Hawkes Process: Fast Calibration, Application to Trade Clustering, and Diffusive Limit

JOSÉ DA FONSECA* and RIADH ZAATOUR

This study provides explicit formulas for the moments and the autocorrelation function of the number of jumps over a given interval for a self-excited Hawkes process. These computations are possible thanks to the affine property of this process. Using these quantities an implementation of the method of moments for parameter estimation that leads to an fast optimization algorithm is developed. The estimation strategy is applied to trade arrival times for major stocks that show a clustering behavior, a feature the Hawkes process can effectively handle. As the calibration is fast, the estimation is rolled to determine the stability of the estimated parameters. Lastly, the analytical results enable the computation of the diffusive limit in a simple model for the price evolution based on the Hawkes process. It determines the connection between the parameters driving the high-frequency activity to the daily volatility. © 2013 Wiley Periodicals, Inc. Jrl Fut Mark 34:548–579, 2014

1. INTRODUCTION

Trading activity leads to time series of irregularly spaced points that show a clustering behavior. This stylized property suggests the use of the Hawkes process, a point process mathematically defined by Hawkes (1971), which is an extension of the classical Poisson process that possesses this clustering property. It explains the large number of works on trading activity and more generally high-frequency econometrics based on this process as a modeling framework. To name only a few let us quote Hewlett (2006), Bowsher (2007), Large (2007), Bacry, Delattre, Hoffmann, and Muzy (2013a), or Muni Toke and Pomponio (2011).[1]

There are other stochastic processes possessing this clustering property; they are often more sophisticated than the Hawkes process in the sense that their dynamic involves several

José Da Fonseca is a Senior Lecturer at the Department of Finance, Auckland University of Technology, New Zealand. Riadh Zaatour is a PhD candidate at the Chair of Quantitative Finance, Ecole Centrale Paris, Châtenay-Malabry, France. We are thankful to the participants of the Recent Advances in Algo and HF Trading conference (University College London, April 2013), the participants and our discussant, Yuen Jung Park, of the 9th Annual Conference of the Asia-Pacific Association of Derivatives (APAD 2013) in Busan. All remaining errors are ours.

JEL Classification: C13, C32, C58

*Correspondence author, Department of Finance, Business School, Auckland University of Technology, Private Bag 92006, 1142 Auckland, New Zealand. Tel: +64-9-9219999 ext. 5063, Fax: +64-9-9219940, e-mail: jose.dafonseca@aut.ac.nz

Received May 2013; Accepted September 2013

[1]For non-financial applications see Vere-Jones (1970), Veen and Schoenberg (2008), Lewis and Mohler (2011), and Mohler, Short, Brantingham, Schoenberg, and Tita (2011).

The Journal of Futures Markets, Vol. 34, No. 6, 548–579 (2014)

Published online 23 October 2013 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/fut.21644

lags and strong nonlinearities. They are actively studied in the econometrics literature; see Hautsch (2012) for a general overview. As for these sophisticated stochastic processes the Hawkes process has a likelihood function that is known in closed form. As a consequence, most of the existing literature focuses on the estimation of the dynamics. However, despite or because of its simplicity the Hawkes process has several advantages from an analytical point of view, which we will develop in this study, that allow for very interesting applications.[2]

First, we show how to compute in closed form the moments of any order of the number of jumps over a given time interval. This analytical tractability even extends to the autocorrelation function of the number of jumps. As such, we can develop an estimation strategy based on these quantities, which, compared with the likelihood estimation strategy, is extremely fast. This aspect is crucial when it comes to applications such as high-frequency trading activity which requires a fast estimation procedure. The maximization of the likelihood function, which can take several minutes, cannot be used in real applications. What is more, with the estimation being immediate, we can roll the estimation procedure and study the parameter stability, which is an essential aspect in practice.

Second, thanks to its analytical tractability we can explicitly compute the impulse function of the Hawkes process, and therefore the trading activity, modelized by this process, can easily be analyzed.

Third, within a simple toy model for a stock based on the Hawkes process we compute the diffusive limit for the asset and therefore make the link between the microscopic activity (i.e., the trading activity at high frequency) to the macroscopic activity (i.e., the daily volatility as used in the Black–Scholes model) explicit. To perform such analysis the analytical tractability of the Hawkes process turns out to be essential and underlines the advantages related to the simplicity of this process. Our work falls in a new trend of the literature developed by Cont, Stoikov, and Talreja (2010), Cont and De Larrard (2011, 2012), Bacry et al. (2013a), Abergel and Jedidi (2013), Bacry, Delattre, Hoffmann, and Muzy (2013b), and Kirilenko, Sowers, and Meng (2013) aiming at connecting these two scales (the high-frequency quantities and the daily quantities).

The structure of the study is as follows. In the first section, we describe the analytical framework that comprises the basic properties of the Hawkes process as well as the Dynkin formula that will be our main mathematical tool. Using these results, the computation of the moments and the autocorrelation function of the number of jumps over a given time interval is provided. It also contains the usual optimization algorithms used in the literature and the method of moments based on the analytical results. In a second part, we present the data, various estimation results, and an impulse response analysis allowed by the model. This second part is completed with a toy model for a stock for which we derive the limit properties. Finally, we conclude and provide some technical results, tables, and figures that are gathered in the Appendix.

