p + \alpha_q)e^{-\beta_p x}$.

We assume $\beta_p > 0$ for any p to guarantee the existence of a finite asymptotic stationary measure for $\bar{\lambda}$, when $T \to +\infty$. If it was $\beta_{\bar{p}} = 0$ for a specific $\bar{p}$, then the condition $\int_0^{+\infty} \Phi(x)dx < 1$ would not be satisfied and the instantaneous increase by $\sum_p \alpha_p$ of $\bar{\lambda}$ after the occurrence of a jump of N would in fact contain a part $\alpha_{\bar{p}}$ not decaying and staying persistent. Thus, at any jump of N, $\bar{\lambda}$ would increase with a permanent effect, ending up with explosion for $T \to +\infty$. The stationarity

condition is satisfied iff $\sum_{p=1}^{P} \frac{\alpha_p}{\beta_p} < 1$.

In the power case, we have $\Phi(x) = \frac{\alpha\beta}{(1+\beta x)^{1+\gamma}}$, with $\alpha, \beta > 0$, so that after the occurrence of a jump of N the intensity $\bar{\lambda}$ increases by the amount $\alpha\beta$ and, as soon as $1 + \gamma > 0$, the impact of such an increase decays as a power of the time elapsed from the jump arrival. The higher either β or γ, the higher the speed of decay. The stationarity condition is equivalent to $\alpha < \gamma$, implying $\gamma > 0$.

4 Exponential Hawkes properties: quantification of jumps risk

In this section, we present the theoretical results for the model with one exponential term kernel because it is the most appropriate for our JPM dataset. The proofs of all the stated theoretical results are postponed to the web Appendix. These properties are then exploited to measure the probability that a cluster of jumps that we are experiencing is going to produce further jumps in the near future.

Although we use the term *cluster* of jumps to indicate a generic group of close jumps, we consider particular types of clusters, where the jumps assume specific configurations: we distinguish *clusters*$^+$ and *clusters*$^\circ$, while in the Appendix we also introduce *clusters**.

4.1 Bounds for $\bar{\lambda}_t$ and λ_t

By its definition, $\lambda_t\delta$ approximates, for sufficiently small $\delta > 0$, the probability $P\{N_{t+\delta} - N_t = 1|F_{t-}\}$ that, given the state of the system up to time $t-$, one jump could occur immediately after. Furthermore, $\lambda_{n\delta}dt$ also equals the expected value $E[N_{n\delta+dt} - N_{n\delta}|\mathcal{F}_{n\delta-}]$, thus a quantification of $\lambda_{n\delta}$ also gives the number of jumps that we can expect just after $n\delta$, if we are at time $n\delta-$. A high value of λ_t indicates a high risk of having jumps just after t. Thus, in practice it is important to evaluate the amplitude of λ_t in terms of its parameters because, once we have estimated them, we have a clear quantification of the mentioned risk. In particular, it is of interest to have bounds for the highest and lowest values that λ_t can assume. At a time instant $t > 0$, the observation in continuous time of the past realization λ of $\bar{\lambda}$ is not available to us, while it is feasible to estimate how many jumps $\bar{\lambda}$ experienced up to $t-$, we thus use bounds for $\bar{\lambda}_t$ conditionally to $N_{t-} = n$ in what follows. Note that λ and $\bar{\lambda}$ a posteriori realize the same paths, so the obtained bounds are also valid for λ.

If T_1 is the first jump time of N, then we have $\bar{\lambda}_t = \lambda_0$ for all $t \in [0, T_1]$, while $\bar{\lambda}_{T_1^+} = \lambda_0 + \alpha$.

Conditionally on the occurrence of n jumps before t, we have that the supremum of the possible values that $\bar{\lambda}_t$ can reach along all the possible paths of N is $\sup \bar{\lambda}_t = \lambda_0 + \alpha n$, because $T_1, T_2, ..., T_n$ could happen arbitrarily close to t. Analogously, for $t > T_1$, λ_0 is the infimum value for $\bar{\lambda}_t$ along all the possible paths of N because the subsequent jump time (certainly occurring) could be arbitrarily far from t. Finding narrow bounds for $\bar{\lambda}_t$ is not easy. Next Lemma 1 is known, however it is stated because it clarifies that the bounds crucially depend on how concentrated the jump times are on $[0, t)$, and will be much used in what follows. The bounds in two extremal cases are explicit, in all the other situations $\bar{\lambda}_t$ can assume all the intermediate values.

Lemma 1. Conditionally on $N_{t-} = n$ we have $\lambda_0 + \alpha n e^{-\beta t} < \bar{\lambda}_t < \lambda_0 + \alpha n$. Furthermore, for any $\eta > 0$,

1. conditionally on $T_i < \eta \ \ \forall i = 1, ..., n$: $\lambda_0 + \alpha n e^{-\beta t} < \bar{\lambda}_t < \lambda_0 + \alpha n e^{-\beta(t-\eta)}$;
2. conditionally on $T_i > t - \eta \ \ \forall i = 1, ..., n$: we have $\lambda_0 + \alpha n e^{-\beta \eta} < \bar{\lambda}_t < \lambda_0 + \alpha n$.

The given bounds do not depend on the exact locations of the occurred jumps before t because we do not know them. In fact, it is not possible to improve the given bounds for $\bar{\lambda}$ if the exact jumps locations are not available. Lemma 1 is also used for the statement just above (4), for (11) and Proposition 4.

4.2 Decay instant and clusters amplitude

It is likely that a default of a firm is manifested through a jump of the firm asset price which is triggered by previous unusual instabilities. This can lead us to classify the whole set of the price large variations as a cluster of jumps. Then, it is natural to wonder whether the phenomenon having produced such a cluster is going or not to produce further effects. In order to detail our main results, we now formally introduce a more specific notion of cluster, the cluster$^+$.

We identify as a *father* a jump happening when the value of λ_t is close to λ_0 with an accuracy of ϵ, i.e. $\lambda_t < \lambda_0(1 + \epsilon)$, where $\epsilon > 0$ is small. We define a *cluster*$^+$ as the set of all the jumps arrived after a father and before the subsequent time t when λ_t comes again very close to λ_0. We thus consider the jumps within a cluster$^+$ as being (direct or indirect) descendants of a father jump, occurred at some previous time instant V. We are interested in understanding *how long* the event happened at V can produce consequences. If a jump occurred at a time $S > V$ belongs to the cluster$^+$ then its descendants are also descendants of the father event at V. Our interpretation is

that if the event at time S produces some descendants, then the cluster$^+$ triggered by the father event is further extended, and the consequences of the event in V did not exhaust yet. When instead after time V the intensity is again close to λ_0 then no events of the cluster has anymore the capability of producing descendants, so the father event has already been absorbed by the market.

For $\lambda_0 > 0$, given the path of $\bar{\lambda}$ up to any jump time S (note that given $\mathcal{F}_S$ we have $\lambda_u = \bar{\lambda}_u$ for all $u \leq S$) such that $\lambda_S + \alpha > \lambda_0(1+\epsilon)$ we define *decay instant* the deterministic number

$$t_\epsilon(S) := \min\{t > 0|\ \lambda_0 + (\lambda_S - \lambda_0 + \alpha)e^{-\beta t} < \lambda_0(1+\epsilon)\}.$$

Note that, if two consecutive jumps occurred at times τ_{i-1} and τ_i, then for any $s \in (\tau_{i-1}, \tau_i)$ we have that $\lambda_s = \lambda_0 + (\lambda_{\tau_{i-1}} - \lambda_0 + \alpha)e^{-\beta(s-\tau_{i-1})}$, thus, with $t = s - \tau_{i-1}$, $t_\epsilon(\tau_{i-1})$ measures the amount of time we would have to wait for obtaining $\lambda_s < \lambda_0(1+\epsilon)$ *if* no further jumps occurred. It turns out that

$$t_\epsilon(S) = \frac{1}{\beta}\log\left(\frac{\lambda_S - \lambda_0 + \alpha}{\epsilon\lambda_0}\right).$$

If instead S is not a jump time and $\lambda_S > \lambda_0(1+\epsilon)$ the decay instant is defined by

$$t_\epsilon(S) = \min\{t > 0|\ \lambda_0 + (\lambda_S - \lambda_0)e^{-\beta t} < \lambda_0(1+\epsilon)\},$$

and

$$t_\epsilon(S) = \tfrac{1}{\beta}\log\left(\tfrac{\lambda_S - \lambda_0}{\epsilon\lambda_0}\right).$$

Let us now consider the case where a cluster$^+$ occurred with at least k jumps, arrived at times $\tau_i, .., \tau_{i+k-1}$. By our definition of cluster$^+$, regardless of whether the father occurred at or before τ_i, the intensity is larger than $\lambda_0(1+\epsilon)$ during the whole period $[\tau_i, \tau_{i+k-1}]$. Then the $(k+1)$-th jump belongs to the same cluster$^+$ if and only if for $s \in [\tau_{i+k-1}, T_{i+k}]$ we still have $\lambda_s > \lambda_0(1+\epsilon)$, or equivalently

$$\lambda_{\tau_{i+k-1}+t} = \lambda_0 + (\lambda_{\tau_{i+k-1}} - \lambda_0 + \alpha)e^{-\beta t} > \lambda_0(1+\epsilon)$$

for $t \in [0, T_{i+k} - \tau_{i+k-1}]$. That is, the cluster$^+$ did not exhaust yet if and only if at T_{i+k} the intensity did not reach yet the level $\lambda_0(1+\epsilon)$, or $T_{i+k} < \tau_{i+k-1} + t_\epsilon(\tau_{i+k-1})$. That allows us to interpret $t_\epsilon(\tau_{i+k-1})$ as the *residual length* of the cluster$^+$.

More generally, if we are at any $S \in [\tau_{i+k-1}, T_{i+k})$, then the probability that the next jump is still a direct or indirect effect of the father of the cluster$^+$ is $P\{T_{i+k} < S + t_\epsilon(S)\}$.

Note that the event "the $(i+k)$-th jump belongs to the previous cluster$^+$" is different from the event "the jump occurred at τ_{i+k} is a descendant". First of all, in the latter case τ_{i+k} is known, while in the first one the jump time is still random. Secondly, the probability of the first event coincides with the probability that the intensity at the random time T_{i+k} is greater than $\lambda_0(1+\epsilon)$; while the probability that a jump already occurred at τ_{i+k} is a descendant is given by the ratio $q \doteq \frac{\lambda_{\tau_{i+k}} - \lambda_0}{\lambda_{\tau_{i+k}}}$

([27]). The method proposed in this paper allows us to obtain a more precise measure, with respect to q, of how long an agglomeration of jumps triggered by a specific event is going to last. For example, following the simulation exercise in Section 6, the estimation of $P\{T_{i+k} < S + t_\epsilon(S)\} \approx 1$ is close to the theoretical probability. On the contrary, assuming to know the future location T_{i+k} we would obtain $q = 0.49$. Otherwise, if we approximate $\lambda_{T_{i+k}}$ with λ_S then we obtain the even smaller probability $\frac{\lambda_S - \lambda_0}{\lambda_S} = 0.19$.

The explicit expression for $P\{T_{i+k} < S + t_\epsilon(S)\}$ is available in closed form (see Proposition 2 for the case $S \in (\tau_{i+k-1}, T_{i+k})$) and it is a function of λ_S and the parameters of the kernel. It follows that to implement this formula we need to know either the whole path of $\bar{\lambda}$ up to time S or both λ_S and the triplet $(\lambda_0, \alpha, \beta)$. However $\bar{\lambda}$ is not observable, due to the discrete observation of the stock log-price X, and since S comes after a cluster of jumps, the estimation of the value of $\bar{\lambda}$ at S is more uncertain than the estimation of its value at a time U within a more tranquil period. That is, we could know: with more precision the value λ_U for a certain $U < S$, for instance before the turbulent period started ($U < V < S$); and that a cluster$^+$ of k jumps occurred within $[U, S)$, without knowing the exact time locations. In this case, we provide below bounds for $t_\epsilon(S)$ allowing for an approximation of the probability $P(T_{i+k} < S + t_\epsilon(S))$ that T_{i+k} is part of the cluster$^+$ preceding it.

Let us consider the case $S \in (\tau_{i+k-1}, T_{i+k})$ (the case $S = \tau_{i+k-1}$ being analogous). With $S - U = s$, using a recursive formula for λ_S we find that (see A.26 in Appendix) $t_\epsilon(S)$ belongs to the interval $(\underline{t}_\epsilon, \bar{t}_\epsilon)$, where

$$\underline{t}_\epsilon = \max\left\{\frac{1}{\beta}\log\left(\frac{(\lambda_U - \lambda_0 + k\alpha)e^{-\beta s}}{\lambda_0 \epsilon}\right), 0\right\},\ \bar{t}_\epsilon = \frac{1}{\beta}\log\left(\frac{(\lambda_U - \lambda_0)e^{-\beta s} + k\alpha}{\lambda_0 \epsilon}\right): \tag{4}$$

$\underline{t}_\epsilon$ is $t_\epsilon(S)$ in the case where all the k jumps of the cluster$^+$ occurred at U (limit case 1 of Lemma 1), while $\bar{t}_\epsilon$ is $t_\epsilon(S)$ in the case in which all those jumps occurred at S (case 2). It follows that

$$P\{T_{i+k} < S + t_\epsilon(S)\} \in \Big(P\{T_{i+k} < S + \underline{t}_\epsilon\}, P\{T_{i+k} < S + \bar{t}_\epsilon\}\Big). \tag{5}$$

The lower bound corresponds to ascribing to the cluster$^+$ a shorter residual length with respect to the theoretical one, and this implies considering that the cluster$^+$ has a lower number of jumps. A symmetric consideration holds for the upper bound. Proposition 1 makes the bounds explicit.

Note that, however the jump times are distributed on (U, S), if s increases then (U, S) is larger, and the contribution to λ coming from the k jumps of the cluster$^+$ has more time to decay. This means that the level reached by λ at S is lower, and the residual length $t_\epsilon(S)$ of the cluster is lower. In particular $\bar{t}_\epsilon$ is lower. In turn, if $(S, S + \bar{t}_\epsilon)$ is shorter, then one would expect that the

probability of a jump on $(S, S+\bar{t}_\epsilon)$ is lower. In contrast, an increase of the number k of elements in the cluster$^+$ implies further contributions (each one by α) to the intensity process within $[U, S)$, any the distribution of the jump times is. In particular $(S, S+\bar{t}_\epsilon)$ is larger, and the probability of a further jump in $[S, S+\bar{t}_\epsilon)$ should have to be higher. Proposition 1 ensures that this in fact is the case, and, with respect to Proposition 2, has the advantage of making explicit the dependence on $s = S - U$ and on k our quantification of the length of the cluster triggered by the father event. Recall that S is assumed not to be a jump time. To simplify the next formulas (6) and (7), we omit writing the conditioning on knowing both the value $\bar{\lambda}_U$ of the intensity at time U and that k jumps occurred within (U, S).

Proposition 1. Let $S \in (\tau_{i+k-1}, T_{i+k})$ such that $\lambda_S > \lambda_0(1+\epsilon)$ then, conditionally on $N_S - N_U = k$ and on $\bar{\lambda}_U = \lambda_U$, we have

$$LB \doteq P\{T_{i+k} < S + \underline{t}_\epsilon\} = 1 - e^{-\frac{1}{\beta}\left(e^{-\beta s}(\alpha k + \lambda_U - \lambda_0) - \lambda_0 \epsilon\right)} \left((\alpha k + \lambda_U - \lambda_0)\frac{e^{-\beta s}}{\lambda_0 \epsilon}\right)^{-\frac{\lambda_0}{\beta}}, \qquad (6)$$

$$UB \doteq P\{T_{i+k} < S + \bar{t}_\epsilon\} = 1 - e^{-\frac{1}{\beta}\left(\alpha k + e^{-\beta s}(\lambda_U - \lambda_0) - \lambda_0 \epsilon\right)} \left((e^{\beta s}\alpha k + \lambda_U - \lambda_0)\frac{e^{-\beta s}}{\lambda_0 \epsilon}\right)^{-\frac{\lambda_0}{\beta}}, \qquad (7)$$

and the two probabilities are strictly decreasing in s and increasing in k.

Proposition 2. Let $S \in (\tau_{i+k-1}, T_{i+k})$ and $\epsilon > 0$ such that $\lambda_S > \lambda_0(1+\epsilon)$, then

$$P \doteq P(T_{i+k} < S + t_\epsilon(S)) = 1 - e^{-\frac{\lambda_S - \lambda_0(1+\varepsilon)}{\beta}} \left(\frac{\lambda_0 \varepsilon}{\lambda_S - \lambda_0}\right)^{\frac{\lambda_0}{\beta}}.$$

4.3 Durations between consecutive jump times

For the exponential Hawkes process the decay instant, and thus the probability that an observed cluster did not exhaust yet, is linked to the stochastic increasingness property between consecutive jump interarrivals. Accordingly, the longer we have to wait for the arrival of the n-th jump, the longer we will have to wait for the arrival of the $(n+1)$-th jump. Theorem 1 ensures the validity of this property and sheds light on the mentioned link. Proposition 3 is crucial for proving the Theorem.

Definition 1. *([16]) Let X, Y be two random variables. Y is stochastically increasing (resp. decreasing) in X if, for any y, the conditional law $P(Y > y|X = x)$ is increasing (resp. decreasing) in x. We abbreviate the term stochastically increasing by SI.*

The next result clarifies how the interarrival time $T_{n+1} - T_n$ depends on the past history through $\bar{\lambda}_{T_n}$.

Proposition 3. For any $\ell \geq \lambda_0$, conditionally on $\bar{\lambda}_{T_n} = \ell$, the density and the distribution function of the interarrival times $T_{n+1} - T_n$ are, respectively, given by

$$P(T_{n+1} - T_n = t | \bar{\lambda}_{T_n} = \ell) = \Big[\lambda_0 + (\ell - \lambda_0 + \alpha)e^{-\beta t}\Big] e^{(e^{-\beta t}-1)(\ell-\lambda_0+\alpha)/\beta - \lambda_0 t}; \tag{8}$$

$$P(T_{n+1} - T_n > \tau | \bar{\lambda}_{T_n} = \ell) = e^{\left[e^{-\beta\tau}(\ell+\alpha-\lambda_0) - \beta\lambda_0\tau - \ell - \alpha + \lambda_0\right]/\beta}. \tag{9}$$

Corollary 1. *$T_{n+1} - T_n$ is stochastically decreasing in $\bar{\lambda}_{T_n}$.*

The following Theorem is the main result of the present subsection.

Theorem 1. *For any $\ell \geq \lambda_0$, conditionally on $\bar{\lambda}_{T_{n-1}} = \ell$, $T_{n+1} - T_n$ is stochastically increasing in $T_n - T_{n-1}$.*

In particular, this last result formalizes, as shown in the web Appendix at (A.30), how the interarrival jump times are dependent. As well known, if we specify λ by taking either $\alpha = 0$ or $\beta \to +\infty$, we describe a simple Poisson process. Consistently, starting from the conditional distributions of the interarrival times we reach the same conclusion.

For the jump times T_i estimated on the JPM dataset and then modeled by an exponential Hawkes process, Figure 4, left-hand, in the web Appendix shows the copula of the conditional joint law of $T_{n+1} - T_n$ and $T_n - T_{n-1}$. A sensible conditional positive dependence is displayed when the model has the parameters of the local maximizer of the likelihood function. In this case, the self-excitation is significant. In contrast, we checked that the dependence is significant, but weaker, with the parameters of the global maximizer, due to the very high value of β.

We cannot deduce from the conditional SI, in general, that $T_{n+1} - T_n$ is *unconditionally* SI in $T_n - T_{n-1}$. However, our data display a certain unconditional positive dependence, even if weak, as shown in Figure 4, right-hand. On simulated paths of the Hawkes processes, we observe a similar behavior for the copula of the unconditional joint law in both the cases of local or global maximizers.

The fact that Theorem 1 only delivers (within its proof) formulas for $P\{T_{n+1} - T_n = t | T_n - T_{n-1} = s, \bar{\lambda}_{T_{n-1}}\}$, rather than $P\{T_{n+1} - T_n = t | T_n - T_{n-1} = s\}$, is useful in practice. In fact, once we estimated the jump times and the parameters of the fitted Hawkes model, we have some numerical bounds for the values assumed by $\bar{\lambda}_{T_{n-1}}$ so we can approximate $P\{T_{n+1} - T_n = t | T_n - T_{n-1} = s, \bar{\lambda}_{T_{n-1}}\}$. Furthermore, even if we could directly observe the whole history of $\bar{\lambda}$ up to T_{n-1} (i.e.

number and exact locations of the jumps), it could still be the case that we need to save memory space and computational time. In such a case we could decide to store only the partial information about the level reached by $\bar{\lambda}_{T_{n-1}}$, rather than the whole path of $\bar{\lambda}$.

The connection of the SI property with the decay instant is that $P\{T_{n+1} - T_n > \tau | T_n - T_{n-1} = s, \bar{\lambda}_{T_{n-1}}\}$ is decreasing in $t_\epsilon(T_{n-1})$. This means that, if by T_{n-1} we observed a cluster, a bigger $t_\epsilon(T_{n-1})$ corresponds to a farther end, and thus to a higher probability that T_n and T_{n+1} belong to the same cluster. Since the jumps of a cluster tend to be close, the conditional probability that $T_{n+1} - T_n > \tau$ has to be smaller. This is detailed in Appendix.

5 Jumps risk measurement with discrete observations

In this section, we expressly account for the discreteness of our observations and we measure the probability of the occurrence of specific jumps configurations. The Hawkes processes are designed in continuous time, while in our application the data are only available at the discrete times $i\delta$ with $\delta > 0$ (in Section 6 we have $\delta = 5$ minutes). This leads us to study some properties of the *discretized* version $(\lambda_{i\delta})_{i=1...[T/\delta]}$ of the conditional intensity λ. These will be used to compute an approximation for the first quantity of our interest in this section, namely

$$\tilde{\lambda}_n \doteq P(\text{ at least one jump in } [n\delta, (n+1)\delta) | \mathcal{F}_{n\delta-}).$$

For fixed δ and n, we define

$$\omega_i = \mathbf{1}_{\{\text{at least one jump occurred in } [i\delta,(i+1)\delta)\}}, \ i = n, n+1, ...,$$

thus

$$\tilde{\lambda}_n = P(\omega_n = 1 | \mathcal{F}_{n\delta-}).$$

Recall that estimating the occurrence of one jump of X (in (1)) within $[n\delta, (n+1)\delta)$ does not tell us exactly when it happened inside the interval. In addition, the occurrence of two or more distinct jumps within the interval will be registered as only one jump by the discretized process. [4]

To obtain $\tilde{\lambda}_n$, we can do a bit better than simply using the relation $\tilde{\lambda}_n \approx \lambda_{n\delta}\delta$ following by the definition of N. Consistently with the Markov property of our process, the conditional probability $\tilde{\lambda}_n$ can be expressed through the following exact function of $\lambda_{n\delta}$:

$$\tilde{\lambda}_n = P(\text{at least one jump in } [n\delta, (n+1)\delta) | \mathcal{F}_{n\delta-}) = \int_{n\delta}^{(n+1)\delta} \lambda_s e^{-\int_{n\delta}^{s} \lambda_\tau d\tau} ds$$

[4] However, this last event has a low probability to happen, especially if the observation interval δ is small, as $P(N_{t+dt} - N_t > 1 | \mathcal{F}_{t-}) = o(dt)$. In the web Appendix we quantify this probability for our JPM database, with δ equal to 5 minutes.

$$= 1 - \exp\left(\frac{1}{\beta}(\lambda_{n\delta} - \lambda_0)(e^{-\beta\delta} - 1) - \delta\lambda_0\right). \tag{10}$$

In particular, $\tilde{\lambda}_0 = 1 - e^{-\delta\lambda_0}$.

However, we usually do not know the value of $\lambda_{n\delta}$ for generic times $n\delta$, and we cannot compute the right-hand term in (10). Theorem 2 thus provides a recursive formula where $\tilde{\lambda}_n$ is approximated in terms of $\tilde{\lambda}_{n-1}$. Given that we know the exact expression for $\tilde{\lambda}_0$, once we have estimated the parameters of the model, we can reach a clear quantification of $\tilde{\lambda}_n$ for any n.

We need some intermediate steps. The first step follows from Lemma 1 and is a quantification of $\bar{\lambda}_{n\delta}$ given information on $\lambda_{(n-1)\delta}$ and the variable ω_{n-1}. However the same quantification can be given for the conditional $\lambda_{n\delta}$, because the paths realized at time $n\delta-$ by $\bar{\lambda}_{n\delta}$ and $\lambda_{n\delta}$ are the same. We have

$$\lambda_{n\delta} \begin{cases} \in \left[\lambda_0 + (\lambda_{(n-1)\delta} - \lambda_0 + \alpha)e^{-\beta\delta}, \lambda_0 + (\lambda_{(n-1)\delta} - \lambda_0)e^{-\beta\delta} + \alpha\right) & \text{if } \omega_{n-1} = 1 \\ = \lambda_0 + (\lambda_{(n-1)\delta} - \lambda_0)e^{-\beta\delta} & \text{if } \omega_{n-1} = 0. \end{cases} \tag{11}$$

In particular, if within $[(n-1)\delta, n\delta)$ no jumps occurred, then we have an exact expression for $\lambda_{n\delta}$. When instead a jump occurred, then the infimum of the approximating interval is the value that the intensity would have reached if the jump happened at $(n-1)\delta$, while the supremum is the one if the jump happened at $n\delta$. The approximation of $\lambda_{n\delta}$ is given with an absolute error smaller than $\alpha(1 - e^{-\beta\delta}) \approx \alpha\beta\delta$.

To approximate λ_t for $t > n\delta$, given $\lambda_{(n-1)\delta}$ and ω_{n-1}, we can use an analogous of (11). In this case, as t increases, the error can decay or not depending on two different situations: if no further jumps occur until t, then the given error decays; when instead a new jump occurs at $\tau' \in [n\delta, t)$ then the previous error, decayed until τ', increases by a new error because we do not know where exactly the jump occurred.

Jointly using (10) and (11), we obtain the following recursive relation, which is the main result of this part and concerns the forecasting of a jump in $[n\delta, (n+1)\delta)$.

Theorem 2. *Given* $\tilde{\lambda}_{n-1}$,

if $\omega_{n-1} = 1$, *then*

$$\tilde{\lambda}_n \in \left[1 - (1 - \tilde{\lambda}_{n-1})^{e^{-\beta\delta}} \exp\left((e^{-\beta\delta} - 1)\left(\delta\lambda_0 + \frac{\alpha e^{-\beta\delta}}{\beta}\right)\right), \tag{12}$$

$$1 - (1 - \tilde{\lambda}_{n-1})^{e^{-\beta\delta}} \exp\left((e^{-\beta\delta} - 1)\left(\delta\lambda_0 + \frac{\alpha}{\beta}\right)\right)\right); \tag{13}$$

if $\omega_{n-1} = 0$, $\quad \tilde{\lambda}_n = 1 - (1 - \tilde{\lambda}_{n-1})^{e^{-\beta\delta}} \exp\left((e^{-\beta\delta} - 1)\delta\lambda_0\right)$.

This result allows us to compute the probability of a jump in a given interval $[n\delta, (n+1)\delta)$ either given $\tilde{\lambda}_{n-1}$ or by subsequently computing $\tilde{\lambda}_j$, for $j = 0, ..., n-1$. In the latter case, the approximation error from $\tilde{\lambda}_{j-k}$ to $\tilde{\lambda}_j$ amplifies at each step if the number of occurred jumps within $[(j-k)n, jn)$ increases and if the jumps within $[(j-k)n, jn)$ are closer each other. In contrast, the error decreases if no jumps occurred.

From a practical point of view, the approximation error seems reasonable. For instance, under the exponential Hawkes model for the estimated JPM jump times we obtain that the amplitude of the approximating interval is about 0.015 (parameters of the global maximizer, first line for model 1 in Table 1) or 1.1×10^{-7} (parameters of the local maximizer). For this computation, we substituted $\delta E[\lambda_\infty] = \delta\lambda_0/(1-\alpha/\beta)$ for $\tilde{\lambda}_{n-1}$ in the above formula, and considered the case $\omega_{n-1} = 1$.

The recursion given in Theorem 2 (also rephrased at Remark 6 in the Appendix) is implementable but cannot be expressed in closed form. To allow an easier implementation, for the next formula we used the rougher approximations $\tilde{\lambda}_n \approx \delta\lambda_{n\delta}$, $\delta\lambda_{(n-1)\delta} \approx \tilde{\lambda}_{n-1}$ and formula (11). The result that we obtained is rougher but is more useful to prove an approximation for $\tilde{\lambda}_n$ (following from (15) below) which in turn is needed to reach the second aim of this section, namely the bounds (16) for the probability of occurrence of k consecutive jumps.

Remark 1. The following recursive approximation for $\tilde{\lambda}_n$ holds true:

$$\tilde{\lambda}_n \begin{cases} \in \left[\tilde{\lambda}_0 + (\tilde{\lambda}_{n-1} - \tilde{\lambda}_0 + \alpha\delta)e^{-\beta\delta}, \tilde{\lambda}_0 + (\tilde{\lambda}_{n-1} - \tilde{\lambda}_0)e^{-\beta\delta} + \alpha\delta\right) & \text{if } \omega_{n-1} = 1 \\ = \tilde{\lambda}_0 + (\tilde{\lambda}_{n-1} - \tilde{\lambda}_0)e^{-\beta\delta} & \text{if } \omega_{n-1} = 0 \end{cases} \tag{14}$$

The amplitude of the approximating interval in (14) is 0.037 with the parameters of the global maximizer and 1.2×10^{-7} with those of the local maximizer.

5.1 Probability of a cluster$^\circ$ and of a jump immediately after

We use here the terminology *consecutive jumps*, meaning that within an interval $[n\delta, (n+k)\delta)$ we estimated the occurrence of one jump on each one of the k subintervals $[i\delta, (i+1)\delta), i = n, \ldots, n+k-1$. We also define *cluster*$^\circ$ as a group of consecutive jumps. We extend the previous results of the Section with the aim of obtaining the two conditional probabilities: of the occurrence of k consecutive jumps; and of the arrival of a further jump after a cluster$^\circ$. Other than giving us practical indications on the risk of future jumps, these results are also useful to verify whether the statistical properties of our data match those of the model. For the empirical application to our JPM data, the check is done in Table 2.

We first compute the following relation between $\bar{\lambda}_{(n+k)\delta}$ and $\lambda_{n\delta}$ (see the proof in Appendix).

Proposition 4. Given $\lambda_{n\delta}$ and conditionally on $\{\omega_i = 1 \ \forall \ i = n, \ldots, n+k-1\}$, for $k = 1, 2, ...$ we have

$$\bar{\lambda}_{(n+k)\delta} \in \left[\lambda_0 + e^{-\beta\delta k}(\lambda_{n\delta} - \lambda_0) + \alpha \frac{1 - e^{-\beta\delta k}}{e^{\beta\delta} - 1}, \lambda_0 + e^{-\beta\delta k}(\lambda_{n\delta} - \lambda_0) + \alpha \frac{1 - e^{-\beta\delta k}}{1 - e^{-\beta\delta}} \right). \tag{15}$$

The second quantity of our interest in Section 5 is the conditional probability of k consecutive jumps, defined as

$$\tilde{\lambda}_{(n+k)} \doteq P(\text{ at least one jump in each } [i\delta, (i+1)\delta), i = n, ..., n+k-1 | \mathcal{F}_{n\delta-}).$$

This coincides with the probability of $\{\omega_i = 1, \ i = n, \ldots, n+k-1\}$ conditional on $\tilde{\lambda}_n$. In contrast from the case where we deal with the product of independent binomial distributions, here the single events $\{\omega_j = 1\}$ are dependent on the outcomes of the events $\{\omega_i = 1\}, i = n, ..., j-1$ preceding them.

Using the approximation $\lambda_{n\delta}\delta \approx \tilde{\lambda}_n$ for any n, and equations (14) and (15), we obtain that

$$\tilde{\lambda}_{(n+k)} = \tilde{\lambda}_n \tilde{\lambda}_{n+1|\{\omega_n=1\}} \cdots \tilde{\lambda}_{n+k-1|\{\omega_i=1, \ i=n,...,n+k-2\}} \in$$

$$\left[\prod_{i=0}^{k-1} \left(\tilde{\lambda}_0 + (\tilde{\lambda}_n - \tilde{\lambda}_0) e^{-\beta\delta i} + \alpha\delta e^{-\beta\delta} \frac{1 - e^{-\beta\delta i}}{1 - e^{-\beta\delta}} \right), \prod_{i=0}^{k-1} \left(\tilde{\lambda}_0 + (\tilde{\lambda}_n - \tilde{\lambda}_0) e^{-\beta\delta i} + \alpha\delta \frac{1 - e^{-\beta\delta i}}{1 - e^{-\beta\delta}} \right) \right). \tag{16}$$

Formula (16) is used to deliver Table 2, in the empirical application section, to approximate the probability of *at least* k consecutive jumps for the JPM asset prices; that is, the probability of the union of the sets $\{\ell$ consecutive jumps on $[n\delta, (n+\ell)\delta)\}$, with $\ell \geq k$.

A second direct consequence of (15) is that the conditional probability of a further jump after the occurrence of a cluster° of k jumps is obtainable, given $\tilde{\lambda}_n$ and for any integer $k \geq 1$. It is sufficient to apply the previous result and the relation

$$P(\omega_n = 1 | \omega_{n-1} = 1, ..., \omega_{n-k} = 1) = \frac{P(\omega_n = 1, \omega_{n-1} = 1, ..., \omega_{n-k} = 1)}{P(\omega_{n-1} = 1, ..., \omega_{n-k} = 1)}. \tag{17}$$

Note that the bounds furnished in Proposition 1 are different in that they do not explicitly account for the discreteness of the observations and the cluster can potentially have arbitrarily sparse jumps.

6 Application to financial data

The aim of this section is to exploit the theoretical results to obtain practical indications about jumps risks underlying our data. This section proceeds as follows: the jumps of the general model (1) are estimated (Section 6.2), the behavior of the detected jumps is investigated and reveals that

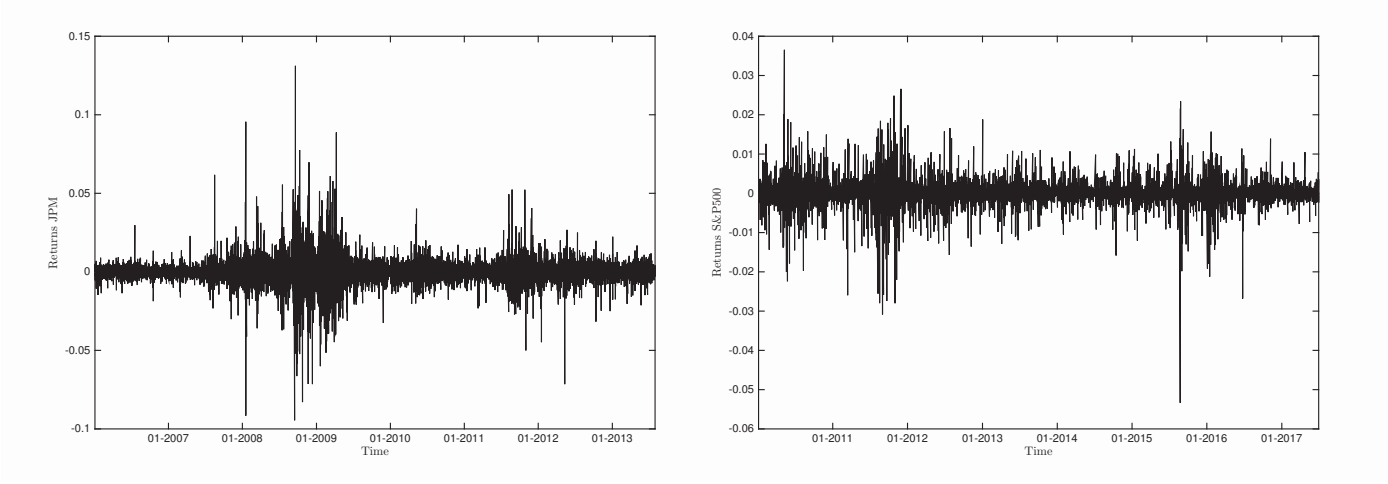

Figure 1: JPM (left-hand) and S&P500 (right-hand) returns

they are not independent but rather cluster significantly (Sec. 6.3). This motivates our use of a Hawkes model: in Section 6.4, after a comparison of three different types of Hawkes models, it is shown that the best fit is given by the one with one-term exponential kernel. Our theoretical results are then applied to quantify, under the best fitting model, the following risks: the risk of the occurrence of a cluster of consecutive jumps and the risk that an observed cluster is going to produce further jumps.

6.1 Our data

We considered the 5 minutes prices of the assets JPM, from 3/1/2006 to 31/7/2013, and S&P500, from 4/1/2010 to 30/6/2017: we have $\delta = 1/(252 \cdot 80)$ years. The price at each t_i is the one of the first transaction on $(t_{i-1}, t_i]$.