2. THE ANALYTICAL FRAMEWORK

2.1. Dynamics and Affine Structure of the Moment-Generating Function

The Hawkes process was defined in Hawkes (1971) and is a self-excited point process whose intensity depends on the path followed by the point process. More precisely, the point process

[2]We stress the fact that in this work we focus only on the univariate Hawkes process, which is, therefore, a *self-excited* process. The multivariate Hawkes process, which can be self-excited and mutually excited, is not covered in this work.

is determined by the intensity process $(\lambda_t)_{t\geq 0}$ through the relations:

$$\mathbb{P}[N_{t+h} - N_t = 1|\mathcal{F}_t] = \lambda_t h + o(h), \tag{1}$$

$$\mathbb{P}[N_{t+h} - N_t > 1|\mathcal{F}_t] = o(h), \tag{2}$$

$$\mathbb{P}[N_{t+h} - N_t = 0|\mathcal{F}_t] = 1 - \lambda_t h + o(h), \tag{3}$$

where $(\mathcal{F}_t)_{t\geq 0}$ is a filtration on the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ containing the filtration generated by $(N_t)_{t\geq 0}$. The intensity follows the dynamic:

$$d\lambda_t = \beta(\lambda_\infty - \lambda_t)\,\mathrm{d}t + \alpha\,\mathrm{d}N_t. \tag{4}$$

A jump of N_t at a given time will increase the intensity, which increases the probability of another jump thanks to Equation (1) and justifies the use of the term "self-exciting" to qualify this process. The jumps tend to cluster but the process does not blow up because the drift becomes negative whenever the intensity is above $\lambda_\infty > 0$ (β is by hypothesis positive) and prevents any explosion. Furthermore, applying Ito's lemma to $\mathrm{e}^{\beta t}\lambda_t$ yields:

$$\lambda_t = \mathrm{e}^{-\beta t}(\lambda_0 - \lambda_\infty) + \lambda_\infty + \int_0^t \alpha\,\mathrm{e}^{-\beta(t-s)}\,\mathrm{d}N_s. \tag{5}$$

From (5) we also observe that the impact on the intensity of a jump dies out exponentially as time passes. For the existence and uniqueness results we refer to chapter 14 of Daley and Jones (2008) and references therein; of particular interest is Brémaud and Massoulié (1994).

As t gets larger the impact of λ_0, the initial value for the intensity, vanishes leaving us with:

$$\lambda_t \sim \lambda_\infty + \int_0^t \alpha\,\mathrm{e}^{-\beta(t-s)}\,\mathrm{d}N_s.$$

Our presentation differs slightly from the usual one found in the literature where the Hawkes intensity is written as:

$$\lambda_t = \lambda_\infty + \int_{-\infty}^t \alpha\,\mathrm{e}^{-\beta(t-s)}\,\mathrm{d}N_s. \tag{6}$$

Equation (6) leads to a stochastic differential equation similar to (4); the process starts infinitely in the past and is at its stationary regime. In our case we have a dependency with respect to the initial position λ_0 in Equation (5) but, as mentioned above, for t large enough its impact will vanish.

Our presentation for the Hawkes process follows closely Errais, Giesecke, and Goldberg (2010) and is motivated by the fact that we want to perform stochastic differential calculus.

The process $X_t = (\lambda_t, N_t)$ is a Markov process in the state space $D = \mathbb{R}_+ \times \mathbb{N}$. This is a key property that will give us very powerful tools to investigate the distributional properties of the process. Among these tools is the infinitesimal generator. Consider a sufficiently regular function $f : D \to \mathbb{R}$, the infinitesimal generator of the process, denoted $\mathcal{L}$, is the operator acting on f such that:

$$\mathcal{L}f(x) = \lim_{h\to 0} \frac{\mathbb{E}_t^x[f(X_{t+h})] - f(x)}{h}$$

with $\mathbb{E}_t^x[\cdot] = \mathbb{E}^x[\cdot|\mathcal{F}_t]$ and $X_t = x$ ($\mathbb{E}_0[\cdot] = \mathbb{E}[\cdot]$). In the case of a Hawkes process, this writes:

$$\mathcal{L}f(x) = \beta(\lambda_\infty - \lambda_t)\frac{\partial f}{\partial \lambda}(x) + \lambda_t[f(\lambda_t + \alpha, N_t + 1) - f(x)]. \tag{7}$$

For every function f in the domain of the infinitesimal generator, the process:

$$M_t = f(X_t) - f(X_0) - \int_0^t \mathcal{L}f(X_u)\,\mathrm{d}u$$

is a martingale relative to its natural filtration (see, for example, Proposition 1.6 of chapter VII in Revuz and Yor (1999)); thus, for $s > t$ we have:

$$\mathbb{E}_t\left[f(X_s) - \int_0^s \mathcal{L}f(X_u)\,\mathrm{d}u\right] = f(X_t) - \int_0^t \mathcal{L}(X_u)\,\mathrm{d}u$$

by the martingale property and we finally obtain the important Dynkin formula:

$$\mathbb{E}_t[f(X_s)] = f(X_t) + \mathbb{E}_t\left[\int_t^s \mathcal{L}f(X_u)\,\mathrm{d}u\right]. \tag{8}$$

This formula allows the computation of conditional expectations of functions of the Markov process (λ_t, N_t), which turns out to be very useful when the expectation on the right-hand side can be easily computed. In the following section, we will rely heavily on this formula to compute some distributional properties of Hawkes process.

As underlined in Errais et al. (2010) the process $X_t = (\lambda_t, N_t)$ is a Markov process that is *affine*, which implies that a closed form solution for the moment-generating function is available. Let $u = (u_1, u_2)^\top \in \mathbb{R}^2$, the conditional moment-generating function of $X_T = (\lambda_T, N_T)$ is defined as $f(t, X_t) = \mathbb{E}_t^x[\mathrm{e}^{u^\top X_T}] = \mathbb{E}_t^x[\mathrm{e}^{u_1\lambda_T + u_2 N_T}]$. Clearly, $f(t, X_t)$ must be a martingale and the function f satisfies:

$$\frac{\partial f}{\partial t}(t, X_t) + \mathcal{L}f(t, X_t) = 0 \tag{9}$$

with boundary condition $f(T, X_T) = \mathrm{e}^{u^\top X_T}$. As $X_t = (\lambda_t, N_t)$ is a Markov affine point process, we guess the solution of (9) is an exponential affine form of the state variable, that is to say:

$$f(t, X_t) = \mathrm{e}^{a(t) + b(t)\lambda_t + c(t)N_t}. \tag{10}$$

Setting this guess into Equation (9) we obtain the system of ordinary differential equations:

$$\frac{\partial a}{\partial t} = -\beta\lambda_\infty b(t), \tag{11}$$

$$\frac{\partial b}{\partial t} = \beta b(t) + 1 - \mathrm{e}^{\alpha b(t) + c(t)}, \tag{12}$$

$$\frac{\partial c}{\partial t} = 0 \tag{13}$$

with terminal conditions $a(T) = 0$, $b(T) = u_1$, and $c(T) = u_2$.