We cleaned the JPM data in the following way: in the case where the first observation of a day was zero, we eliminated it; in the case where we found a zero observation during a day, because that typically happened at the beginning of afternoons when the market was officially closed, we eliminated all the following observations in the day. The eliminated prices are 769 out of 152560, which is a negligible number.

The available S&P500 dataset is instead complete, with 153836 prices. Figure 1 shows the plots of the JPM and of the S&P500 log returns.

After having analyzed both the JPM dataset including the overnight and over weekend returns and the JPM dataset excluding them, we decided to focus on the first choice: in view of forecasting the occurrence of crashes the overnight and over weekend returns could be relevant because they

account for the reaction of the traders to the diffusion of information circulating after the closing and before the opening hours.

If a model $dX = b_t dt + \eta_t dW_t$ with continuous paths and cadlag coefficients, with η bounded from below by a positive constant, was adequate for our datasets, then the *standardized returns* $\frac{\Delta_i X}{\hat{\eta}_{t_i}\sqrt{\delta}}$ would have to display a more or less standard Gaussian behavior. We estimated spot η for both the JPM and the S&P500 datasets and in Figure 5 in the Appendix we plotted the estimates for η and the standardized returns in each case. The figure tells us that a stochastic volatility model with continuous paths does not fit well our data because the fluctuations of the standardized returns present amplitudes and picks which are incompatible with a standard Brownian motion. In fact the fluctuations of a standard Brownian motion are mostly included into the range $[-4.5, 4.5]$.

6.2 Filtering out the jumps

Under model (1), we estimated the times of jump by the times t_{i-1} such that $(\Delta_i X)^2$ is above the threshold $r_i(\delta) = 2\hat{\sigma}^2_{t_{i-1}}\delta \log\frac{1}{\delta}$, where $\hat{\sigma}_{t_{i-1}}$ is an estimator of spot σ at t_{i-1} and the function $2\delta \ln\frac{1}{\delta}$ is the squared modulus of continuity of the Brownian motion paths. Note that $(\Delta_i X)^2 > r_i(\delta)$ shows that the standardized return $\frac{\Delta_i X}{\hat{\sigma}_{t_i}\sqrt{\delta}}$ has an absolute value that is larger than the theoretical level $\sqrt{2\log\frac{1}{\delta}}$, which makes it incompatible with a standard Brownian motion. We denote by $\{\tau_\ell, \ell = 1, 2, ...N\}$ the estimated jump times. The threshold $r_i(\delta)$ has been shown to be optimal in minimizing the error when estimating the integrated variance $IV = \int_0^T \sigma_s^2 ds$ by the Truncated Realized Variance $\hat{IV} = \sum_{i=1}^m (\Delta_i X)^2 \mathbf{1}_{\{(\Delta_i X)^2 < r_i(\delta)\}}$. More precisely, when J only has finite jump activity (as it is the case for any semimartingale whose squared jumps sizes are larger than $2\underline{\sigma}^2\delta\log\frac{1}{\delta}$) then this choice minimizes both the Mean Squared Error $E[(\hat{IV} - IV)^2]$ and the Mean Squared Error conditional to the realized paths of σ and of J ([11]). To estimate spot σ, we chose a first step kernel-truncated estimator based on [18] and applied to the pre-truncated returns $\Delta_j X \cdot \mathbf{1}_{\{(\Delta_j X)^2 < 9\delta^{0.99}\}}$, as follows:

$$\tilde{\sigma}^2_{t_{i-1}} = \frac{\sum_{j=2}^m K_h(t_{j-1} - t_{i-1})(\Delta_j X)^2 \cdot \mathbf{1}_{\{(\Delta_j X)^2 \leq 9\delta^{0.99}\}}}{\delta \sum_{j=2}^m K_h(t_{j-1} - t_{i-1})}, \tag{18}$$

where $K_h(x) = \frac{1}{h}K\left(\frac{x}{h}\right)$, K is the doubly exponential kernel $K(x) = e^{-|x|}/2$, and the bandwidth is $h = 200\delta$. For each t_i, σ_{t_i} is estimated using a bilateral window of observations, where only those close to t_i have the most relevant weight[5]. The choice of the double exponential kernel is

[5]The highest value of $K_h(t_{j-1} - t_{i-1})$ is 100.8 reached for $t_{j-1} = t_{i-1}$. For $|t_{j-1} - t_{i-1}| > 200$ the weight is below 37.5 and for $|t_{j-1} - t_{i-1}| > 2000$ the weight is below 0.6.

motivated by [10] because it has been shown to be optimal in estimating the spot σ of a Brownian semimartingale, if we want to minimize the Mean Squared Error. [6] We also applied a second recursion by determining

$$\hat{\sigma}^2_{t_{i-1}} = \frac{\sum_{j=2}^{m} K_h(t_{j-1} - t_{i-1})(\Delta_j X)^2 \cdot \mathbf{1}_{\{(\Delta_j X)^2 \leq 2\tilde{\sigma}^2_{t_{i-1}} \delta \log \frac{1}{\delta}\}}}{\delta \sum_{j=2}^{m} K_h(t_{j-1} - t_{i-1})}, \tag{19}$$

however, the second recursion did not change the estimation of spot sigma. We then identified as jump times those t_{i-1} such that $(\Delta_i X)^2 > 2\hat{\sigma}^2_{t_{i-1}} \delta \log \frac{1}{\delta}$.

Remark 2. In our framework, the chosen name (*jump* or *large variation*) for a return above the threshold is irrelevant. What is crucial is that its occurrence is a special, unusual and sometimes dangerous event. We analyze/model the arrival of unusual events with the conviction that for some assets they are correlated.

We found 816 jumps for the JPM returns and 844 for the S&P500 returns. Figure 6 in the web Appendix shows the estimated spot σs, the estimated jumps and the standardized returns for both the datasets. The latter seem to be quite compatible with the increments of a standard Brownian motion.

6.3 Statistical analysis of the large variations

The aim of this section is to show that the estimated JPM jumps display persistency. In particular, we are interested in the possible presence of self-excitation, which is important for forecasting purposes. For this reason, we compared our empirical data with the most commonly used model for the arrival times of an asset jumps, namely the Poisson process. We performed three types of comparisons and we concluded that a Poisson process does not represent well the features of both our empirical datasets. However, while the statistical analysis of the S&P500 revealed that the estimated jumps do not display significant self-excitation, the JPM estimated jumps show features characterizing a Hawkes process, such as large clusters and dependence of the increments of their counter. This inspired us to follow the modeling idea in ([1]), as well as other papers in the literature, and considering Hawkes processes for the JPM jumps.

The mentioned comparisons with a Poisson model consist in: 1. checking the QQ-plot of the jump interarrivals in our data versus the ones of a Poisson process; 2. tabulating a comparison for

[6] The optimal bandwidth $h^{opt}_{t_i}$ suggested in [10] is different at each t_i : considering a Heston model for the truncated X, with parameters as in [26], on average $h^{opt}_{t_i}$ is about half the chosen h.

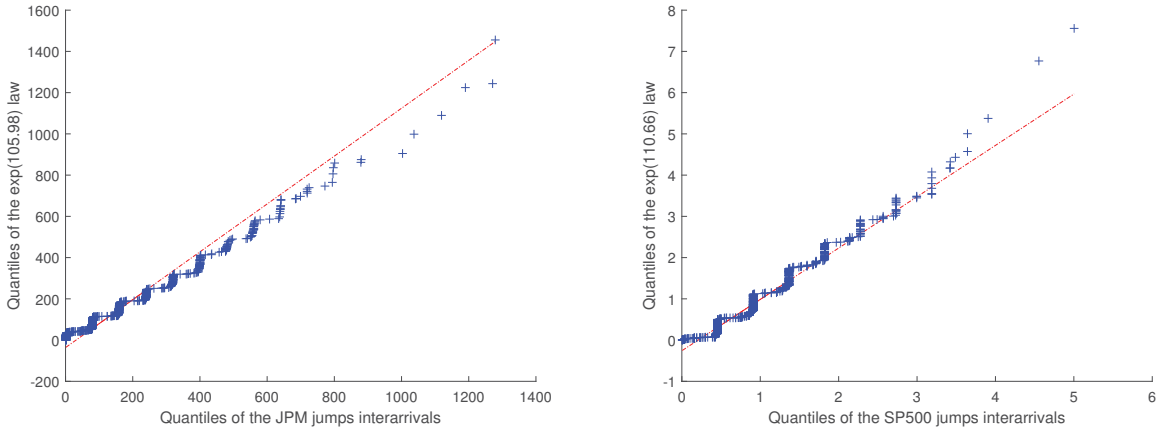

Figure 2: QQ-plots of the jump times *interarrivals* (temporal duration between two consecutive jump times) against the best fitting exponential laws. Left-hand: JPM jumps interarrivals against $\mathcal{E}(105.9823)$; right-hand: S&P500 jumps interarrivals against $\mathcal{E}(110.6058)$ (annual parameters).

the presence of close jumps in our data versus the presence of close jumps in a Poisson process; 3. checking statistically the dependence of the increments of the jumps counter N for our data.

1. Figure 2 shows, for each record of jumps, the QQ-plot of the empirical quantiles of the estimated jumps interarrivals against the theoretical quantiles of the best fitting exponential law. In annual units of measure, the parameter for the JPM data is 105.9823, that for the S&P500 data is 110.6058.

None of the datasets is compatible with a homogeneous Poisson process. For example, the frequent almost vertical segments appearing in the QQ-plot of the JPM data mean that many interarrivals have approximately the same length (for instance the first long vertical segment is at a level of about 80, in units of 5 minutes, meaning 1 trading day), which has no counterpart in the observations produced by an exponential law. Some consecutive jumps occur at a distance of 5 minutes, many others at a distance of 1 day, many others of two days and so on. This corresponds to the occurrence of jumps at the opening and at the closing of many trading days, where the volatility also increases. In other words, the turbulence of the prices at the beginning and at the end of many days generate jumps on top of the (already high) values of the volatility, thus exacerbating the usual daily U shaped volatility picture. However, we cannot attribute the large variations of the prices that we classified as jumps to the sole volatility because the standardized returns $\Delta_i X/(\hat{\eta}_{t_i}\sqrt{\delta})$ of the Brownian semimartingale model were shown to be incompatible with the increments of a Brownian motion.

For a comparison, the QQ-plot of the interarrival times of the 394 jumps obtained from the JPM dataset *without overnight returns* was similar in shape but with lower steps. This means that the frequency of the interarrivals of length multiple of 1 day is not due only to the opening and closing effects of each trading day. Moreover, we cannot assert that the presence of a jump at the beginning and at the end of a day is a deterministic phenomenon because it does not happen every day. We argue that the jumps at the beginning of the days are the result of a higher turbulence than usual, plausibly in response to some information that was circulating during the night or the day (or some days) before.

2. The average interarrival jump time for JPM is of 185.63 units of 5 minutes, corresponding to about 2.5 trading days. The jump interarrivals below the average point out the possible presence of clusters and Tables 4 and 5 in the web Appendix document that this is exactly the case.

3. Table 6 in the web Appendix clearly shows the existence of a significant dependence among the increments of the counter N of the estimated JPM jumps. That also is at odd with a Poisson model and in favor of a self-exciting model.

We now argue that clusters of consecutive jumps (precisely defined *clusters*$^\circ$ at Sec.5.1) have a significant forecasting potential for the JPM jumps. We implemented a regression of the occurrence of a jump at a time instant t on the occurrence of jumps at some previous times, and we found significant coefficients. More precisely we estimated through OLS the parameters of the regression $I_{t_i} = \beta + \alpha_1 I_{t_{i-1}} + \alpha_2 I_{t_{i-2}} + \alpha_3 I_{t_{i-3}} + \epsilon_{t_i}$, where t_i are the observation times and $I_{t_i} = \mathbf{1}_{(\Delta_i X)^2 > r_i(\delta)}$. We found a low R^2, but we caught the dependence feature leading us to consider Hawkes models.

6.4 Estimation of the parameters and practical indications

We performed maximum likelihood estimation for the parameters of a Hawkes process modeling the jump times τ_ℓ of our datasets obtained as described in Section 6.2. The details of the estimation are given in Appendix.

For the JPM jump times, we compared the results obtained with three models: 1. a model with three parameters, having a single exponential function kernel ($P = 1$); 2. a model with five parameters, with kernel having two exponential terms ($P = 2$); 3. a model with four parameters, having a single power function kernel. Model 3 is implemented for a comparison with the approach in [4]. In all cases we imposed the positivity of the parameters. Namely, for models 1 and 2 we adopted the constraints $\lambda_0 \geq 0, \alpha_i \geq 0$, $\beta_i > 0, \forall i$; we also imposed the asymptotic stationarity

condition $\alpha_i < \beta_i, \forall i$ [7]. For model 3 we set the constraints $\alpha, \beta > 0$ and, for the stationarity, $\alpha < \gamma$. Table 1 reports the MLE for each model. The standard error of each estimate is reported in brackets.

Table 1: ML estimates of the parameters of the following three Hawkes models fitting the JPM jump times: 1. single exponential term kernel; 2. kernel with two exponential terms ($P = 2$); 3. single power term kernel. The standard errors are reported within brackets.

ker	λ_0	α_1	β_1	α_2	β_2	$\log \mathcal{L}$
1. 1 exp	91.41	1711.29	10'931.23			3186.07
	(3.71)	(253.77)	(1481.30)			
1. 1 exp	53.27	4.72	9.14			3020.92
	(13.26)	(2.22)	(12.34)			
2. 2 exp	58.86	1706.83	11'000.7	1.42	4.57	3190.20
	(13.33)	(256.24)	(1521.11)	(34.16)	(341.19)	
2. 2 exp	91.41	1711.28	10'931.21	1.05×10^{-7}	208.15	3186.07
	(5.86)	(249.27)	(2730.70)	(96.23)	(1022.66)	
ker	λ_0	α	β	γ		$\log \mathcal{L}$
3. power	91.41	1164.49	1.471	7430.25		3186.1
	(3.47)	(113.10)	(0.18)	(17.74)		

For models 1 and 2, besides the global maximizer, we reported a local maximizer of $\mathcal{L}$ (second line of each model in Table 1). The two sets of the local and the global maximizers fit the same record of jumps but classify them differently. Model 1 with the parameters of the local maximizer has a lower λ_0 than in the case of the global maximizer: the former identifies more jumps as descendants and fewer jumps as fathers. In fact, even if the intensity jumps are much smaller (α is low), their effect is much more persistent (β is small). This combination tends to keep the intensity higher than λ_0 for much more time. With respect to the local maximizer, the global one delivers a Hawkes model which is close to a Poisson process. This is because even if the increase α of λ is large when a jump occurs, the very high β coefficient reduces the persistency. We argue later that the Hawkes model with the parameters of the local maximizer gives a better fit of the JPM jumps.

In both the cases of local and global maximizers, the estimated α is significantly different from zero. Thus, we can conclude that the jumps of the JPM stock in the given period displayed

[7] However, the found maximizers vary negligibly if we remove the constraint $\alpha < \beta$.

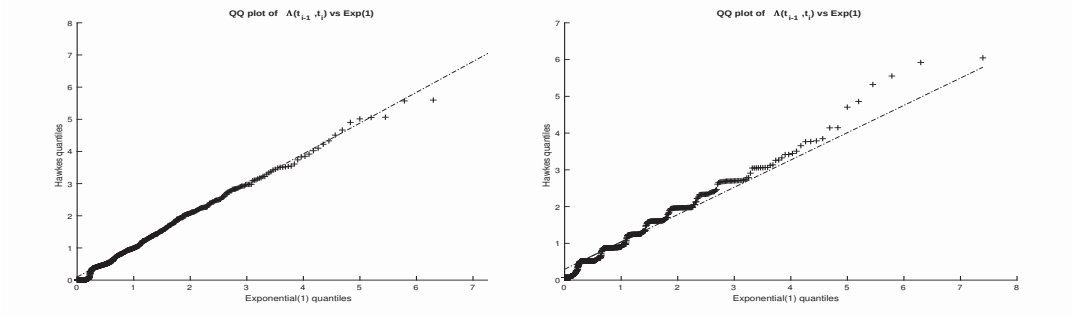

Figure 3: Quantiles of the values $\int_{\tau_{j-1}}^{\tau_j} \lambda_t dt$ against quantiles of 816 values assumed by an $\mathcal{E}(1)$ random variable: (left) for model 1, with one exponential term kernel, parameters of the local maximizer; (right) for model 3, with one power term kernel, parameters of the unique maximizer. The models are fitted by ML to the estimated jump times of the JPM returns.

a clustering behavior which is caught by a Hawkes model. For the global maximizer, all the parameters are strongly significantly different from zero, while for the local maximizer only $\hat{\lambda}_0$ is strongly significant. The proximity to zero of the confidence interval for $\hat{\alpha}$ makes the estimation of β much more uncertain than in the previous case.

As for model 2, the local maximizer reported in Table 1 has values of $\lambda_0, \alpha_1, \beta_1$ similar to the ones of the global maximizer for model 1, while $\hat{\alpha}_2$ is not significantly different from 0, and therefore the estimated $\hat{\beta}_2$ has no tangible impact on λ. In fact, adding more kernels does not improve the description of the data. However, according to the QQ-plot analysis that we mention later on, not even model 2 with the parameters of the global maximizer of $\log \mathcal{L}$ improves the fit.

For model 3, we only found a global maximum value for the log likelihood, but the QQ-plot analysis showed that the fit is not good.

Any Hawkes process N with $\lambda_0 > 0$ is transformed in a simple Poisson process $PP(1)$ with parameter 1 through the time change $\Lambda_t = \int_0^t \lambda_s ds$. ¡more precisely, N is such that $\{N_{\Lambda_t}\}_{t\geq 0} \sim PP(1)$, and thus the *transformed jumps interarrivals* $\Lambda_{\tau_j} - \Lambda_{\tau_{j-1}}$ are independent exponential random variables with parameter 1; that is, $\int_{\tau_{j-1}}^{\tau_j} \lambda_t dt \sim \mathcal{E}(1)$. From the analysis, for each one of the above models, of the obtained QQ-plot of the values $\int_{\tau_{j-1}}^{\tau_j} \lambda_t dt$ against a record of independent values assumed by an $\mathcal{E}(1)$ random variable, we found that the exponential Hawkes model (model 1) with the parameters of the local maximizer is the best. We show its QQ-plot in Figure 3, left-hand panel.