The above system of ODE (ordinary differential equation) fully characterizes the moment-generating function and the Laplace transform of the process, which completely

determines its distribution. However, an explicit solution for Equation (12) is usually not available. From the moment-generating function (10) we can retrieve the moments of the process *after* differentiating it with respect to u_1 or u_2, depending on which variable is considered, and evaluating the resulting function for $(u_1, u_2) = 0$. This computation leads to the differentiation of the system of ODE with respect to u_1 or u_2. Therefore, we can compute the different moments of the number of jumps by successively differentiating the system of ordinary differential equations. In our particular case (i.e., for a one-dimensional Hawkes process) the computations are feasible, but for a multidimensional Hawkes process, they soon become tedious. Also, whenever we wish to compute moments for the intensity process λ_t the resolution of (12) is not possible. We provide in the Appendix an example illustrating the difficulties raised by the computations of the moments.

Lastly, the computation of the autocovariance function of the number of jumps increments, that is to say:

$$\mathbb{E}_t^x[(N_{t_4} - N_{t_3})(N_{t_2} - N_{t_1})] \tag{14}$$

with $t < t_1 < t_2 < t_3 < t_4$ can be obtained from (10) by performing successive conditioning. In that case it will introduce the intensity process, which appears on the right-hand side of (10), and implies that the joint moment-generating function (i.e., (λ_t, N_t)) has to be evaluated. The resulting system of ODE and its differentiation become more complicated. Because the quantity (14) is essential, it carries the clustering property of the Hawkes process; we need to develop a simpler approach to perform the computation.

In Errais et al. (2010), the authors rely on this approach to compute the expected number of jumps and higher moments, but they rapidly become tedious beyond the first moment. To overcome these computational difficulties that obviously increase with the dimension of the process or when the autocovariance property of the process is needed, we develop another approach that we now present.

2.2. Computing the Moments and the Autocovariance Function

Our aim in this section is to compute the moments of the process $X_t = (\lambda_t, N_t)$ and also the autocovariance of the number of jumps over a period τ. To achieve this we rely on the infinitesimal generator of the process given by (7) and Dynkin's formula (8). In order to obtain the expected number of jumps and the expected intensity we use the following lemma:

Lemma 1. *Given a Hawkes process $X_t = (\lambda_t, N_t)$ with dynamic given by* (4) *the expected number of jumps $\mathbb{E}[N_t]$ and the expected intensity $\mathbb{E}[\lambda_t]$ satisfy the set of ODE:*

$$d\mathbb{E}[N_t] = \mathbb{E}[\lambda_t]\,\mathrm{d}t, \tag{15}$$

$$d\mathbb{E}[\lambda_t] = (\beta\lambda_\infty + (\alpha - \beta)\mathbb{E}[\lambda_t])\,\mathrm{d}t. \tag{16}$$

Note that (16) can be explicitly computed and once obtained another integration leads to the expression for $\mathbb{E}[N_t]$. If these quantities are known, a similar procedure will give higher-order moments as we have:

Lemma 2. *Given a Hawkes process $X_t = (\lambda_t, N_t)$ with dynamic given by* (4) $\mathbb{E}[\lambda_t^2]$, $\mathbb{E}[\lambda_t N_t]$ *and* $\mathbb{E}[N_t^2]$ *satisfy the set of ODE:*

$$d\mathbb{E}[N_t^2] = 2\mathbb{E}[\lambda_t N_t]\,\mathrm{d}t + \mathbb{E}[\lambda_t]\,\mathrm{d}t, \tag{17}$$

$$d\mathbb{E}[\lambda_t N_t] = \beta\lambda_\infty \mathbb{E}[N_t]\,\mathrm{d}t + (\alpha-\beta)\mathbb{E}[\lambda_t N_t]\,\mathrm{d}t + \mathbb{E}[\lambda_t^2]\,\mathrm{d}t + \alpha\mathbb{E}[\lambda_t]\,\mathrm{d}t, \tag{18}$$

$$d\mathbb{E}[\lambda_t^2] = (\alpha^2 + 2\beta\lambda_\infty)\mathbb{E}[\lambda_t]\,\mathrm{d}t + 2(\alpha-\beta)\mathbb{E}[\lambda_t^2]\,\mathrm{d}t. \tag{19}$$

These two lemmas allow the computation of different quantities useful to perform the estimation of the process. In fact, we have:

Proposition 1. *Given a Hawkes process $X_t = (\lambda_t, N_t)$ with dynamic given by* (4) *we have the following equalities:*

$$\lim_{t\to\infty} \mathbb{E}[N_{t+\tau} - N_t] = \frac{\beta\lambda_\infty}{\beta-\alpha}\tau = \Lambda\tau \tag{20}$$

with $\Lambda = \frac{\lambda_\infty}{1-\alpha/\beta}$ (the stationary regime expected intensity) gives the long-run expected value of the number of jumps during a time interval of length τ. The variance is:

$$\begin{aligned} \mathrm{V}(\tau) &= \lim_{t\leftarrow\infty} \mathbb{E}[(N_{t+\tau} - N_t)^2] - \mathbb{E}[N_{t+\tau} - N_t]^2 \\ &= \Lambda\left(\tau\kappa_-^2 + (1-\kappa_-^2)\frac{(1-\mathrm{e}^{-\tau\gamma_-})}{\gamma_-}\right), \end{aligned} \tag{21}$$

where

$$\Lambda = \frac{\lambda_\infty}{1-\alpha/\beta}, \quad \kappa_- = \frac{1}{1-\alpha/\beta}, \quad \textit{and} \quad \gamma_- = \beta - \alpha.$$