The initial step of such a QQ-plot indicates that there are almost no data for values of the jumps interarrival times above 5 minutes and below 1 day, or less data than the model would produce. Instead, the high number of interarrivals of length 1 day allows the model to fall into

line with the data. This means that the model misses some features of the data, however it represents a remarkable improvement over the Poisson model. The self-excitation feature of the exponential Hawkes model captures much better many relevant characteristics of our empirical data, as confirmed also in Table 2. Thus the chosen model has the advantage of providing easily implementable formulas delivering much more realistic probabilistic measures of the risk of jumps. Only, we have to keep in mind that the model predicts more couples of consecutive jumps having interarrival time below 1 day than displayed by the data, while the model predicts less couples having interarrivals around 1 day. Thus, we have to remember that the probability that a cluster is going to produce a further jump, computed from such a model, is an overestimate if the further jump is close, while it is an underestimate if the jump has a distance of 1 day. Similarly, note that for values of the interarrivals in the range [6, 7.5] years there are less data than the model would produce.

The QQ-plots obtained under either model 1 with the parameters of the global maximizer, or under model 2 or under model 3, reported in Table 1, are similar to the one in Figure 3, right-hand. In particular, in contrast to what happens in [4] for Bayer and Djind data, our QQ-plot under model 3 shows that, for modeling the sole counter of the large variations of the JPM prices, a Hawkes process with power kernel does not fit the jump times well. [8]

For a comparison, based on the dataset of the estimated S&P500 jump times, the MLE for an exponential Hawkes model is $\hat{\lambda}_0 = 110.6061$ $(SE = 9.32), \hat{\alpha} = 3.78 \times 10^{-5}$ $(2 \times 10^5), \hat{\beta} = 2.2564$ (4×10^{10}), which implies $\log \mathcal{L} = 3127.84$. The negligibility of $\hat{\alpha}$ indicates that, the data are incompatible with a Poisson model (Figure 2, right), but also with a self-exciting exponential Hawkes process. The QQ-plot of the transformed jump interarrivals under this model against iid $\mathcal{E}(1)$ rvs is similar to the one plotted in Figure 3, right, but has even more pronounced steps. The MLE of the parameters of a Hawkes model having power kernel similarly indicates the absence of the self-excitation feature in the S&P500 jumps $\Big(\hat{\lambda}_0 = 110.606$ $(86.482), \hat{\alpha} = 2.914 \times 10^{-5}$ $(0.4897), \hat{\beta} = 0.054$ (0.0125), $\hat{\gamma} = 33.149(11.214), \log \mathcal{L} = 3127.8\Big)$. Note that our observation period for the index S&P500 excludes the 2008 credit crisis. For this index and in the considered period, we cannot use the obtained properties of an exponential Hawkes model to have information on the arrival of

[8] Interestingly, a further element in favor of the chosen fitting model is given in [12], where some measures of discrepancy between paths generated by different data generating processes are provided. It turns out that the discrepancy of the jump times we estimated from our JPM data versus the best fitting Poisson process is substantially the same of the one of the jump times of a simulated exponential Hawkes process with parameters $\lambda_0 = 53.27, \alpha = 4.72, \beta = 9.14$.

future jumps, given the observation of previous jump clusters.

6.5 Practical indications

In the remaining part of this section, we illustrate the practical consequences of the main theoretical results of this paper. Namely, we show a quantification of the conditional probability $\tilde{\lambda}_{(n+k)}$, of the occurrence of a cluster$^\circ$ of consecutive jumps, and of the probability $P\{T_{i+k} < S + t_\epsilon(S) | N_S - N_U = k, \lambda_U\}$ that a given cluster$^+$ (constituted by jumps occurring when λ is at a high level) is not finished. The first conditional probability is an application of (16), which follows from Theorem 2. The second conditional probability is an application of Proposition 1. We mainly focus on the results offered by the model giving the best QQ-plot for our dataset of the JPM estimated jumps, which is the Hawkes model with one exponential term kernel and parameters of the local maximizer of the log likelihood.

Table 2 quantifies the approximation that we can give of the conditional probability of *at least* k consecutive jumps in the JPM prices; that is, the probability of the union of the sets $\{\ell$ consecutive jumps on $[n\delta, (n+\ell)\delta)\}$, as $\ell \geq k$. The first and second columns are computed using (16). The conditioning is to $\tilde{\lambda}_n$, which is approximated by $\delta\lambda_{n\delta}$, and $\lambda_{n\delta}$ is taken within the specified interval. For a comparison, we also look at the results obtained with the global maximizer parameters (second panel).

With the parameters of the local (respectively global) maximizer, the most frequent values of λ_t, 62 times out of 816, fall within $[98.5, 101.5]$ (resp. $[91.4, 91.41]$). Thus in order to compute the lower bounds we fixed $\lambda_{n\delta} = 98.5$ (resp. $\lambda_{n\delta} = 91.4$), while for the upper bounds $\lambda_{n\delta} = 101.5$ (resp. $\lambda_{n\delta} = 91.41$).

The third column of the Table offers a comparison with the empirical probabilities (*EmpirP*) and the last column gives the mean empirical probabilities obtained on 100 simulated paths of the Hawkes process with the parameters specified in each panel.

In the first panel we see that for $k = 1$ the bounds, the empirical data and the empirical probability computed on the simulated data are all consistent. For $k = 2$ the empirical probability computed on the simulated data is in line with the bounds, which underestimate the empirical probability computed on the real data. However, among the 100 simulated paths we can find some paths giving values similar to the ones in column 3. The empirical probabilities for large values of k are zero due to the narrow size of the dataset. This is consistent with the corresponding bounds because a probability around 10^{-6} means that to see a cluster of k consecutive jumps, we

Table 2: Bounds (16) for the probability that the JPM prices have at *least* k consecutive jumps (starting from $n\delta$) conditional to $\tilde{\lambda}_n$. We approximate $\tilde{\lambda}_n \approx \lambda_{n\delta}\delta$ for $n > 0$, while take $\tilde{\lambda}_0$ as in (10). $\lambda_{n\delta}$ is taken within the specified interval. *First panel*: exponential Hawkes model with parameters of the local maximizer for the likelihood function. *Second panel*: same model but with the parameters of the global maximizer.

k	Lower bound	Upper bound	EmpirP, data	EmpirP, simuls
local maximizer $\lambda_0 = 53.27, \alpha = 4.72, \beta = 9.14; \lambda_{n\delta} \in [98.5, 101.5]$				
1	4.89×10^{-3}	5.03×10^{-3}	5.28×10^{-3}	5.01×10^{-3}
2	2.50×10^{-5}	2.65×10^{-5}	3.41×10^{-4}	2.82×10^{-5}
3	1.34×10^{-7}	1.46×10^{-7}	0	3.31×10^{-7}
4	7.47×10^{-10}	8.36×10^{-10}	0	0
5	4.35×10^{-12}	4.99×10^{-12}	0	0
6	2.63×10^{-14}	3.09×10^{-14}	0	0
7	1.65×10^{-16}	1.99×10^{-16}	0	0
8	1.08×10^{-18}	1.33×10^{-18}	0	0
9	7.26×10^{-21}	9.14×10^{-21}	0	0
10	5.07×10^{-23}	6.51×10^{-23}	0	0
global maximizer $\lambda_0 = 91.4, \alpha = 1709.7, \beta = 10851.2; \lambda_{n\delta} \in [91.4, 91.41]$				
1	4.53×10^{-3}	4.53×10^{-3}	4.40×10^{-3}	4.30×10^{-3}
2	2.45×10^{-4}	4.05×10^{-4}	5.13×10^{-4}	2.40×10^{-5}
3	2.03×10^{-5}	5.62×10^{-5}	2.70×10^{-5}	1.65×10^{-6}
4	2.03×10^{-6}	9.43×10^{-6}	0	1.10×10^{-7}
5	2.22×10^{-7}	1.74×10^{-6}	0	3.47×10^{-8}
6	2.57×10^{-8}	3.39×10^{-7}	0	0
7	3.05×10^{-9}	6.78×10^{-8}	0	0
8	3.69×10^{-10}	3.68×10^{-8}	0	0
9	4.48×10^{-11}	2.84×10^{-9}	0	0
10	5.49×10^{-12}	5.86×10^{-10}	0	0

need on average to have at least 10^6 observations, while the number of observations at our disposal ($m = 151792$) is less than 10^6.

In contrast, in the second panel, for $k \geq 2$ the bounds for the conditional probabilities are in line with the empirical probabilities computed on the real data. This is consistent with the fact that the likelihood is higher but the bounds overestimate the empirical probabilities computed on the simulated data. In this case, all of the 100 simulated paths show values very close to the reported mean empirical probabilities. For $k = 1$ again all the columns give consistent results.

Our interpretation is that the model with the parameters of the global maximizer seems less reliable on finite samples because, even if the bounds catch the empirical probabilities of k consecutive jumps implied in the data, the empirical probabilities on simulated data are not consistent and the quantiles fit (QQ-plot) is not very good. In contrast, the model with the parameters of the local maximizer is more consistent with the simulated data, it produces some paths for which all the columns of the Table are consistent and the quantiles fit is much better. The situation seems similar to when we estimate a quantity with a consistent estimator but we find that on finite samples the estimator multiplied by a correction term performs better.

For both the models, we also computed the bounds using a higher conditioning value of $\lambda_{n\delta}$, and both the lower and the upper bounds turn out to be much higher, which is in line with the fact that the leading term of the probability of one jump in an interval is proportional to $\lambda_{n\delta}$. The comparisons of the empirical probabilities computed either on our JPM data or on the simulated data and the consistency with the theoretical bounds are similar.

Table 3 reports the conditional probability that a cluster$^+$ is not yet finished by the observation time S, that is, the following jump is still a consequence of the same causes having produced the cluster$^+$. As in Proposition 1, we assume to know that a cluster of k jumps at times $\tau_i, \ldots, \tau_{i+k-1}$ occurred within a time interval $[U, S = U + s)$. We argued that the following jump at T_{i+k} is part of the same cluster$^+$ iff $T_{i+k} < S + t_\epsilon(S)$. The results are produced under the exponential Hawkes model, for the JPM estimated jumps, with the parameters of the local maximizer of log $\mathcal{L}$, and with $\epsilon = 0.01$. The table shows the lower bound (LB) for $P\{T_{i+k} < S + t_\epsilon(S)\}$ given by $P\{T_{i+k} < S + \underline{t}_\epsilon\}$ and the upper bound (UB) given by $P\{T_{i+k} < S + \bar{t}_\epsilon\}$. The conditioning level $\lambda_U = 55$ (quite close to λ_0) is where the decay can better be caught (otherwise the probabilities do not vary from 1).

From table 3, we can see the following. As expected, the higher is the number of jumps in the cluster$^+$, the higher is the probability that the following jump is still a product of the same

Table 3: Probability that a cluster$^+$ of k jumps produces a further jump. All the values "1" below are in fact approximated to 1 with an error less than 10^{-7}. Exponential Hawkes model with parameters of the local maximizer.

k	1	2	3	4	5	6	7	8	9
$\lambda_U = 55$, $s = 80\delta$ (1 day)									
LB	0.9999997	1	1	1	1	1	1	1	1
UB	0.9999997	1	1	1	1	1	1	1	1
$\lambda_U = 55$, $s = 400\delta$ (one week)									
LB	0,9999991	1	1	1	1	1	1	1	1
UB	0,9999997	1	1	1	1	1	1	1	1
$\lambda_U = 55$, $s = 0,25$ years (3 months)									
LB	0,7074953	0,9886529	0,9986173	0,9997115	0,9999176	0,9999711	0,9999883	0,9999947	0,9999974
UB	0,9999985	1	1	1	1	1	1	1	1

increment in the intensity λ that caused the cluster$^+$. The closer is the time horizon S to U, the higher is the probability. This happens because the cluster accumulates closer jumps, which generate a larger increase in the intensity of jump.

Remark 3. Reliability check on simulations. The theoretical results presented in Sec.4.2 require that the parameters that qualify the Hawkes process are known. Actually, we do not know the true values of $\lambda_0,\ \alpha$ and β but we estimate them from the jump times which in turn are identified starting from the observed asset prices. The two measurement errors could deliver a relevant bias in our estimate of $P \doteq P\{T_{i+k} < S + t_\epsilon(S)\}$. To check for this we implemented the following simulation exercise. A dataset of log prices $X_{t_i},\ t_i = i\delta,\ \delta = 1/(252 \times 80),\ i = 1..n,\ n = 150'000$, $T = n\delta$, is obtained through discretization (Euler scheme) of the path of a jump-diffusion process evolving in continuous time by

$$dX_t = -\frac{\sigma_t^2}{2}dt + \sigma_t dW_t + dJ_t \tag{20}$$

$$\sigma_t = e^{v_t}, \quad dv_t = 0.09\big(\log(0.25) - v_t\big)\ dt + 0.05\ dW_t^v, \quad v_0 = \log(0.3),$$

with: $d[W, W^v]_t = -0.7dt$, jump part $J_t = \sum_{j=1}^{N_t} Z_j$ independent on (W, W^v); N exponential Hawkes process with the same parameters $\lambda_0 = 53.27, \alpha = 4.72, \beta = 9.14$ that we estimated for the JPM data; and Z_j iid $N(0, 0.03^2)$. The continuous martingale part of X is as in [7] (p. 792), and similar to [15] (model SV1FJ), while the variance of the jump sizes has been chosen so as the generated paths of X resemble the one of the JPM log-price at our disposal. Such a variance is particularly low with respect to values normally used in the literature simulations studies, but we

purposedly took it to make the jumps identification more difficult so to stress the stability of our results. Further details are in Appendix. We simulated 500 paths, for each path we estimated spot σ and the jump times through the method described in Sec.6.2. We also implemented a second recursion to improve the jump times identification. Then using the estimated jump times we implemented MLE for the exponential Hawkes parameters. We found that the average percentage estimation error of the Hawkes kernel parameters λ_0, α, β is about 19%, 40% and 54%, respectively. Nevertheless, with $\epsilon = 0.01$ and two different choices for (U, S), i.e. (0.18, 0.215) and (0.02, 0.07), by substituting the estimated parameters for the true ones in P, LB and UB, we obtained that the percentage errors in identifying P and the true bounds is of order $10^{-3}, 5 \times 10^{-3}$ and 6×10^{-4} respectively. Moreover, the estimation of the bounds is really close to the true P, with a percentage error of order 5×10^{-3} for $\hat{LB}$ and 6×10^{-4} for $\hat{UB}$.

The numbers reported in table 3 tell us that for any $k = 1, ..., 9$ the probability, implied by our data, that an observed cluster of jump did not exhaust yet is very high (always above 70%). In fact, the selected model displays a high degree of persistency in the values of the intensity (low β). Under the model with the parameters of the global maximizer, the probabilities (not shown) are lower because the parameter β is much higher. However, the probabilities are still important because they vary from 5% to 70%. This conclusion may be a useful indication when studying contagion phenomena and for portfolio management.

7 Conclusions

Many relevant movements of financial assets prices are not isolated episodes but are rather preceded by exceptional price instabilities. Indeed a cluster of jumps can be interpreted as an anticipatory signal of a future extreme event.

Supposing a jump diffusion model for the log-price of an asset, the price variations are classified as jumps, or large variations, if they are above a given threshold, namely they are incompatible with a Brownian semimartingale. In this model, we have two independent sources of risk: volatility and jump risks. We focus on the jumps (or, more precisely, on the jumps above the threshold), and the counting process of their estimated arrival times is modeled according to a Hawkes process.

Motivated by the results of our empirical analysis, we investigate the probabilistic properties of the exponential Hawkes process. In particular, we aim to measure the probability of occurrence of a cluster of jumps and the residual length of an observed cluster. This is clearly of interest for an

investor who is managing her portfolio risk.

Through the newly defined intensity *decay instant*, we compute bounds for the probability that the next jump following a cluster will belong to it, rather than having been generated by an independent cause. For the exponential Hawkes process, we prove the stochastic increasingness between consecutive interarrival times, strengthening their positive correlation. Moreover, by explicitly accounting for the error in locating the jump times induced by the discreteness of our data, we furnish bounds for the probability of the occurrence of future consecutive jumps.

Finally, we apply our formulas to the counting process of the estimated jump arrivals for a record of JPM 5 minutes prices. We find that the large variations that occurred display dependence and clusters, and the best fitting Hawkes model is exponential. We show that the empirical probability of the occurrence of a cluster of many jumps is in some cases enormously higher for the JPM prices than for the best fitting Poisson model, while it is quite compatible with the theoretical bounds obtained for an exponential Hawkes process. We find that, under the best fitting exponential Hawkes model, the probability that an observed cluster of jumps did not exhaust yet is very high, mostly close to 1.

In contrast, we found that the S&P500 index jumps during the observation period do not display self-excitation, which is not that surprising considering that our observation period for the index S&P500 excludes the 2008 crash period.

This paper confirms that for some financial assets the exponential Hawkes model reveals the presence of self-excitation among the estimated jumps, thus the formulas that we provide represent an important improvement in measuring the risk of jumps in asset prices and can be relevant for a variety of problems in economics and in finance.

Acknowledgements: we thank Giacomo Bormetti, Francesco Calvori, Fabrizio Cipollini, Giampiero M. Gallo, Jean Jacod, Fabrizio Lillo, Roberto Renó, Simone Scotti, Peter Tankov and Nakahiro Yoshida for their precious comments and suggestions. This project benefited from funds given by the INdAM-GNAMPA association of Italian mathematicians, sited in Rome, and from the Europlace Institute of Finance of the Institut Louis Bachelier, in Paris.

References

[1] Yacine Aït-Sahalia, Julio Cacho-Diaz, and Roger J.A. Laeven. Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, 117(3):585–606,

2015.

[2] E. Bacry, I. Mastromatteo, and J.F. Muzy. Hawkes processes in finance. *Market Microstructure and Liquidity*, 1(1):59, 2015.

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

[4] V. Chavez-Demoulin, A.C. Davison, and A.J. McNeil. Estimating value-at-risk: a point process approach. *Quant. Finance*, 5(2):227–234, 2005.

[5] Y. Chen. Likelihood Function for Multivariate Hawkes Processes. *Preprint (5 pages) available on https://www.math.fsu.edu/ ychen/research/HawkesLikelihood.pdf*, 2016.

[6] S. Clinet and N. Yoshida. Statistical inference for ergodic point processes and application to Limit Order Book. *Stochastic Processes and their Applications*, 127:1800–1839, 2017.

[7] R. Cont and C. Mancini. Nonparametric tests for pathwise properties of semimartingales. *Bernoulli*, 17(2):781–813, 2011.

[8] Daryl J. Daley and David Vere-Jones. *An introduction to the theory of point processes. Vol. II: General theory and structure.* New York, NY: Springer, 2nd revised and extended edition, 2008.

[9] Jun Pan Durrel Duffie and Kenneth Singleton. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376, 2000.