The covariance is given by:

$$\begin{aligned} \mathrm{Cov}(\tau,\delta) &= \lim_{t\to\infty} \mathbb{E}[(N_{t+\tau} - N_t)(N_{t+2\tau+\delta} - N_{t+\tau+\delta})] - \mathbb{E}[(N_{t+\tau} - N_t)]\mathbb{E}[(N_{t+2\tau+\delta} - N_{t+\tau+\delta})] \\ &= \frac{\lambda_\infty\beta\alpha(2\beta-\alpha)(\mathrm{e}^{(\alpha-\beta)\tau}-1)^2}{2(\alpha-\beta)^4}\,\mathrm{e}^{(\alpha-\beta)\delta} \end{aligned} \tag{22}$$

for $\delta > 0$.

The strategy of taking the limit to simplify the dependency of the results with respect to the initial value of the process, which is unknown, is borrowed from Aït-Sahalia, Cacho-Diaz, and Laeven (2010) who used the Hawkes process for modeling contagion effects between stocks. These results require the stability condition $\frac{\alpha}{\beta} < 1$, which allows us to put the process in its long-run stationary regime.

The autocovariance function contains information regarding the self-exciting or clustering property of the Hawkes process but it is more convenient to derive the autocorrelation function. We do not provide a proof for the following result as it is straightforward to obtain from the previous proposition.

Proposition 2. *Given a Hawkes process* $X_t = (\lambda_t, N_t)$ *with dynamic given by* (4) *the autocorrelation function of the number of jumps over a given interval* τ *is:*

$$\begin{aligned}\mathrm{Acf}(\tau,\delta) &= \lim_{t\to\infty} \frac{\mathbb{E}[(N_{t+\tau}-N_t)(N_{t+2\tau+\delta}-N_{t+\tau+\delta})] - \mathbb{E}[(N_{t+\tau}-N_t)]\mathbb{E}[(N_{t+2\tau+\delta}-N_{t+2\tau+\delta})]}{\sqrt{\mathrm{var}(N_{t+\tau}-N_t)\ \mathrm{var}(N_{t+2\tau+\delta}-N_{t+\tau+\delta})}} \\ &= \frac{\mathrm{e}^{-2\beta\tau}(\mathrm{e}^{\alpha\tau}-\mathrm{e}^{\beta\tau})^2\alpha(\alpha-2\beta)}{2(\alpha(\alpha-2\beta)(\mathrm{e}^{(\alpha-\beta)\tau}-1)+\beta^2\tau(\alpha-\beta))}\,\mathrm{e}^{(\alpha-\beta)\delta}, \end{aligned} \tag{23}$$

where the lag is δ.

The above expression is always positive when $\alpha < \beta$, which is the stability condition of the process and decays exponentially with the lag δ. The half-life depends on the difference $\alpha - \beta$. Note also that the background intensity λ_∞ is not involved in the autocorrelation, a property that could have been expected. The complete expressions for the third and fourth moments are provided, without proof, in the Appendix.

The computation performed above allows us to determine the moments up to the second order of $(X_t)_{t\geq 0}$ as well as the autocorrelation function for the number of jumps over an interval τ. Following this approach, we can compute higher-order moments. The key ingredient underlying the computations is the stability of the polynomial functions with respect to the infinitesimal generator of the Hawkes process. More precisely, the expected value of a polynomial function of the process $(X_t)_{t\geq 0}$ (i.e., $\sum_{i\leq n j\leq m} a_{ij}x^i x^j$) can be expressed as a function of polynomial functions of same or lower degree. This property is a consequence of the affine structure of the Hawkes process and has been used for the classical standard affine model of Duffie and Kan (1996) in Cuchiero, Keller-Ressel, and Teichmann (2012) and Filipović, Mayerhofer, and Schneider (2013). To be more precise, if we denote $Z_t = (\mathbb{E}[\lambda_t], \mathbb{E}[N_t], \mathbb{E}[\lambda_t^2], \mathbb{E}[\lambda_t N_t], \mathbb{E}[N_t^2])^\top$ then using the ODE obtained above the vector Z_t satisfies the ODE:

$$\frac{\mathrm{d}Z_t}{\mathrm{d}t} = AZ_t + B \tag{24}$$

with:

$$A = \begin{pmatrix} \alpha-\beta & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ \alpha^2+2\beta\lambda_\infty & 0 & 2(\alpha-\beta) & 0 & 0 \\ \alpha & \beta\lambda_\infty & 1 & \alpha-\beta & 0 \\ 1 & 0 & 0 & 2 & 0 \end{pmatrix} \qquad B = \begin{pmatrix} \beta\lambda_\infty \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

The solution of this ODE is given by:

$$Z_t = \mathrm{e}^{At}Z_0 + \int_0^t \mathrm{e}^{A(t-s)}B\,\mathrm{d}s, \tag{25}$$

where the exponential of the matrix is computed using classical algorithms; see Golub and Van Loan (1996).

2.3. Inference Strategies

In this section, we first present the classical Maximum Likelihood approach usually used to calibrate the Hawkes process and underline the numerical difficulties. Then using the explicit expression for the moments and the autocorrelation function computed in the previous section we develop a method of moment estimation strategy whose computational speed appears to be very fast compared to the existing alternatives.