[10] J. E. Figueroa-López and C. Li. Optimal Kernel Estimation of Spot Volatility of Stochastic Differential Equations. *Preprint 2018 (33 pages) available on https://pages.wustl.edu/files/pages/imce/figueroa/kernelestimationofspotvolatility_0618.pdf.*

[11] José E. Figueroa-López and Cecilia Mancini. Optimum thresholding using mean and conditional mean squared error. *Journal of Econometrics*, 208(1):179–210, 2019.

[12] R. Foschi. Measuring discrepancies between Poisson and exponential Hawkes processes. *Available at SSRN*, 2019.

[13] Lawrence R Glosten and Paul R. Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of financial economics*, 14(1):71–100, 1985.

[14] Sanford Grossman. On the efficiency of competitive stock markets where trades have diverse information. *The Journal of finance*, 31(2):573–585, 1976.

[15] X. Huang and G. Tauchen. The relative contribution of jumps to total price variance. *Journal of Financial Econometrics*, 3(4):456–499, 2005.

[16] H. Joe. *Multivariate Models and Multivariate Dependence Concepts*. New York: Chapman and Hall/CRCr, 1997.

[17] C. Mancini. Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scand. J. Stat.*, 36(2):270–296, 2009.

[18] C. Mancini, V. Mattiussi, and R. Renó. Spot volatility estimation using delta sequences. *Finance Stoch.*, 19:261–293, 2015.

[19] H. Manner O. Grothe, V. Korniichuk. Modeling multivariate extreme events usign self-exciting point processes. *Journal of econometrics*, 182(2):269–289, 2014.

[20] Yosihiko Ogata. The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Ann. Inst. Stat. Math.*, 30:243–261, 1978.

[21] Yosihiko Ogata. On Lewis Simulation Method for Point Processes. *IEEE Transactions on Information Theory*, IT-27(1):23–31, 1981.

[22] L. Lin P. Embrechts, T. Liniger. Multivariate hawkes processes: an application to financial data. *Journal of applied probability*, 48(A):367–378, 2011.

[23] M. Rambaldi, P. Pennesi, and F. Lillo. Modeling foreign exchange market activity around macroeconomic news: Hawkes-process approach. *Phys. Rev. E*, 91(1):012819 – 012833, 2015.

[24] A. Rangan and R.E. Grace. A non-Markov model for the optimum replacement of self-repairing systems subject to shocks. *J. Appl. Probab.*, 25(2):375–382, 1988.

[25] J.G. Rasmussen. Lecture Notes: Temporal Point Processes and the Conditional Intensity Function. *http://arxiv.org/abs/1806.00221*, 2011.

[26] L. Zhang, P. Mykland, and Y. Ait-Sahalia. A tale of two time scales. *Journal of the American Statistical Association*, 100:472, 2005.

[27] Jiancang Zhuang, Yosihiko Ogata, and David Vere-Jones. Stochastic declustering of space-time earthquake occurrences. *J. Am. Stat. Assoc.*, 97(458):369–380, 2002.

A Web Appendix

A.1 Preliminary remarks

Remark. From [24] we have that

$$P(T_2 = t_2 | T_1 = t_1) =$$

$$\lim_{dt_2 \to 0} \frac{1}{dt_2} P\big\{\text{jump in } [t_2, t_2 + dt_2) | T_1 = t_1, \text{no jumps in } (t_1, t_2)\big\} P\big\{\text{no jumps in } (t_1, t_2) | T_1 = t_1\big\}$$

$$= \lambda_{t_2} \exp\left(- \int_{t_1}^{t_2} \lambda_s ds \right). \tag{A.21}$$

Note that considering $P(T_2 = t_2 | T_1 = t_1)$ means that on (t_1, t_2) there are no jumps, because the first jump occurs at t_1 and the *second* jump occurs at t_2. Thus, $\mathcal{F}_{t_1} = \mathcal{F}_{t_2 -}$, and the right-hand side of (A.21) is $\mathcal{F}_{t_1}$ measurable, other than $\mathcal{F}_{t_2 -}$ measurable, which is what we would expect for $P(T_2 = t_2 | T_1 = t_1)$. □

By (A.21) we can compute the finite-dimensional distribution of the process,

$$P(T_1 = t_1, ..., T_n = t_n) = \lambda_{t_1} \cdots \lambda_{t_n} \exp\left(- \int_0^{t_n} \lambda_t dt \right),$$

and, by changing variables, the distribution of the first n interarrival times,

$$P(T_1 = s_1, T_2 - T_1 = s_2, ..., T_n - T_{n-1} = s_n) = \lambda_{s_1} \cdots \lambda_{s_1 + \cdots + s_n} \exp\left(- \int_0^{s_1 + \cdots + s_n} \lambda_t dt \right).$$

Note that because the last quantity is not factorisable, this proves that the interarrival times for a Hawkes process with non-constant λ are not independent. By making further explicit the joint distribution, we find that they are not even exchangeable.

We remark that for $n > 2$ we have

$$P(T_n = t_n | T_{n-1} = t_{n-1}) = \bar{\lambda}_{t_n} \exp\left(- \int_{t_{n-1}}^{t_n} \bar{\lambda}_{t | \{T_{n-1} = t_{n-1}\}} dt \right),$$

which preserves the same form of Eq. (A.21) only if expressed in terms of the unconditional intensity. In contrast, the expression of $P(T_n = t_n | T_{n-1} = t_{n-1})$ in terms of the conditional intensity $\lambda_t = E[\bar{\lambda}_t | \mathcal{F}_{t-}]$ is more complicated because λ_t requires conditioning on the history until t. In this case, we have

$$P(T_n = t_n | T_{n-1} = t_{n-1}) =$$

$$\int P(T_n = t_n | T_1 = t_1, ..., T_{n-1} = t_{n-1}) \cdot P(T_1 = t_1, ..., T_{n-2} = t_{n-2} | T_{n-1} = t_{n-1}) d(t_1, ..., t_{n-2})$$

$$= \int_{\mathbf{\Delta}_{n-2}\cap(0,t_{n-1})^{n-2}} \lambda_{t_n} \exp\left(-\int_{t_{n-1}}^{t_n} \lambda_t dt\right)$$
$$\cdot \frac{\lambda_{t_1}\cdots\lambda_{t_{n-1}} \exp\left(-\int_0^{t_{n-1}} \lambda_t dt\right)}{\int_{\mathbf{\Delta}_{n-2}\cap(0,t_{n-1})^{n-2}} \lambda_{t_1}\cdots\lambda_{t_{n-1}} \exp\left(-\int_0^{t_{n-1}} \lambda_t dt\right) d(t_1,...,t_{n-2})} d(t_1,...,t_{n-2}),$$

where $\mathbf{\Delta}_m$ denotes the simplex $\{\mathbf{t} \in \mathbb{R}_+^m | t_1 < t_2 < ... < t_m\}$ of $\mathbb{R}^m$. We also have

$$P(N_t = n) =$$
$$\int_{\mathbf{\Delta}_n\cap[0,t]^n} P\left(\text{no jumps in } (t_n, t] | T_1 = t_1, ..., T_n = t_n\right) P(T_1 = t_1, ..., T_n = t_n) d(t_1, ..., t_n)$$
$$= \int_{\mathbf{\Delta}_n\cap[0,t]^n} \lambda_{t_1}\cdots\lambda_{t_n} \exp\left(-\int_0^t \lambda_s ds\right) d(t_1, ..., t_n). \quad \square$$

A.2 Recursive formula for λ_t in terms of $\lambda_{\tau_{i-1}}$

For the exponential Hawkes model with kernel $\Phi(x) = \alpha e^{-\beta x}$, for $t \in (\tau_{i-1}, \tau_i]$ it holds that

$$\lambda_t = \lambda_0 + [\lambda_{\tau_{i-1}} - \lambda_0 + \alpha] e^{-\beta(t-\tau_{i-1})}; \tag{A.22}$$

while for $S \in (\tau_{i-1}, \tau_i)$ and $t \in [S, \tau_i]$ it holds that

$$\lambda_t = \lambda_0 + (\lambda_S - \lambda_0) e^{-\beta(t-S)}. \tag{A.23}$$

It follows that the increments $\Delta_{\tau_i}\Lambda$ of the integrated λ are given by

$$\Delta_{\tau_i}\Lambda = \Lambda(\tau_i) - \Lambda(\tau_{i-1}) = \lambda_0(\tau_i - \tau_{i-1}) + \frac{1}{\beta}[\lambda_{\tau_{i-1}} - \lambda_0 + \alpha]\left[1 - e^{-\beta(\tau_i - \tau_{i-1})}\right]. \tag{A.24}$$

A.3 Bounds for λ, Section 4.1

Proof of Lemma 1. Consider the case where the n jumps on $(0, t]$ all occurred within $(0, \eta)$. The lower bound $\lambda_0 + \alpha n e^{-\beta t}$ is a trivial consequence of the fact that $e^{\beta T_i} > 1$ for all i. Meanwhile, if $T_i < \eta$ for all i, then

$$\lambda_t = \lambda_0 + \alpha \sum_{k=1}^{n} e^{-\beta(t-T_i)} < \lambda_0 + \alpha n e^{-\beta(t-\eta)}.$$

Analogously, while $e^{-\beta(t-T_k)} < 1$ holds in general, if the n jumps on $(0, t]$ all occurred within $(t - \eta, t]$, then $t - T_i < \eta$, so $e^{-\beta(t-T_i)} > e^{-\beta\eta}$ and the stated lower bound follows. $\square$

Remark 4 (Lemma 1)**.** Item 1 describes the case when the n jumps all occurred at the beginning of the observation period $[0, t)$. When η is close to 0, the previously mentioned upper bound $\lambda_0 + \alpha n$

is not sharp. Item 2 describes the case when the n jumps have all occurred at the end of the observation period; however, this time, when η is close to 0, the upper bound $\lambda_0 + \alpha n$ is accurate.

Lemma 1 also allows to assess the probability that one or more jumps occur after a given observation period, namely $P(N_{t+dt} - N_t = 1 | N_t - N_{t-a} = k)$ and $P(N_{t+a} - N_t = m | N_t - N_{t-a} = k)$, because these two probabilities may be computed as $P(N_{t+b} - N_t = m)$, starting from the intensity $\bar{\lambda}_{t|N_t - N_{t-a} = k}$. E.g. $P(N_{t+dt} - N_t = 1 | N_t - N_{t-a} = k) = \bar{\lambda}_{t|N_t - N_{t-a} = k} dt + o(dt)$.

A.4 Mean length of a cluster

Let us fix a tolerance $\varepsilon^\star$. Given the occurrence of a jump at τ_n, we define *decay time* of $\bar{\lambda}$ the time $s_{\varepsilon^\star}$ needed to the increase $\alpha e^{-\beta(t-\tau_n)}$, experienced by the intensity, to become negligible. More precisely, we define $s_{\varepsilon^\star} := t_{\varepsilon^\star} - \tau_n$, where $t_{\varepsilon^\star} := \inf\{t > \tau_n : \alpha e^{-\beta(t-\tau_n)} < \varepsilon^\star\}$. As expected, the time of decay is increasing in α and decreasing in β because $s_{\varepsilon^\star} = \dfrac{1}{\beta} \log\left(\dfrac{\alpha}{\varepsilon^\star}\right)$. Moreover, we have the same time of decay for couples (α, β) and (α', β') such that $\frac{\beta'}{\beta} = \frac{\log(\alpha'/\varepsilon^\star)}{\log(\alpha/\varepsilon^\star)}$.

We then define *cluster** the set of descendants of τ_n, intended as the set of the jump times provoked by the increase of $\bar{\lambda}$ at τ_n. Let us define also the *length of the cluster** as the length $\ell_{\varepsilon^\star} \doteq b - \tau_n$ of the time interval which on average contains the $100 \cdot (1 - \varepsilon^\star)\%$ of the elements of the cluster*. As shown below,

$$\ell_{\varepsilon^\star} = s_{\varepsilon^\star} - \frac{\log(\alpha)}{\beta}. \tag{A.25}$$

So, as expected, the smaller $s_{\varepsilon^\star}$, the shorter any cluster* is.

To have a concrete idea, with $\varepsilon^\star = 5\%$ the exponential Hawkes process fitting our JPM dataset, with the parameters of the global maximizer of $\mathcal{L}$, delivers a time of decay equal to $s_{\varepsilon^\star} \approx 0,000955$ years, corresponding to 1.6 hours. The length of the interval containing on average the 95% of each cluster* is $\ell_{\varepsilon^\star} \approx 0,000274$ years, which means that a cluster* almost exhausted after about 27 minutes. With the parameters of the local maximizer we would obtain $s_{\varepsilon^\star} \approx 0,498$ years; that is, 125 days, and $\ell_{\varepsilon^\star} \approx 0,328$ years, 83 days.

Proof of (A.25). $\ell_{\varepsilon^\star}$ is obtained from the condition

$$E[N^{(\tau_n)}([\tau_n, b] - \tau_n)] / E[N^{(\tau_n)}([\tau_n, +\infty) - \tau_n)] = 1 - \varepsilon^\star,$$

where $N^{(\tau_n)}$ is the part of N only counting the direct descendants of τ_n. Indicating by $\mu^{(\tau_n)}$ the mean measure of $N^{(\tau_n)}$, we have

$$E[N^{(\tau_n)}([\tau_n, b] - \tau_n)] / E[N^{(\tau_n)}([\tau_n, +\infty) - \tau_n)] = \mu^{(\tau_n)}([\tau_n, b] - \tau_n) / \mu^{(\tau_n)}([\tau_n, +\infty) - \tau_n),$$

and

$$\mu^{(\tau_n)}([\tau_n,+\infty)-\tau_n)=\int_{\tau_n}^{+\infty}\alpha e^{-\beta(x-\tau_n)}dx=\int_0^{+\infty}\alpha e^{-\beta z}dz=\frac{\alpha}{\beta},$$

$$\mu^{(\tau_n)}([\tau_n,b]-\tau_n)=\int_{\tau_n}^{b}\alpha e^{-\beta(x-\tau_n)}dx=\int_0^{b-\tau_n}\alpha e^{-\beta z}dz=\frac{\alpha}{\beta}(1-e^{-\beta(b-\tau_n)}),$$

thus $E[N^{(\tau_n)}([\tau_n,b]-\tau_n)]/E[N^{(\tau_n)}([\tau_n,+\infty)-\tau_n)]=1-\varepsilon^\star$ iff $\mu^{(\tau_n)}([\tau_n,b]-\tau_n)=\frac{\alpha}{\beta}(1-\varepsilon^\star)$, iff $e^{-\beta(b-\tau_n)}=\varepsilon^\star$, iff $b=\frac{1}{\beta}\log\left(\frac{1}{\varepsilon^\star}\right)+\tau_n$.

It follows that $\ell_{\varepsilon^\star}\doteq\frac{1}{\beta}\log\left(\frac{1}{\varepsilon^\star}\right)$, and thus $\ell_{\varepsilon^\star}=s_{\varepsilon^\star}-\frac{\log(\alpha)}{\beta}$. □

A.5 Decay *instant* of $\bar{\lambda}$, Section4.2

The given definition of $t_\epsilon(S)$ generalizes the previous one of $t_{\varepsilon^\star}$. Consistently, the concept of cluster is extended by the one of cluster$^+$. In the latter case, the small distance ϵ of λ from λ_0 is taken in percentage of λ_0 rather than as an absolute quantity. Note that the decay instant $t_\epsilon(\tau_1)$ of the first jump of the Hawkes process coincides with $t_{\varepsilon^\star}$ if, rather than at λ, we look at $\lambda-\lambda_0$ and $\varepsilon^\star=\lambda_0\epsilon$.

Proof that $t_\epsilon(S)\in(\underline{t}_\epsilon,\bar{t}_\epsilon)$, with $\underline{t}_\epsilon,\bar{t}_\epsilon$ as in (4). We first reach the following recursive relation between λ_U and λ_S, we then establish the inequalities.

We have $\lambda_U-\lambda_0=\alpha\sum_{T_\ell<U}e^{-\beta(U-T_\ell)}$, thus

$$\begin{aligned}\lambda_S&=\lambda_0+\alpha\sum_{T_\ell<S}e^{-\beta(S-T_\ell)}\\&=\lambda_0+e^{-\beta(S-U)}\alpha\sum_{T_\ell<U}e^{-\beta(U-T_\ell)}+\alpha\sum_{U\le T_\ell<S}e^{-\beta(S-T_\ell)}\\&=\lambda_0+e^{-\beta s}(\lambda_U-\lambda_0)+\alpha\sum_{U\le T_\ell<S}e^{-\beta(S-T_\ell)}.\end{aligned}\tag{A.26}$$

The cluster observed on $[U,S)$ has k jumps, and obviously for any jump of the cluster we have $T_\ell\ge U$, so the third term above satisfies

$$k\alpha e^{-\beta s}=k\alpha e^{-\beta(S-U)}\le\alpha\sum_{U\le T_\ell<S}e^{-\beta(S-T_\ell)}\le k\alpha.$$

Note that the right extreme $k\alpha$ would be obtained if all the jumps of the cluster were concentrated at S ($T_\ell=S$ for any T_ℓ of the cluster), which is the limit for $\eta\to0$ of case 2 in Lemma 1. Analogously, the left extreme $k\alpha e^{-\beta(S-U)}$ would be obtained if all the jumps were concentrated at U, which is the limit of case 1 in the Lemma.

It follows that

$$\lambda_S-\lambda_0\in\Big(e^{-\beta s}(\lambda_U-\lambda_0)+k\alpha e^{-\beta s},\quad e^{-\beta s}(\lambda_U-\lambda_0)+k\alpha\Big).$$

Because S was assumed not to be a jump time then $t_\epsilon(S) = \frac{1}{\beta}\log\left(\frac{\lambda_S-\lambda_0}{\epsilon\lambda_0}\right)$, and the stated inequality is straightforward. □

Proof of Proposition 1. We have $P(T_{i+k} < U+s+t_\epsilon(S)) =$

$$\int_0^{t_\epsilon(S)} \bar{\lambda}_{U+s+t|\{\text{no jumps in } [U+s,U+s+t)\}} \exp\left(-\int_0^t \bar{\lambda}_{U+s+\tau|\{\text{no jumps in } [U+s,U+s+\tau)\}} d\tau\right) dt. \quad \text{(A.27)}$$

Similarly as in (A.26), with $S+\tau$ in place of S, we have

$$\bar{\lambda}_{U+s+\tau} = \lambda_0 + (\lambda_U - \lambda_0)e^{-\beta(s+\tau)} + \alpha \sum_{T_\ell \in [U,U+s+\tau)} e^{-\beta(U+s+\tau-T_\ell)},$$

and if there are no jumps within $(U+s, U+s+\tau]$, then within $[U, U+s+\tau)$ we still have k jumps, and the last sum has k terms.