2.3.1. Maximum likelihood estimation

Let $(X_t)_{t\geq 0}$ be a simple point process on $[0, T]$ and $t_1 \cdots t_{N_T}$ denote a realization of $(N_t)_{t\geq 0}$ over $[0, T]$, then, as established in proposition 7.2.III of Daley and Jones (2002), the log-likelihood of $(X_t)_{t\geq 0}$ is of the form:

$$\begin{aligned} L &= \int_0^T (1-\lambda_s)\,\mathrm{d}s + \int_0^T \ln(\lambda_s)\,\mathrm{d}N_s \\ &= \int_0^T (1-\lambda_s)\,\mathrm{d}s + \sum_{i=1}^{N_T} \ln(\lambda_{t_i}). \end{aligned}$$

In the case of a Hawkes process we have:

$$\begin{aligned} L &= \int_0^T (1-\lambda_t)\,\mathrm{d}t + \int_0^T \ln(\lambda_t)\,\mathrm{d}N_t \\ &= \int_0^T \left(1-\lambda_\infty + \int_0^t \alpha\,\mathrm{e}^{-\beta(t-s)}\,\mathrm{d}N_s\right)\mathrm{d}t + \int_0^T \ln\left(\lambda_\infty + \int_0^t \alpha\,\mathrm{e}^{-\beta(t-s)}\,\mathrm{d}N_s\right)\mathrm{d}N_t \\ &= T - T\lambda_\infty + \sum_{i=1}^{N_T}\int_0^T \alpha\,\mathrm{e}^{-\beta(t-t_i)} \times 1_{\{t\geq t_i\}}\,\mathrm{d}t + \sum_{i=1}^{N_T} \ln\left(\lambda_\infty + \sum_{t_j\leq t_i} \alpha\,\mathrm{e}^{-\beta(t_i-t_j)}\right) \end{aligned}$$

and simplifying the above integral

$$\begin{aligned} \int_0^T \alpha\,\mathrm{e}^{-\beta(t-t_i)} \times 1_{\{t\geq t_i\}}\,\mathrm{d}t &= \left[-\frac{\alpha}{\beta}\,\mathrm{e}^{-\beta(t-t_i)} \times 1_{\{t\geq t_i\}}\right]_0^T - \int_0^T -\frac{\alpha}{\beta}\,\mathrm{e}^{-\beta(t-t_i)} \times \delta_{\{t=t_i\}}\,\mathrm{d}t \\ &= \frac{\alpha}{\beta} - \frac{\alpha}{\beta}\,\mathrm{e}^{-\beta(T-t_i)}, \end{aligned}$$

we end with:

$$L = T - T\lambda_\infty - \sum_{i=1}^{N_T} \frac{\alpha}{\beta}(1-\mathrm{e}^{-\beta(T-t_i)}) + \sum_{i=1}^{N_T} \ln(\lambda_\infty + \alpha A(i)), \tag{26}$$

where $A(i) = \sum_{t_j\leq t_i} \mathrm{e}^{-\beta(t_i-t_j)}$.

The estimation leads to a nonlinear optimization algorithm such as Nelder–Mead to find the maximum of this function. We stress that for each set of parameters the evaluation of this function requires a loop over the observations that, for the problem at hand, trade clustering, is very large.

Some authors, such as Ozaki (1979), pointed out that $A(i)$ used in (26) satisfies a recursive relation. Indeed, defining $A(1) = 0$ then for $i \geq 1$, $A(i+1) = \mathrm{e}^{-\beta(t_{i+1}-t_i)} \times (1 + A(i))$, simplifies the calculation of the likelihood function and speeds up evaluation. However, the calibration still takes a few minutes and a large number of function calls are performed. Any simplification of the calibration procedure is therefore of interest.

2.3.2. *Fast Hawkes process calibration*

Even with the improvement previously presented the parameter estimation procedure based on the maximum likelihood function is still very time-consuming. Having computed explicitly the moments as well as the autocorrelation function for the Hawkes process a natural estimation strategy is the generalized method of moments; see Hall (2004) for an exhaustive treatment, and Bollerslev and Zhou (2002, 2004) for an application to finance. The inference problem now writes as:

$$\hat{\theta} = \operatorname{argmin}\{(M - f(\theta))^{\top} W (M - f(\theta))\}, \tag{27}$$

where M is the vector of empirical moments (eventually related to the autocorrelation function), $f(\theta)$ is the vector of corresponding theoretical moments, and W is a symmetric positive definite weighting matrix. The optimization problem can then be solved very quickly by Levenberg–Marquardt algorithm (we use the implementation provided by Lourakis (2004)).

This method of estimation is known to be consistent and asymptotically normal. It can also be shown that with a suitable choice of the weighting matrix W, the estimator is asymptotically efficient in the sense that it has the smallest covariance matrix (the usual iterative algorithm to select this matrix can be used). As we are primarily interested in speed we fix ex ante this matrix W to reduce the discrepancies between the components of the vector involved in the objective function and in that case the optimization problem turns out to be the simple least squares method (LS in the sequel). Some numerical experiments lead us to the following conclusions: the optimization problem (27) based on the mean and variance of number jumps during an interval τ (i.e., Eqs. (20) and (21)), and autocorrelation function (23) gives good results if calibration quality and speed are taken into account. Computation time remains negligible and is faster than the MLE.

From a numerical point of view we found it simpler and more robust to work with normalized quantities, in which case the optimization problem is:

$$\hat{\theta} = \operatorname{argmin} \left\{ \left(1 - \frac{f(\theta)}{M} \right)^{\top} W \left(1 - \frac{f(\theta)}{M} \right) \right\}, \tag{28}$$

where the components of the vector $\left(1 - \frac{f(\theta)}{M}\right)$ are $\left(1 - \frac{f_i(\theta)}{M_i}\right)$ and involve the relative ratio of the theoretical moment to its empirically estimated counterpart. This is made possible because all the considered moments are different from zero.[3] In that case we can choose for W the identity, thus simplifying the specification of this matrix tremendously. Let us emphasize that the evaluation of the empirical moments is only made *once* during the optimization procedure and explains why this estimation strategy is intrinsically faster than the MLE.

This gives us a very appealing estimation procedure. Not only is it instantaneous and this point is crucial if the objective is to apply a model to a high-frequency problem (optimal

[3] A similar scaling appears in Aït-Sahalia et al. (2010) where the authors use the matrix W to control the discrepancies between the moments.