When all the k jumps in the cluster within $[U,S)$ happen at U, then $T_\ell = U$ for all ℓ and $t_\epsilon(S) = \underline{t}_\epsilon$, thus $\bar{\lambda}_{U+s+\tau|\{\text{no jumps in } [U+s,U+s+\tau)\}} = \lambda_0 + (\alpha k + \lambda_U - \lambda_0)e^{-\beta(s+\tau)}$, and we reach

$$P(T_{i+k} < U+s+\underline{t}_\epsilon) = \int_0^{\underline{t}_\epsilon} \left(\lambda_0 + (\alpha k + \lambda_U - \lambda_0)e^{-\beta(s+t)}\right) \exp\left(-\int_0^t \lambda_0 + (\alpha k + \lambda_U - \lambda_0)e^{-\beta(s+\tau)} d\tau\right) dt$$

$$= \int_0^{\underline{t}_\epsilon} (\lambda_0 + (\alpha k + \lambda_U - \lambda_0)e^{-\beta(s+t)}) \exp(-\frac{1}{\beta}(e^{-\beta s}(\alpha k + \lambda_U - \lambda_0) - e^{-\beta(s+t)}(\alpha k + \lambda_U - \lambda_0) + \lambda_0 t\beta))$$

$$= e^{-\frac{e^{-\beta s}}{\beta}(\alpha k + \lambda_U)} \left[-\left((\alpha k + \lambda_U - \lambda_0)\frac{e^{-\beta s}}{\lambda_0 \epsilon}\right)^{-\frac{\lambda_0}{\beta}} e^{\frac{\lambda_0}{\beta}(e^{-\beta s}+\epsilon)} + e^{\frac{e^{-\beta s}}{\beta}(\alpha k + \lambda_U)}\right].$$

If instead all the k jumps occurred at the end of the interval $[U, U+s)$, then $T_\ell = U+s$ for all ℓ and $t_\epsilon(S)$ coincides with $\bar{t}_\epsilon$, so $\bar{\lambda}_{U+s+\tau|\{\text{no jumps in } [U+s,U+s+\tau)\}} = \lambda_0 + (\lambda_U - \lambda_0)e^{-\beta(s+\tau)} + k\alpha e^{-\beta\tau}$, and we reach that $P(T_{i+k} < U+s+\bar{t}_\epsilon)$ coincides with

$$\int_0^{\bar{t}_\epsilon} (\lambda_0 + (\alpha k + (\lambda_U - \lambda_0)e^{-\beta s})e^{-\beta t}) \exp\left(-\int_0^t \lambda_0 + (\alpha k + (\lambda_U - \lambda_0)e^{-\beta s})e^{-\beta\tau} d\tau\right) dt$$

$$= \int_0^{\bar{t}_\epsilon} \left(\lambda_0 + (\alpha k + \lambda_U - \lambda_0)e^{-\beta t}\right) \exp\left(\frac{1}{\beta}\Big((\lambda_U - \lambda_0)e^{-\beta(s+t)} + e^{-\beta t}\alpha k - \lambda_0 t\beta - e^{-\beta s}(\lambda_U - \lambda_0) - k\alpha\Big)\right)$$

$$= 1 - \exp\left(-\frac{1}{\beta}\Big(-(\lambda_U - \lambda_0 + \alpha k)e^{-\beta(s+\bar{t}_\epsilon)} + \lambda_0 \bar{t}_\epsilon \beta + e^{-\beta s}(\lambda_U - \lambda_0) + k\alpha\Big)\right)$$

$$= \exp\left(-\frac{e^{-\beta s}\lambda_U + k\alpha}{\beta}\right)\left(-\left((e^{\beta s}\alpha k + \lambda_U - \lambda_0)\frac{e^{-\beta s}}{\lambda_0 \epsilon}\right)^{-\frac{\lambda_0}{\beta}} \exp\left(\frac{\lambda_0}{\beta}(e^{-\beta s}+\epsilon)\right) + \exp\left(\frac{e^{-\beta s}\lambda_U + k\alpha}{\beta}\right)\right).$$

As for the monotonicity, for $P(T_{i+k} < U+s+\underline{t}_\epsilon)$ we have:

$$\frac{d}{ds}P(T_{i+k} < U + s + \underline{t}_\epsilon) = -\exp\left(-e^{-\beta s}\frac{\alpha k + \lambda_U}{\beta}\right)\left(\frac{\alpha k + \lambda_U - \lambda_0}{\lambda_0 \epsilon}e^{-\beta s}\right)^{-\frac{\lambda_0}{\beta}} \cdot \cdot \exp\left(\frac{\lambda_0}{\beta}(e^{-\beta s} + \epsilon)\right)(e^{-\beta s}(\alpha k + \lambda_U - \lambda_0) + \lambda_0),$$

which is negative, as expected. Furthermore,

$$\frac{d}{dk}P(T_{i+k} < U + s + \underline{t}_\epsilon) = \alpha \exp\left(-e^{-\beta s}\frac{\alpha k + \lambda_U}{\beta}\right)\left(\frac{\alpha k + \lambda_U - \lambda_0}{\lambda_0 \epsilon}e^{-\beta s}\right)^{-\frac{\lambda_0}{\beta}} \cdot \exp\left(\frac{\lambda_0}{\beta}(e^{-\beta s} + \epsilon)\right)\frac{e^{-\beta s}(\alpha k + \lambda_U - \lambda_0) + \lambda_0}{\beta(\alpha k + \lambda_U - \lambda_0)}$$

is strictly positive as soon as $\alpha > 0$. Finally,

$$\frac{d}{ds}P(T_{i+k} < U + s + \bar{t}_\epsilon) = -\exp\left(-\frac{e^{-\beta s}\lambda_U + k\alpha}{\beta}\right)\left((e^{\beta s}\alpha k + \lambda_U - \lambda_0)\frac{e^{-\beta s}}{\lambda_0 \epsilon}\right)^{-\frac{\lambda_0}{\beta}} \cdot \cdot \exp\left(\frac{\lambda_0}{\beta}(e^{-\beta s} + \epsilon)\right)(\lambda_U - \lambda_0) \cdot \frac{\alpha k + e^{-\beta s}(\lambda_U - \lambda_0) + \lambda_0}{e^{-\beta s}\alpha k + \lambda_U - \lambda_0} < 0,$$

$$\frac{d}{dk}P(T_{i+k} < U + s + \bar{t}_\epsilon) = \alpha \exp\left(-\frac{1}{\beta}(e^{-\beta s}\lambda_U + k\alpha)\right)\left((e^{\beta s}\alpha k + \lambda_U - \lambda_0)\frac{e^{-\beta s}}{\lambda_0 \epsilon}\right)^{-\frac{\lambda_0}{\beta}} \cdot \cdot \exp\left(\lambda_0 \frac{e^{-\beta s} + \epsilon}{\beta}\right) \cdot \frac{e^{\beta s}\alpha k + \lambda_0 e^{\beta s} + \lambda_U - \lambda_0}{(e^{\beta s}\alpha k + \lambda_U - \lambda_0)\beta} > 0. \quad \square$$

Proof of Proposition 2. If there are no jumps within $[S, S+t)$, then $\bar{\lambda}_{S+t} = \lambda_0 + (\lambda_S - \lambda_0)e^{-\beta t}$, and with $U + s = S$ (A.27) becomes

$$\int_0^{t_\epsilon(S)} \left[\lambda_0 + (\lambda_S - \lambda_0)e^{-\beta t}\right] e^{-\int_0^t \lambda_0 + (\lambda_S - \lambda_0)e^{-\beta v}dv} dt =$$

$$\int_0^{t_\epsilon(S)} \left[\lambda_0 + (\lambda_S - \lambda_0)e^{-\beta t}\right] e^{-\lambda_0 t + (\lambda_S - \lambda_0)\frac{e^{-\beta t} - 1}{\beta}} dt.$$

With $c = \lambda_S - \lambda_0$ and $b = t_\epsilon(S)$ the above display equals

$$1 - e^{\frac{ce^{-b\beta}}{\beta}} e^{-b\lambda_0 - \frac{c}{\beta}}.$$

Since $e^{-b\beta} = \frac{\lambda_0 \varepsilon}{\lambda_S - \lambda_0}$ and $e^{-b\lambda_0} = \left(\frac{\lambda_0 \varepsilon}{\lambda_S - \lambda_0}\right)^{\frac{\lambda_0}{\beta}}$ then

$$P = 1 - e^{\frac{\lambda_S - \lambda_0}{\beta}\frac{\lambda_0 \varepsilon}{\lambda_S - \lambda_0}}\left(\frac{\lambda_0 \varepsilon}{\lambda_S - \lambda_0}\right)^{\frac{\lambda_0}{\beta}} e^{-\frac{\lambda_S - \lambda_0}{\beta}} = 1 - e^{-\frac{\lambda_S - \lambda_0(1+\varepsilon)}{\beta}}\left(\frac{\lambda_0 \varepsilon}{\lambda_S - \lambda_0}\right)^{\frac{\lambda_0}{\beta}}.$$

Note that under $\lambda_S > \lambda_0(1+\epsilon)$ we have $\lambda_S > \lambda_0$, so that $P < 1$. Further, we have both $e^{-\frac{\lambda_S - \lambda_0(1+\varepsilon)}{\beta}} < 1$ and $\frac{\lambda_0 \varepsilon}{\lambda_S - \lambda_0} < 1$, so $P > 0$. $\square$

A.6 Duration between consecutive jump times, Section 4.3

A.6.1 Proofs

Proof of Proposition 3. First, we prove (8). Note that conditioning on $\bar{\lambda}_{T_n} = \ell$ does not lead us to know the value of the r.v. T_n. However, if we also condition on the event "no jumps occurred within $(T_n, T_n + t)$", by the recursive formula, we obtain

$$\bar{\lambda}_{T_n+t|\{\bar{\lambda}_{T_n}=\ell,\text{ no jumps in }(T_n,T_n+t)\}} = \lambda_0 + (\ell - \lambda_0 + \alpha)e^{-\beta t}.$$

Thus we can write

$$P(T_{n+1} - T_n = t|\bar{\lambda}_{T_n} = \ell) =$$

$$P(\text{jump at } T_n + t|\text{no jumps in } (T_n, T_n + t),\ \bar{\lambda}_{T_n} = \ell)P(\text{no jumps in } (T_n, T_n + t)|\bar{\lambda}_{T_n} = \ell)$$

$$= \bar{\lambda}_{T_n+t|\{\text{no jumps in }(T_n,T_n+t),\ \bar{\lambda}_{T_n}=\ell\}} \exp\left(-\int_{T_n}^{T_n+t} \bar{\lambda}_{s|\{\bar{\lambda}_{T_n}=\ell,\text{ no jumps in }(T_n,s)\}} ds\right)$$

$$= \left(\lambda_0 + (\ell - \lambda_0 + \alpha)e^{-\beta t}\right) \exp\left(-\int_0^t \left(\lambda_0 + (\ell - \lambda_0 + \alpha)e^{-\beta\sigma}\right) d\sigma\right)$$

$$= \left(\lambda_0 + (\ell - \lambda_0 + \alpha)e^{-\beta t}\right) \exp\left(\frac{1}{\beta}(e^{-\beta t} - 1)(\ell - \lambda_0 + \alpha) - \lambda_0 t\right), \tag{A.28}$$

as we claimed.

Finally, the equality

$$P(T_{n+1} - T_n > \tau|\bar{\lambda}_{T_n} = \ell) = \exp\left(\frac{1}{\beta}(e^{-\tau\beta}(\ell + \alpha - \lambda_0) - \beta\lambda_0\tau - \ell - \alpha + \lambda_0)\right)$$

is obtained by integrating in t the expression (A.28). □

Proof of Corollary 1. By differentiating the expression in (9) wrt ℓ, we reach

$$\frac{d}{d\ell}P(T_{n+1} - T_n > \tau|\bar{\lambda}_{T_n} = \ell) = \frac{1}{\beta}(e^{-\tau\beta} - 1) \exp\left(e^{-\tau\beta}(\ell - \lambda_0 + \alpha) - \beta\lambda_0\tau - \ell - \alpha + \lambda_0\right),$$

which is negative $\forall\tau > 0$. □

Remark on Corollary 1. The conclusion of Corollary 1 could seem to be obvious in view of the fact that $P\{\text{no jumps in } (T_n, T_{n+1})\} = e^{-\int_{T_n}^{T_{n+1}} \bar{\lambda}_u du}$ and one could think to approximate $e^{-\int_{T_n}^{T_{n+1}} \bar{\lambda}_u du}$ by $e^{-(T_{n+1}-T_n)\bar{\lambda}_{T_n}}$. However, this is not correct in this framework for two reasons: $\bar{\lambda}$ is cáglád, thus, if $T_{n+1} - T_n$ was small, for $t \in (T_n, T_{n+1}]$ the values assumed by $\bar{\lambda}_t$ would be closer

to $\bar{\lambda}_{T_n+} = \bar{\lambda}_{T_n} + \alpha$ rather than to $\bar{\lambda}_{T_n}$; moreover not necessarily $T_{n+1} - T_n$ is small, thus the decay of $\bar{\lambda}_t$, from the value $\bar{\lambda}_{T_n} + \alpha$ reached on T_n+, can lead the values $\bar{\lambda}_u$ far away from $\bar{\lambda}_{T_n}$.

Proof of Theorem 1. For brevity, we omit the conditioning on $\bar{\lambda}_{T_{n-1}} = \ell$, which has to be understood. Let us write j, js for indicating "jump", "jumps", respectively.
We first compute $P(T_{n+1} - T_n = t | T_n - T_{n-1} = s)$ because, after that, obtaining $P(T_{n+1} - T_n > \tau | T_n - T_{n-1} = s)$ is straightforward through the relation

$$P(T_{n+1} - T_n > \tau | T_n - T_{n-1} = s) = \int_\tau^{+\infty} P(T_{n+1} - T_n = t | T_n - T_{n-1} = s) dt.$$

Note that

$$P(T_{n+1} - T_n = t | T_n - T_{n-1} = s) =$$

$$P(\text{j at } T_{n-1} + s + t, \text{no js on } (T_{n-1} + s, T_{n-1} + s + t) | \text{j at } T_{n-1} + s, \text{no js in } (T_{n-1}, T_{n-1} + s))$$

$$= P\Big(\text{j at } T_{n-1} + s + t | \text{no js on } (T_{n-1} + s, T_{n-1} + s + t) \cup (T_{n-1}, T_{n-1} + s),\ \text{j at } T_{n-1} + s\Big) \cdot$$

$$P\Big(\text{no js in } (T_{n-1} + s, T_{n-1} + s + t) \Big| \text{j at } T_{n-1} + s, \text{no js in } (T_{n-1}, T_{n-1} + s)\Big). \tag{A.29}$$

In view of these conditioning events, which have to be understood along the whole proof, we can use the recursive formula (A.22) to compute $\bar{\lambda}$ at the following different times:

- $\bar{\lambda}_{T_{n-1}+s} = \lambda_0 + (\bar{\lambda}_{T_{n-1}} - \lambda_0 + \alpha) e^{-\beta s}$;
- $\bar{\lambda}_{T_{n-1}+s+t} = \lambda_0 + [\bar{\lambda}_{T_{n-1}+s} - \lambda_0 + \alpha] e^{-\beta t} = \lambda_0 + [(\bar{\lambda}_{T_{n-1}} - \lambda_0 + \alpha) e^{-\beta s} + \alpha] e^{-\beta t}$;
- for $\tau \in (T_{n-1} + s, T_{n-1} + s + t)$, we have $\bar{\lambda}_\tau = \lambda_0 + [\bar{\lambda}_{T_{n-1}+s} - \lambda_0 + \alpha] e^{-\beta(\tau - T_{n-1} - s)}$
  $= \lambda_0 + [(\bar{\lambda}_{T_{n-1}} - \lambda_0 + \alpha) e^{-\beta s} + \alpha] e^{-\beta(\tau - T_{n-1} - s)}$. Changing variable by $\sigma = \tau - T_{n-1} - s$, we have $\sigma \in (0, t)$, and we obtain $\bar{\lambda}_\tau = \lambda_0 + [(\bar{\lambda}_{T_{n-1}} - \lambda_0 + \alpha) e^{-\beta s} + \alpha] e^{-\beta \sigma}$.

Using these expressions, from Eq. (A.29) we obtain

$$P(T_{n+1} - T_n = t | T_n - T_{n-1} = s) = \bar{\lambda}_{T_{n-1}+s+t} \exp\left(- \int_{T_{n-1}+s}^{T_{n-1}+s+t} \bar{\lambda}_\tau d\tau \right)$$

$$= \Big\{ \lambda_0 + \Big[(\bar{\lambda}_{T_{n-1}} - \lambda_0 + \alpha) e^{-\beta s} + \alpha \Big] e^{-\beta t} \Big\} \exp\left\{ -\lambda_0 t - \int_0^t \Big[(\bar{\lambda}_{T_{n-1}} - \lambda_0 + \alpha) e^{-\beta s} + \alpha \Big] e^{-\beta \sigma} d\sigma \right\}$$

$$= \Big\{ \lambda_0 + \Big[(\bar{\lambda}_{T_{n-1}} - \lambda_0 + \alpha) e^{-\beta s} + \alpha \Big] e^{-\beta t} \Big\} e^{-\lambda_0 t} \exp\left\{ \frac{1}{\beta} \Big[(\bar{\lambda}_{T_{n-1}} - \lambda_0 + \alpha) e^{-\beta s} + \alpha \Big] (e^{-\beta t} - 1) \right\} \tag{A.30}$$

We compute now

$$P(T_{n+1}-T_n>\tau|T_n-T_{n-1}=s)=\int_\tau^{+\infty}P(T_{n+1}-T_n=t|T_n-T_{n-1}=s)dt$$

$$=\exp\left(\frac{1}{\beta}(\bar\lambda_{T_{n-1}}e^{-\beta(s+\tau)}+e^{-\beta(s+\tau)}\alpha-e^{-\beta(s+\tau)}\lambda_0-\lambda_0\tau\beta-e^{-\beta s}\bar\lambda_{T_{n-1}}+e^{-\beta\tau}\alpha-e^{-\beta s}\alpha+e^{-\beta s}\lambda_0-\alpha)\right)$$

$$-\lim_{x\to+\infty}\exp\left(-\frac{1}{\beta}(-\bar\lambda_{T_{n-1}}e^{-\beta(s+x)}-\alpha e^{-\beta(s+x)}+e^{-\beta(s+x)}\lambda_0+\lambda_0x\beta+e^{-\beta s}\bar\lambda_{T_{n-1}}+e^{-\beta s}\alpha-e^{-\beta s}\lambda_0-\alpha e^{-\beta x}+\alpha)\right)$$

$$=e^{-\lambda_0\tau}\exp\left(\frac{e^{-\beta\tau}-1}{\beta}\Big(e^{-\beta s}(\bar\lambda_{T_{n-1}}-\lambda_0+\alpha)+\alpha\Big)\right). \tag{A.31}$$

In order to study the monotonicity wrt s, we compute

$$\frac{d}{ds}P\Big\{T_{n+1}-T_n>\tau|T_n-T_{n-1}=s\Big\}=$$

$$\Big(e^{-\beta s}-e^{-\beta(s+\tau)}\Big)\Big(\bar\lambda_{T_{n-1}}-\lambda_0+\alpha\Big)e^{-\lambda_0\tau}\exp\Big\{\frac{e^{-\beta\tau}-1}{\beta}\Big(e^{-\beta s}(\bar\lambda_{T_{n-1}}-\lambda_0+\alpha)+\alpha\Big)\Big\},$$

which is greater than 0 for any $s,\tau>0$. □

A.6.2 Figure 4

Figure 4, left-hand, shows the copula of the joint law of $T_{n+1}-T_n$ (on the vertical axis) and T_n-T_{n-1} (on the horizontal axis), conditional on $\bar\lambda_{T_{n-1}}\in(99,102)$. The T_i are the jump times estimated on the JPM dataset and then modeled by an exponential Hawkes process with parameters given by the local maximizer ($\lambda_0=53.27,\alpha=4.72,\beta=9.14$) of the likelihood function $\mathcal{L}$ (the local maximum value of $\log\mathcal{L}$ is $3'020.92$). The unconditional intensity $\bar\lambda$ assumes the 8% (66 out of 816) of its values within $(99,102)$, while in any other intervals with the same length $\bar\lambda$ assumes less values. The observed positive dependence is sensible.

In contrast, with the parameters $\lambda_0=91.41,\alpha=1711.92,\beta=10'931.23$ of the global maximizer of $\mathcal{L}$ (the global maximum value of $\log\mathcal{L}$ is approximately 3186.07), the positive dependence of the interarrival times is present but weaker because the value of β is very high. For a comparison, we remark that on simulated data produced by the Hawkes model with the global maximizer parameters, the copula of the joint law of $T_{n+1}-T_n$ and T_n-T_{n-1} conditional to $\bar\lambda_{T_{n-1}}\in(99,102)$ displays a similarly weaker positive dependence, the copula producing a picture similar to Figure 4, right.

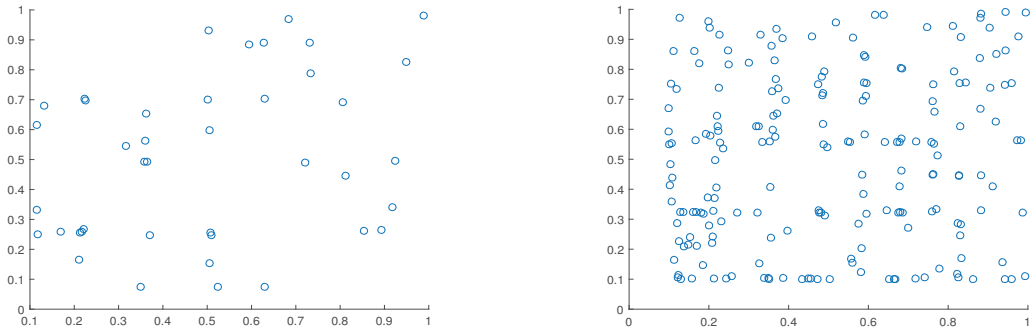

Figure 4: Copulas of joint laws of $T_{n+1} - T_n$ (on the vertical axis) and $T_n - T_{n-1}$ (on the horizontal axis) for the jump times T_i estimated on our JPM dataset. Left-hand: conditional law on $\bar{\lambda}_{T_{n-1}} \in (99, 102)$ for the fitted Hawkes process with the parameters $\lambda_0 = 53.27, \alpha = 4.72, \beta = 9.14$, giving the local maximizer of the likelihood function. The observed positive dependence is sensible. Right-hand: unconditional law. The observed positive dependence is weaker.

A.6.3 Link between the decay instant and the stochastic increasingness property of the interarrival times of an exponential Hawkes process

To check the relationship, we write $P(T_{n+1} - T_n > \tau | T_n - T_{n-1} = s, \bar{\lambda}_{T_{n-1}})$ in terms of the decay instant: abbreviating $t_\epsilon(T_{n-1}) = \frac{1}{\beta} \log \left(\frac{\bar{\lambda}_{T_{n-1}} - \lambda_0 + \alpha}{\lambda_0 \epsilon} \right)$ by t_ϵ, (A.31) can be written as

$$P(T_{n+1} - T_n > \tau | T_n - T_{n-1} = s, \bar{\lambda}_{T_{n-1}}) = e^{-\frac{1-e^{-\beta\tau}}{\beta}(\lambda_0 \epsilon \cdot e^{\beta t_\epsilon} e^{-\beta s} + \alpha) - \lambda_0 \tau},$$

and we find that the conditional probability that, given the length s of an interarrival $T_n - T_{n-1}$, the length of $T_{n+1} - T_n$ is greater than a constant τ is decreasing in t_ϵ. This can be explained by the fact that t_ϵ represents the end of the cluster, thus a larger decay instant indicates that each cluster involves a higher number of jumps, so it is more probable that both the jumps at T_n and T_{n+1} are part of a cluster together with the jump at T_{n-1}. Because the jumps of a cluster are close, the probability to have a large $T_{n+1} - T_n$ is smaller.

A.7 Jumps risk measurement with discrete observations, Section 5

A.7.1 Empirical probability of a double jump of the JPM price in 5 minutes

For our JPM database (with $\delta = 1/(252 \times 80)$) we measured the probability of having a *double jump* (i.e. two jumps within a time interval of length δ) as follows. We simulated a path of the exponential Hawkes with the parameters of the global maximizer of the likelihood function, on a time horizon $10^6 \times \delta$ (equal to about 6.6 times the time horizon $T = 151791 \times \delta$ of our dataset):

the number of non-overlapping 5-minute intervals containing two jumps is 23 out of 4793 intervals containing at least one jump (the total number of jumps being 4816). This amounts to having a probability that an interval contains a double jump equal to 23×10^{-6}, so that on our time horizon $[0, T]$ we expect to have 3.49 double jumps.
With the parameters of the local maximizer, on an analogous simulation we obtain 5490 intervals containing at least one jump, 24 of which contain exactly two jumps (for a total of 5514 jumps). Thus, the probability of a double jump is similar as above (24×10^{-6}).

A.7.2 Theoretical results

Remark 5. Given $\lambda_{(n-1)\delta}$ and that $\omega_{n-1} = 1$, from (11) it follows that, conditionally to the information $\mathcal{F}_{(n-1)\delta-}$, we can approximate the intensity $\bar{\lambda}_{n\delta}$ conditioned on the occurrence of a jump *within* $[(n-1)\delta, n\delta)$ by $\lambda_{n\delta}$ conditioned on the occurrence of a jump exactly *at* $(n-1)\delta$. The error would be

$$\begin{aligned}&\left|\bar{\lambda}_{n\delta|\mathcal{F}_{(n-1)\delta-},\{\text{last jump at } \tau\in[(n-1)\delta,n\delta)\}} - \lambda_{n\delta|\mathcal{F}_{(n-1)\delta-},\{\text{last jump at } (n-1)\delta\}}\right| \\ &\qquad = \alpha\left|e^{-\beta(n\delta-\tau)} - e^{-\beta\delta}\right| < \alpha(1-e^{-\beta\delta}).\end{aligned}$$

Proof of Theorem 2. In view of Eq. (10), we can write $\tilde{\lambda}_n = \Psi(\lambda_{n\delta})$. In order to use a recursive formula for writing $\lambda_{n\delta}$ in terms of $\lambda_{(n-1)\delta}$, we need to compute $\lambda_{n\delta} = \Psi^{-1}(\tilde{\lambda}_n)$. It turns out that

$$\lambda_{n\delta} = \lambda_0 + \frac{\beta}{e^{-\beta\delta}-1}(\delta\lambda_0 + \log(1-\tilde{\lambda}_n)). \tag{A.32}$$

If $\omega_{n-1} = 0$, we know by (A.23) that $\lambda_{n\delta} = \lambda_0 + (\lambda_{(n-1)\delta} - \lambda_0)e^{-\beta\delta}$, which, compared with Eq. (A.32), gives

$$(\lambda_{(n-1)\delta} - \lambda_0)e^{-\beta\delta} = \frac{\beta}{e^{-\beta\delta}-1}(\delta\lambda_0 + \log(1-\tilde{\lambda}_n)). \tag{A.33}$$

In addition, as in Eq. (A.32),

$$\lambda_{(n-1)\delta} = \lambda_0 + \frac{\beta}{e^{-\beta\delta}-1}(\delta\lambda_0 + \log(1-\tilde{\lambda}_{n-1})),$$

which, replaced in Eq. (A.33), leads to

$$e^{-\beta\delta}(\delta\lambda_0 + \log(1-\tilde{\lambda}_{n-1})) = \delta\lambda_0 + \log(1-\tilde{\lambda}_n),$$

hence the thesis.

If $\omega_{n-1} = 1$, the recursive formula only provides upper and lower bounds for $\lambda_{n\delta}$ in terms of $\lambda_{(n-1)\delta}$. Following the same substitution strategy as above the thesis is obtained. □

Remark 6. [About Theormem 2] If we only rely on discrete observation and want to implement the recursion, then we can write Eq. (12) in a more suitable form. Let us indicate by $\mathcal{I}_n$ any interval containing $\tilde{\lambda}_n$. Then, by Theorem 2 we have:

$$\mathcal{I}_n = \left[1-(1-\inf\mathcal{I}_{n-1})^{e^{-\beta\delta}}\exp\left((e^{-\beta\delta}-1)\left(\delta\lambda_0+\omega_{n-1}\frac{\alpha e^{-\beta\delta}}{\beta}\right)\right),\right. \tag{A.34}$$

$$\left. 1-(1-\sup\mathcal{I}_{n-1})^{e^{-\beta\delta}}\exp\left((e^{-\beta\delta}-1)\left(\delta\lambda_0+\omega_{n-1}\frac{\alpha}{\beta}\right)\right)\right]. \tag{A.35}$$

The case when $\tilde{\lambda}_{n-1}$ is given corresponds to $\mathcal{I}_{n-1}=\{\tilde{\lambda}_{n-1}\}$, so that $\inf\mathcal{I}_{n-1}=\sup\mathcal{I}_{n-1}=\tilde{\lambda}_{n-1}$.

Proof of Proposition 4. The proof is straightforward in view of the recursive formula and the application of the bounds. We split the self-exciting term of the intensity into two parts: the first is due to the history before $n\delta$, which gave to λ a contribution that at time $(n+k)\delta$ decayed, and the second is due to the last k jumps. The impact of each one of the k jumps is subject to a progressive decay.

$$\lambda_{(n+k)\delta}=\lambda_0+\alpha\sum_{t_i<n\delta}e^{-\beta((n+k)\delta-t_i)}+\alpha\sum_{i=0}^{k-1}\gamma_i e^{-\beta\delta i},$$

where the parameters $\gamma_i\in[e^{-\beta\delta},1)$ account for the uncertainty on the exact jump times within the intervals $[(n+i)\delta,(n+i+1)\delta)$.

If $\lambda_{n\delta}$ is given, we can write

$$\lambda_{(n+k)\delta}\in\left[\lambda_0+e^{-\beta\delta k}(\lambda_{n\delta}-\lambda_0)+\alpha\sum_{i=0}^{k-1}e^{-\beta\delta(i+1)},\lambda_0+e^{-\beta\delta k}(\lambda_{n\delta}-\lambda_0)+\alpha\sum_{i=0}^{k-1}e^{-\beta\delta i}\right],$$

and since $\sum_{i=0}^{k-1}e^{-\beta\delta i}=\frac{1-e^{-\beta\delta k}}{1-e^{-\beta\delta}}$, we obtain

$$\lambda_{(n+k)\delta}\in\left[\lambda_0+e^{-\beta\delta k}(\lambda_{n\delta}-\lambda_0)+e^{-\beta\delta}\frac{1-e^{-\beta\delta k}}{1-e^{-\beta\delta}},\lambda_0+e^{-\beta\delta k}(\lambda_{n\delta}-\lambda_0)+\frac{1-e^{-\beta\delta k}}{1-e^{-\beta\delta}}\right],$$

which proves the thesis. □

Remark 7. Eq. (15) in particular implies that

$$\left|\bar{\lambda}_{(n+k)\delta|\omega_i=1\ \forall\ i\in\{n,\dots,n+k-1\}}-\lambda_{(n+k)\delta|\bigcap_{i=0}^{k-1}\{\text{jump at }(n+i)\delta\}}\right|<\alpha(1-e^{-k\beta\delta}).$$

Remark 8. The approximation of $\bar{\lambda}_{(n+k)}$ obtained in Prop. 4 could alternatively be obtained by iterating (14) k times.

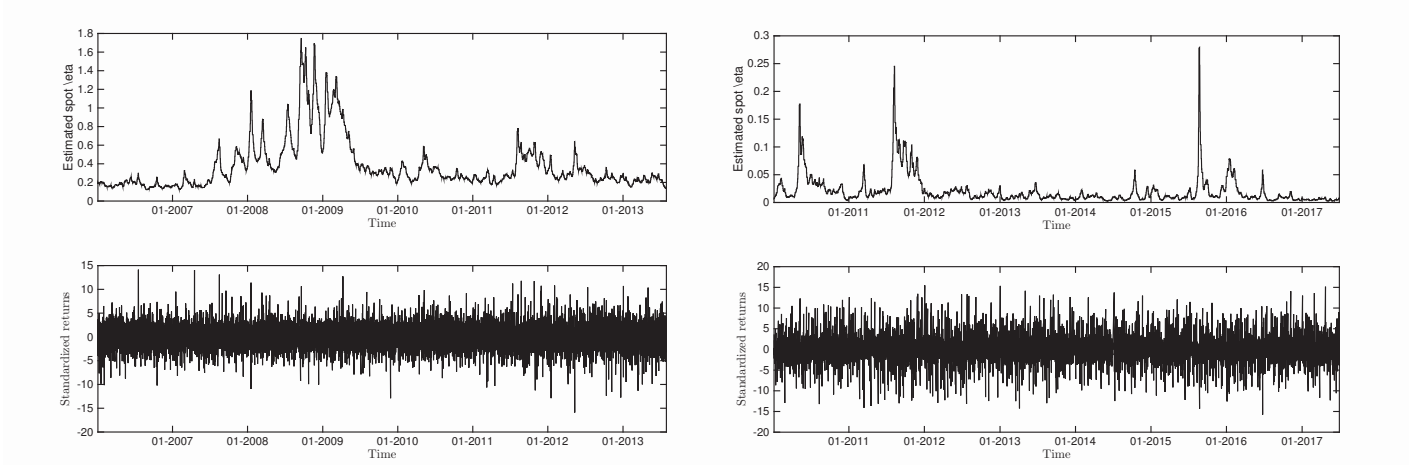

Figure 5: Estimated spot ηs of a model $dX = b_t dt + \eta_t dW_t$, and standardized log returns: JPM (left), S&P500 (right), $\delta = 5$ minutes

Remark 9. Starting from (15), in principle, we can obtain assessment of all the finite-dimensional conditional distributions of the process, which are the probabilities, conditional on $\lambda_{n\delta}$ and ω_n, ω_{n+1}, $..., \omega_{n+k-1}$, that a sequence $(\omega_{n+k}, \omega_{n+k+1}, ..., \omega_{n+k+m-1})$ assumes a given combination of m values within $\{0, 1\}$, as m varies and the combination varies. However, in the presence of zeros in the combination, at moment we can only provide too large bounds and we are unable to deliver useful practical indications.

A.8 Empirical application, Section 6

A.8.1 Importance of studying jumps

As explained in Section 6, Figure 5 evidences that for our datasets some log returns are incompatible with a Brownian semimartingale because they are too large. Our model classifies such large variations as jumps. A stochastic volatility model with continuous paths does not fit well our data (Figure 5), while a jump-diffusion model does it much better (Figure 6).

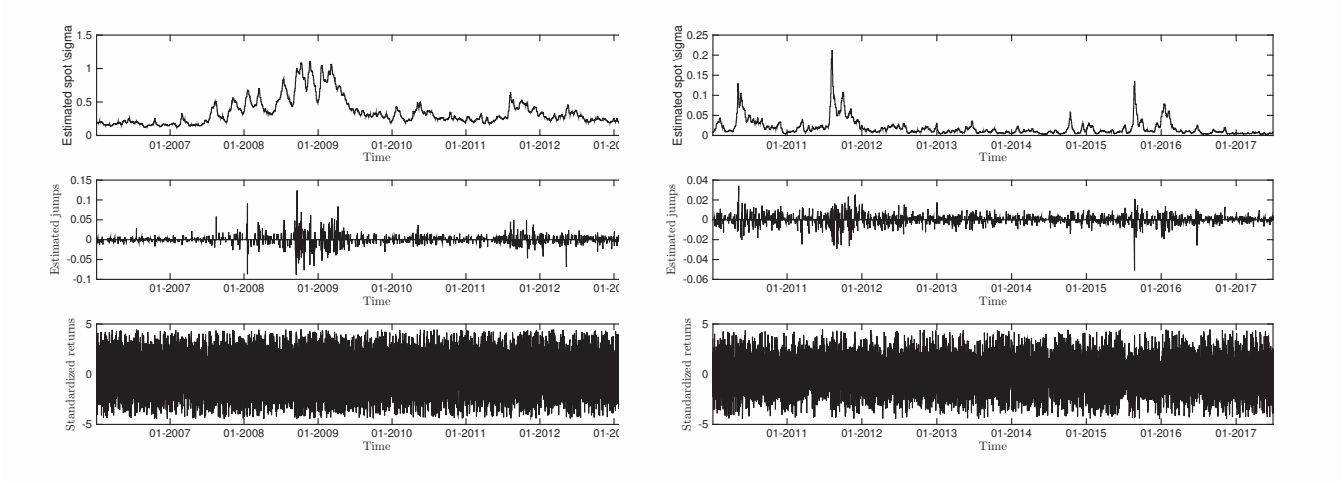

Figure 6: Estimated spot σs of a model $dX = a_t dt + \sigma_t dW_t + dJ_t$, estimated jumps and standardized log returns: JPM (left), S&P500 (right)

A.8.2 Statistical analysis of the JPM prices large variations

The first two lines of Table 4 display the number, both in absolute terms and in percentage, of the *overlapping weeks* [9] containing exactly k jumps, for $k = 0, 1, ..., 9$, while the third and fourth lines show the number of the *non-overlapping* 378 weeks of the dataset with k jumps. We call the relative frequencies that are reported in the tables *empirical probabilities*. They are expressed in percentage (rows labeled a and b).