TABLE I
Estimation Strategies Comparison

	MLE			*Fast Calibration I*			*Fast Calibration II*		
	λ_∞	α	β	λ_∞	α	β	λ_∞	α	β
Mean	0.33	0.16	0.23	0.68	−0.26	−0.23	0.18	0.07	0.25
SD	5.74	7.13	6.69	8.39	10.66	11.04	4.41	7.47	7.69

Note. A Monte Carlo experiment comparing the MLE and fast calibration estimation strategies. Fast calibration I is an LS estimation based on the first, second moments and the autocorrelation function computed for several lags. Fast calibration II is a LS estimation based on the autocorrelation function computed for several lags (λ_∞ is then deduced using (20)). For each method and each parameters we report the mean relative error value and the corresponding standard deviation (both expressed in percent).

execution, price impact analysis of a trade), but it also has the advantage of being robust against data pollution, an aspect that is very common in such data and this point was already underlined in Bacry et al. (2013a).

Lastly, let us also mention that our closed form solutions make it possible to perform forecast experiments as they are usually done for the realized volatility (see Meddahi, 2013).

2.3.3. *Robustness check of the fast calibrations*

To assess the quality of our estimation procedure we perform the following Monte Carlo experiment. We randomly generate the parameters λ_∞, α, and β, with $\beta > \alpha$. We choose them in the interval [0 , 1] because these are the values obtained in our empirical study. We simulate a Hawkes process for a duration of $T = 8$ hours, corresponding to a trading day, using the thinning algorithm as presented in Ogata (1981). We carry out a calibration using both the MLE and the LS estimators. For the MLE, calibration is conducted using Nelder-Mead algorithm as implemented in the open-source library NL-opt.[4] The LS estimations are conducted using the Levenberg–Marquardt algorithm provided by Lourakis (2004). The initial guess for the parameters is their true value multiplied by a uniform random variable on the interval [0.5 1.5] (both algorithms have the same starting point). The estimation test is performed 25,000 times for each algorithm. Table I contains the results. The mean relative error value expressed in percent is reported for each parameter. For the standard deviation we compute the root mean squared relative error value that we also express in percent.

We perform two fast calibrations. The first is based on the least squares optimization problem (28) with W the identity, the first two moments given by (20) and (21) computed with $\tau = 60$ seconds, as well as the autocorrelation function (23) with $\tau = 60$ seconds and lags ranging from 0 to 600 seconds (by step of 60 seconds). For the second fast calibration, the objective function (28) only depends on the autocorrelation function (23) with $\tau =$ 60 seconds and lags ranging from 0 to 600 seconds (by step of 60 seconds). To obtain λ_∞, because the autocorrelation function does not depend on it, we use (20) with $\tau = 60$ seconds. That is, we suppose this equation is satisfied without error.

The MLE leads to an error of 0.33% for λ_∞ and similar magnitudes for α and β as we have 0.16% and 0.23%, respectively. For example, the difference between the estimated λ_∞ and the true λ_∞ is 0.33% of the true value for this estimator. Overall, the estimation procedure is satisfactory as mean values and standard deviations are small. For the first fast calibration, the results are reported in column titled "Fast calibration I", the mean values are

[4]See `http://ab-initio.mit.edu/wiki/index.php/NLopt`.

also small, they are comparable to those obtained with the MLE, and the standard deviation values are roughly 1.5 times those for the MLE but remain acceptable. In conclusion, the fast calibration algorithm performs well and the computational cost is much lower. For the second fast calibration (i.e., column "Fast calibration II") similar conclusions are obtained. Namely, the mean values and the standard deviations are small. Overall, the fast calibration algorithms perform well.

3. APPLICATIONS

3.1. Data

We rely on tick-by-tick data from Thomson Reuters Tick History (TRTH). The data consist of trades and quotes files timestamped in milliseconds. We studied two stocks in particular: BNP Paribas and Sanofi, as well as the futures on the Eurostoxx and the Dax. For each studied day we consider the futures with the shortest maturity date. The data cover the period between January 01, 2010 to December 31, 2011.

All the trading days begin at 9:00 and end at 17:30. We neglect the first and last 15 minutes in order to avoid the open and close auctions. We end with 8 trading hours per day, between 9:15 and 17:15.

Most of our study deals with trade time arrivals and statistics on the number of trades occurring on intervals of fixed length. The fact that the timestamps have a bounded precision, the millisecond in our case, is another appealing feature of our method of estimation. Indeed, many trades will have the same time to the nearest millisecond even if they did not take place at the same time. This millisecond will count as a unique entry in the ML estimation procedure; in the moment-based inference all the trades will be taken into account when computing the moments. This, among other reasons, will make the method of moments more robust to data imprecision, a fact typical to high-frequency data.

3.2. Trade Clustering

As outlined in the Introduction, trading activity is not a completely random and memoryless process. Consequently, the Poisson process is not suitable for modeling trade arrival times. As shown in Figure 1, a QQplot of interarrival times of trades against an exponential distribution clearly rejects the Poisson process as the data-generating process for the order flow.

In fact, trades tend to cluster and an illustration is given in Figure 2 where we plot an histogram of the number of trades occurring every minute during a trading day for the Eurostoxx. The clustering is graphically clear.

Numerous reasons can explain this clustering of trade arrival times, among them liquidity takers splitting their orders so as to minimize their market impact, or insider traders reacting rapidly to take advantage from information they have before it is widespread in the market: these justify a one-sided trade clustering (i.e., either buy or sell initiated trades). On the other hand, heterogeneity of market participants is responsible for the two-sided trade clustering. A complete study can be found in Sarkar and Schwartz (2006).

To quantify this clustering in time we compute the correlation of the number of trades occurring during different time intervals of fixed length separated by a time lag. A plot of this autocorrelation as a function of the lag gives information about the degree of clustering. If the data were generated by a Poisson process this autocorrelation would be equal to zero.

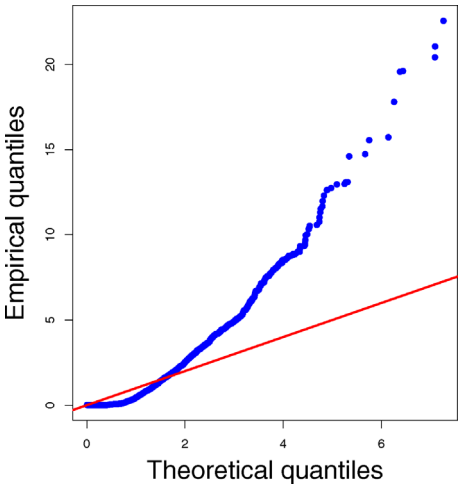

FIGURE 1
QQplot of inter-trade durations against exponential distribution. Interarrival trade times are clearly not exponential. Graph for Eurostoxx futures trades on March 03, 2011 for the first trading hour. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

We want to compute:

$$C(\tau, \delta) = \frac{\mathbb{E}[(N_{t+\tau} - N_t)(N_{t+2\tau+\delta} - N_{t+\tau+\delta})] - \mathbb{E}[(N_{t+\tau} - N_t)]\mathbb{E}[(N_{t+2\tau+\delta} - N_{t+\tau+\delta})]}{\sqrt{\text{var}(N_{t+\tau} - N_t)\text{var}(N_{t+2\tau+\delta} - N_{t+\tau+\delta})}}. \tag{29}$$

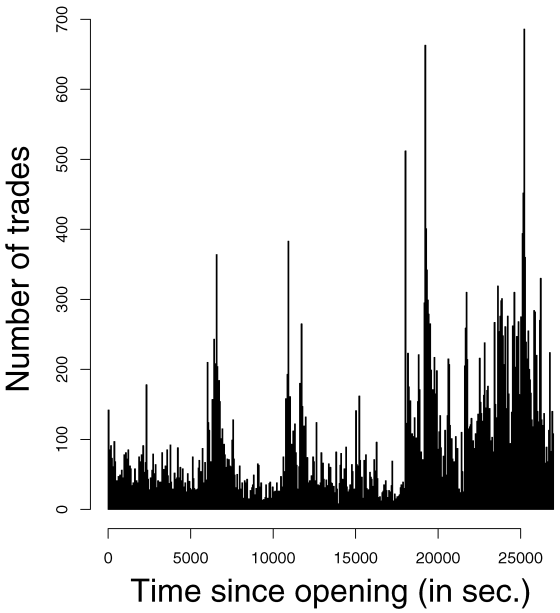

FIGURE 2
Number of trades by time intervals of 1 minute: presence of clusters is graphically apparent. Graph for Eurostoxx futures trades on March 03, 2011.

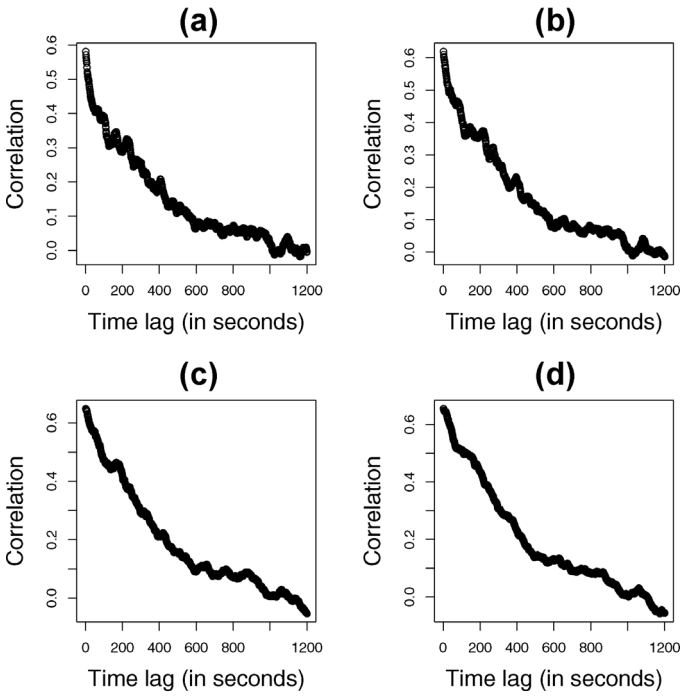

FIGURE 3

Empirical autocorrelation function $\delta \to C(\tau, \delta)$, given by (29), of the number of trades occurring on consecutive intervals of length τ separated by a time lag δ as a function of the lag δ. In the four plots, we change the length of the time interval τ considered. (a) corresponds to $\tau = 20$ seconds, (b) $\tau = 30$ seconds, (c) $\tau = 60$ seconds and (d) $\tau = 90$ seconds. The shape of the function remains identical, even if the plot is noisier when the time interval length decreases. The data are Eurostoxx futures trades on January 07, 2010.

To this end, we count the number of trades occurring during two sliding non-overlapping intervals of length $\tau = 60$ seconds separated by a certain time lag δ. We change this time lag δ from 1 second to 20 minutes by a step of 1 second. This gives 1,200 number of trades autocorrelation points for the function $\delta \to C(\tau, \delta)$ that we report in Figure 3 for different time intervals τ for a given symbol (in that case the Eurostoxx). We clearly see that the autocorrelation is positive and significant and that it decreases with the time lag. It is also remarkable that this is true for all the time interval lengths considered, and that independently from τ this memory effect seems to take around 10 minutes to become insignificant.

Figure 4 confirms that the same phenomenon is observed for all the tested symbols. The absolute value of the autocorrelation is higher for the two futures, which are far more liquid than the stocks, but the same decreasing shape is observed and the time life of this autocorrelation seems to be very close for all the symbols.

Overall, these stylized facts justify the use of the Hawkes process as modeling framework.