Table 4: Number of the overlapping (or non-overlapping) *weeks* containing *exactly* k jumps: in absolute value (nb) and in percentage (a) wrt the 151392 weeks (378, respectively) for the JPM dataset. For comparison, the last line reports the probability$\times$100 (b) that the Poisson process best fitting the JPM jumps (annual parameter 105.9823) makes exactly k jumps on the time interval of length $400 \times \delta$, corresponding to 5 consecutive days. e-p means $\times 10^{-p}$.

k	0	1	2	3	4	5	6	7	8	9	10	11	12	13
JPM jumps, overlapping weeks														
nb	17886	41794	40385	26699	13364	6180	2021	1042	712	100	572	317	242	78
a	11.81	27.61	26.68	17.64	8.83	4.08	1.33	0.69	0.47	0.07	0.38	0.21	0.16	0.05
"P"$(N_S = k)$: JPM jumps, non-overlapping weeks														
nb	42	111	97	67	38	12	4	2	2	0	1	2	1	0
a	11.08	29.29	25.59	17.68	10.03	3.17	1.06	0.53	0.53	0	0.26	0.53	0.26	0
Poisson jumps, non-overlapping weeks														
b	12.21	25.68	27.00	18.92	9.95	4.18	1.47	0.44	0.12	0.03	0.006	0.001	2e-4	3e-5

[9] Groups of 400 observations, which belong to five consecutive working days, starting from the first observation of the dataset of the returns and rolling over the whole record of the observations: each group differs from the previous one by only the first and the last observations, the first one in a group is the second one of the previous group and the last one is just after the last one of the previous group.

The first line labeled a in Table 4 is used to define and compute in Table 6 the probability of a jump at a given t, conditionally on the occurrence of k consecutive jumps in the week preceding t.

The comparison between the second line labeled a (non-overlapping weeks) and the last line of the Table shows that, for corresponding events, the probabilities attributed by the Poisson process best fitting our data to the weeks with numerous jumps ($k \geq 8$) are substantially different from the empirical probabilities: the empirical probabilities of $k \geq 8$ jumps in a week (for both the overlapping and the non-overlapping cases) for our data are higher or extremely higher with respect to the analogous probabilities for the Poisson process. For instance, the probabilities of $k = 11$ (or $k = 12$) jumps in a week are 500 (resp. 1000) times higher in our empirical distribution than in the Poisson one. In our empirical record of jumps, a significant number of large clusters appear, which is in contrast to the case of the Poisson distribution. Thus, the JPM jumps do not behave consistently with a Poisson process.

Table 5: Number of the overlapping (or non-overlapping) days with k jumps: in absolute value (nb) and in percentage (pct $\times 100$) wrt the 151712 (1897, respectively) days of the JPM dataset. For a comparison, the last line reports the probability$\times 100$ (Prob$\times 100$) that the Poisson process best fitting the JPM jumps (annual parameter 105.9823) makes k jumps on the time interval of length $80 \times \delta$, corresponding to the length of most of the days in the dataset (in fact there are some days where the market was officially open only in the morning).

k	0	1	2	3	4	5	6	7	8
JPM jumps, overlapping days									
nb	99967	41099	8817	1338	233	70	113	27	48
pct $\times 100$	65.89	27.09	5.81	0.88	0.154	0.046	0.074	0.018	0.032
JPM jumps, non-overlapping days									
nb	1250	514	111	16	2	1	2	1	0
pct$\times 100$	65.89	27.10	5.85	0.84	0.105	0.053	0.105	0.053	0
Poisson jumps, non-overlapping days									
Prob$\times 100$	65.67	27.62	5.81	0.81	0.086	0.007	5×10^{-4}	3×10^{-5}	2×10^{-6}

Table 5 is analogous to Table 4, but the considered time period is 1 day (rather than 1 week), where two consecutive overlapping days are two groups of 80 observations and differ, as before, by only the first and the last observations. A comparison between the second line labeled pct$\times 100$ and the line labeled Prob$\times 100$ shows that the values for $k = 0, 1, 2, 3$ are similar, while for $k = 5, 6, 7$ the values are more and more discrepant. The empirical dataset displays clusters of numerous

jumps, concentrated in 1 day, much more than the Poisson process. This is consistent with what is observed in Table 4 and indicates that our data are more in line with the behavior of a Hawkes process.

Table 6 shows the frequency, called here *empirical conditional probability*, of the occurrence of exactly k jumps in a week conditional on the occurrence of exactly j jumps in the previous week (non-overlapping weeks), indicated as "P"$(N_S = k|N_{S-1} = j)$, with N_S = number of jumps in week S. Note that in Tables 4, 5 and 6 not necessarily the k (or j) considered jumps are consecutive.

Table 6: Empirical probability "P"$(N_S = k|N_{S-1} = j)$, in percentage form, of $N_S = k$ given that $N_{S-1} = j$, non-overlapping weeks. Note that the line and column relative to $j = 9$ or $k = 9$ are omitted, that is because there are no weeks S such that $N_S = 9$.

j\ k	0	1	2	3	4	5	6	7	8	10	11	12
0	16.67	28.57	19.05	16.67	14.29	0	2.38	0	0	0	2.38	0
1	9.01	30.63	30.64	18.92	5.41	2.70	0.90	0.90	0	0.90	0	0
2	9.38	26.04	27.08	18.75	10.42	5.21	1.04	1.04	1.04	0	0	0
3	11.94	34.33	22.39	16.42	10.45	2.99	1.49	0	0	0	0	0
4	13.16	13.16	26.32	23.68	15.79	5.26	0	0	2.63	0	0	0
5	8.33	50.00	16.67	0	16.67	0	0	0	0	0	8.33	0
6	0	50.00	25.00	0	25.00	0	0	0	0	0	0	0
7	0	100.00	0	0	0	0	0	0	0	0	0	0
8	100.00	0	0	0	0	0	0	0	0	0	0	0
10	0	100.00	0	0	0	0	0	0	0	0	0	0
11	0	0	0	50.00	0	0	0	0	0	0	0	50.00
12	0	100.00	0	0	0	0	0	0	0	0	0	0

A comparison between Table 6 and the empirical probability "P"$(N_S = k)$ of having k jumps in a week, displayed at the second panel of Table 4, proves that the increments of the process counting the jumps of JPM are not independent. In fact, if N_S was independent on N_{S-1}, we would have that "P"$(N_S = k|N_{S-1} = j)$ = "P"$(N_S = k)$ for any j and k. On the contrary we have that, for instance, with $k = 1$, the relative frequencies "P"$(N_S = 1|N_{S-1} = j)$ are quite different as j varies. Compare for instance the cases of $k = 1$ with $j = 4, 5, 7$ and 8: they are very different from each other and different from "P"$(N_S = 1) = 0.2929$. This is also evident for all the other values of k, as e.g. $k = 5$ and $k = 8$.

A.8.3 MLE of the parameters

To perform MLE[10] we constructed and used a Matlab code, which is available upon request, employing the package fmincon for the function $-\log \mathcal{L}$ of the parameters describing the model jump intensity. The expression for $\log \mathcal{L}$ is reported below[11]. For each model, we preliminarily computed the log likelihood function on a huge grid of values of the parameters to have an idea of the regions where $\log \mathcal{L}$ assumes higher values. The fmincon function requires that we indicate a *starting point*, a vector of 3 components for model 1, 5 for model 2, and 4 for model 3. From this, recursively, at each step the routine looks for a subsequent vector where $-\log \mathcal{L}$ reaches a lower value, until such a low value stabilizes. We proceeded by running the routine 100 times, each time with a different starting point, randomly sampled from a uniform distribution[12]. For each starting point the routine produced a vector of estimates and in Table 1 we only reported the estimates that maximize $\log \mathcal{L}$. For each model, we also ran the routine again while using as starting point the set of parameters estimated in the previous maximization and the procedure again converged to it.
The standard error of each estimate is computed as follows. For each model we simulated 1000 paths using the parameters of the model, while for each path we performed MLE, starting from the set of the true values, finally for each parameter we computed the standard deviation of the 1000 obtained estimates.

Expression of the likelihood of the Hawkes process

The density of the law of a Hawkes process N on $[0, T]$ with respect to the Lebesgue measure

[10] Consistency, asymptotic normality and efficiency of the MLE, for a univariate exponential Hawkes process, are provided in [20] and [6]. For the models having kernel either compounded by two exponential terms or by one power function, we assessed the quality of the MLE estimators through a simulation study. We found that the number of observations we dispose of allows us to identify the parameters.

[11] Firstly we checked the reliability of our estimation procedure by applying it to simulated data generated by a Hawkes process having chosen parameters. To simulate the paths of a Hawkes process, we followed [25], which is based on the Ogata modified thinning algorithm [21]. We repeated with many different parameters choices and the estimation results obtained with horizon time length $T = 7.5$ were reasonable. We then repeated the estimation with alternative Matlab routines: we implemented the recursive procedure described in [5] and also a procedure written, and kindly provided, by G. Bormetti and F. Lillo of the University of Bologna, still based on the mentioned recursive procedure, which is described in [23]. The results were found to be similar.

[12] For 50 runs λ_0 and α are sampled from a uniform law on $(0, 1)$, while β from a uniform law on $(\alpha, 2)$, while for the remaining 50 runs λ_0 is sampled from a uniform law on $(30, 100)$, α on $(2, 20)$ and β on $(\alpha, 30)$.

m on $\mathbb{R}_+$ is given by ([8], Proposition 7.3.III)

$$\log \mathcal{L} = -\int_0^T \lambda_t dt + \int_0^T \log(\lambda_t) dN_t.$$

Denote the antiderivative of $\Phi(x)$ by $\tilde{\Phi}(x)$, and note that

$$-\int_0^T \lambda_t dt = -\lambda_0 T - \int_0^T \int_0^t \Phi(t-s) dN_s dt,$$

by Fubini theorem and conditionally on having observed the times of the jumps occurred within $[0,T]$, the last display equals

$$-\lambda_0 T - \int_0^T \int_s^T \Phi(t-s) dt \; dN_s$$

$$= -\lambda_0 T - \int_0^T [\tilde{\Phi}(T-s) - \tilde{\Phi}(0)] dN_s = -\lambda_0 T - \sum_{\ell:\tau_\ell<T} (\tilde{\Phi}(T-\tau_\ell) - \tilde{\Phi}(0)).$$

Furthermore,

$$\int_0^T \log(\lambda_t) dN_t$$

$$= \sum_{\tau_\ell<T} \log(\lambda_{\tau_\ell}) = \sum_{\tau_\ell<T} \log\Big(\lambda_0 + \int_0^{\tau_\ell} \Phi(\tau_\ell - s) dN_s\Big) = \sum_{\tau_\ell<T} \log\Big(\lambda_0 + \sum_{\tau_m<\tau_\ell} \Phi(\tau_\ell - \tau_m)\Big).$$

then

$$\log \mathcal{L} = -\lambda_0 T - \sum_{\ell:\tau_\ell<T} \Big[\tilde{\Phi}(T-\tau_\ell) - \tilde{\Phi}(0) + \log\Big(\lambda_0 + \sum_{\tau_m<\tau_\ell} \Phi(\tau_\ell - \tau_m)\Big)\Big].$$

In particular, for $\Phi(x) = \sum_{p=1}^{\mathcal{P}} \alpha_p e^{-\beta_p x}$ we have $\tilde{\Phi}(x) = -\sum_{p=1}^{\mathcal{P}} \frac{\alpha_p}{\beta_p} e^{-\beta_p x}$, so that

$$\int_0^T \Phi(x) dx = \tilde{\Phi}(T) - \tilde{\Phi}(0) = \sum_{p=1}^{\mathcal{P}} \frac{\alpha_p}{\beta_p}(1 - e^{-\beta_p T}),$$

while for $\Phi(x) = \frac{\alpha\beta}{(1+\beta x)^{1+\gamma}}$ we have

$$\int_0^T \Phi(x) dx = \frac{\alpha}{\gamma}\Big(1 - \frac{1}{(1+\beta T)^\gamma}\Big).$$

□

A.8.4 Reliability check on simulations.

Once obtained $n = 150'000$ simulated increments $\Delta_i X$ of a process X following (20), with $t_i = i\delta, \delta = 1/(252 \times 80), i = 1..n$, $T = n\delta$, we estimated spot $\tilde{\sigma}^2_{t_{i-1}}$ through (18), with bandwidth $h = 100\delta$. We ran a second recursion, by determining $\hat{\sigma}^2_{t_{i-1}}$ as in (19) and then identifying as returns containing jumps the ones such that $(\Delta_i X)^2 > 2\hat{\sigma}^2_{t_{i-1}} \delta \log \frac{1}{\delta}$. Then, using the estimated

jump times $\hat{\tau_k}$, we implemented MLE for the exponential Hawkes parameters. With $\epsilon = 0.01$, for a given choice of U and S, we compared the quantities

$$P \doteq P\{T_{i+k} < S + t_\epsilon(S)\}, LB \doteq P\{T_{i+k} < S + \underline{t}_\epsilon\}, UB \doteq P\{T_{i+k} < S + \bar{t}_\epsilon\}$$

computed with the true parameters as in Propositions 2 and 1 with the estimates $\hat{P}, \hat{LB}, \hat{UB}$ obtained by plugging $\hat{\lambda}_0, \hat{\alpha}, \hat{\beta}$ into the quantities required to implement the respective formulas, namely $\hat{k} = \hat{N}_S - \hat{N}_U$ and $\hat{\lambda}_t = \hat{\lambda}_0 + \hat{\alpha} \sum_{\tau_\ell < t} e^{-\hat{\beta}(t-\tau_\ell)}$, for $t = U, S$.

We only considered the simulated paths of X for which S does not coincide with neither a jump time nor with an estimated one. Under the hypotheses $\lambda_S > \lambda_0(1+\epsilon)$ we are guaranteed that $P \in (0,1)$, $UB \in [0,1]$ and $LB \leq 1$, however to guarantee $LB \geq 0$ an even smaller ϵ has to be taken, for instance such that $\lambda_S > \lambda_0(1 + \epsilon \cdot e^\epsilon)$. We defined the following percentage estimation errors ("E" stands for "error"):

$$E^{Mean(\sigma)} = \frac{|mean(\hat{\sigma}_{t_i}) - mean(\sigma_{t_i})|}{mean(\sigma_{t_i})}, \quad E^{N_T} = \frac{|\hat{N}_T - NT|}{N_T},$$

$$E^{\lambda_0} = \frac{|\hat{\lambda}_0 - \lambda_0)|}{\lambda_0}, \quad E^{\alpha} = \frac{|\hat{\alpha} - \alpha)|}{\alpha}, \quad E^{\beta} = \frac{|\hat{\beta} - \beta)|}{\beta},$$

$$E^{P} = \frac{|\hat{P} - P|}{P}, \quad E^{LB} = \frac{\hat{LB} - LB}{LB}, \quad E^{UB} = \frac{\hat{UB} - UB}{UB}.$$

$$d(P, LB) = \frac{P - LB}{P}, \quad d(P, \hat{LB}) = \frac{P - \hat{LB}}{P}, \quad d(P, UB) = \frac{UP - P}{P}, \quad d(P, \hat{UB}) = \frac{\hat{UB} - P}{P},$$

$$E^{\lambda_S} = \frac{\hat{\lambda}_S - \lambda_S}{\lambda_S}, \quad E^{\lambda_U} = \frac{\hat{\lambda}_U - \lambda_U}{\lambda_U}, \quad rangeBs = \frac{UB - LB}{P}, \quad \hat{rangeBs} = \frac{\hat{UB} - \hat{LB}}{P}$$

We repeated for 500 simulated paths of X such that $S \neq \tau_j, \hat{\tau}_j$ for any j and $\lambda_S > \lambda_0(1+\epsilon e^\epsilon), \hat{\lambda}_S > \hat{\lambda}_0(1+\epsilon e^\epsilon)$. The average estimation errors are reported in table 7.

Table 7: Average percentage estimation errors on 500 simulated paths of X as in (20). Standard errors are within brackets.

Spot σ and point process N		
$E^{Mean(\sigma)}$= 0.0025697 (0.001737)	E^{N_T}=0.24814 (0.021869)	
E^{λ_0}=0.19175 (0.1234),	E^{α}= 0.40574 (0.24497)	E^{β}=0.54079 (0.8463)
U= 0.18, S= 0.215		
LB= 1 (7.1652e-10)	$\hat{LB}$=0.99505 (0.042901)	E^{LB}= 0.0049507 (0.042901)
P=1 (5.7832e-10)	$\hat{P}$=0.99854 (0.011961)	E^{P} =0.0014647 (0.011961)
UB= 1 (5.2035e-10)	$\hat{UB}$= 0.99936 (0.0056831)	E^{UB}= 0.00064008 (0.0056831)
$d(P,\hat{LB})$= 0.0049507 (0.042901)	$d(P,\hat{UB})$= 0.00064008 (0.0056831)	$d(P,LB)$=3.196e-11 (2.026e-10)
$d(P,UB)$= 1.5458e-11 (1.0879e-10)	$rangeBs$= 4.7418e-11 (3.0757e-10)	$range\hat{B}s$= 0.0043106 (0.037304)
E^{λ_S}= -0.26068 (0.098074)	E^{λ_U}= -0.25768 (0.10174)	
U2= 0.02, S2= 0.07		
LB= 1 (1.6599e-07)	$\hat{LB}$= 0.9811 (0.12206)	E^{LB}= 0.0189 (0.12206)
P=1 (1.0129e-07)	$\hat{P}$=0.99445 (0.036556)	E^{P} = 0.0055484 (0.036556)
UB= 1 (8.3005e-08)	$\hat{UB}$= 0.99756 (0.02364)	E^{UB}=0.0024435 (0.02364)
$d(P,\hat{LB})$=0.0189 (0.12206)	$d(P,\hat{UB})$= 0.0024435 (0.02364)	$d(P,LB)$= 4.2277e-08 (1.069e-07)
$d(P,UB)$= 1.5058e-08 (5.1005e-08)	$rangeBs$= 5.7334e-08 (1.3523e-07)	$range\hat{B}s$= 0.016457 (0.11547)
E^{λ_S}= -0.23757 (0.12747)	E^{λ_U}= -0.18247 (0.17135)	