Thanks to the closed form solutions of the previous section we estimate the parameters using the fast calibration algorithm presented in the inference section. To further reduce computational cost the objective function (28) is such that it depends *only* on the empirical autocorrelation function (29) and its analytical counterpart (23).[5] As the autocorrelation

[5]Compared with the robustness check performed previously it means that we drop the first and second moments.

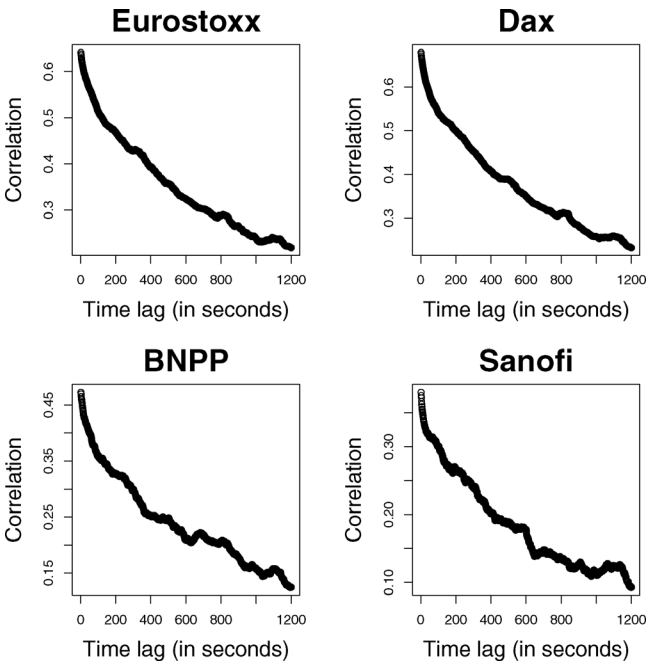

FIGURE 4

Empirical autocorrelation function $\delta \to C(\tau, \delta)$ of the number of trades with $\tau = 60$ seconds for Eurostoxx, Dax, BNPP, and Sanofi, averaged every day for the month January 2010.

function of the number of trades is independent of λ_∞ we rely on Equation (20) to obtain λ_∞ from the other parameters.[6] Based on Figure 5 we choose to fit the autocorrelation function (23) for $\tau = 60$ seconds and δ ranging from 0 to 600 seconds by step of 60 seconds because Figure 4 convinces us that this choice is suitable for all the analyzed stocks. For each symbol we perform a daily calibration using the 2-year sample described in the data section. In Table II, we report the mean and median estimated values as well as the standard deviations. The table also contains in column "Poisson-λ" the intensity we obtain if a Poisson process is calibrated.

A striking fact is that λ_∞ is much smaller than the Poisson-λ as the former amounts to only 5% to 10% of the latter. λ_∞ represents the background activity of the Hawkes process whereas from Proposition 1 the long-term expected intensity, which will be close to the Poisson intensity, is $\Lambda = \frac{\lambda_\infty}{1-\alpha/\beta}$. As such, the discrepancy between λ_∞ and the Poisson intensity is precisely due to α, which controls the self-exciting property of the Hawkes process. Our results suggest that the fraction of events due to the background process activity is small in comparison of the part due to the branching structure of the process (i.e., the self-exciting property captured by α). For the four symbols the parameters have the same size and the standard deviations are small. Note also that the stability condition ($\alpha < \beta$) is satisfied, although the optimization algorithm is performed without constraints. The quality of the fit can be assessed graphically, we report an example of such fit in Figure 6; it suggests that the Hawkes process captures well the empirical property of trade arrival times.

For robustness, we also conduct calibrations of the model on a variety of traded assets ranging from interest rate futures to commodity, energy, and foreign exchange futures. Results are reported in Table III.

[6]From a statistical point of view it means that we suppose Equation (20) is satisfied without error.

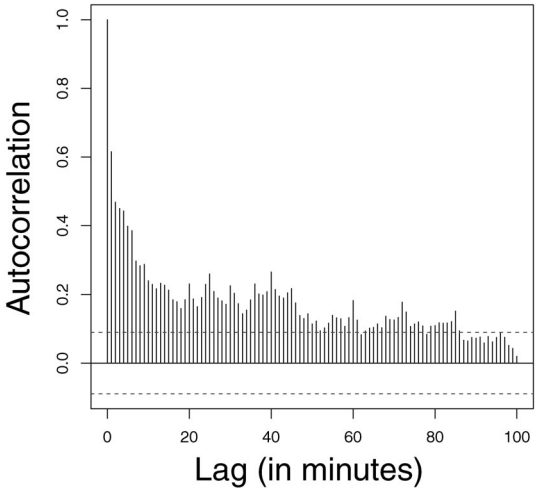

FIGURE 5
Number of trades empirical autocorrelation. The number of trades are computed for an interval $\tau = 60$ seconds, and the lag δ is now measured in *minutes*. The symbol is Eurostoxx on October 26, 2010.

For all the assets, the standard deviation values are small compared to the mean parameter values and the similarity between the mean and median values, denoting the relative stability of the calibration and absence of outliers, confirms the ability of the Hawkes process to fit the data. As for the four previous assets all the calibrations lead to solutions that satisfy the stability constraint $\alpha < \beta$, albeit no constraints were used during the optimization procedure. Lastly, let us underline the fact that $\alpha - \beta$ controls the decreasing shape of the autocorrelation function as it is clear from Equation (23) and quantifies to which extent the data-generating process departs from a Poisson process or equivalently to which extent the clustering behavior of the Hawkes process is fruitful.

TABLE II
Calibration Results

Symbol	*Measure*	*Poisson-*λ	λ_∞	α	β
Eurostoxx	Mean	1.4480	0.0625	0.0869	0.0911
	SD	0.6283	0.0209	0.0229	0.0237
	Median	1.3343	0.0593	0.0843	0.0882
Dax	Mean	1.7814	0.0664	0.0993	0.1034
	SD	0.7322	0.0249	0.0218	0.0226
	Median	1.6214	0.0609	0.0988	0.1028
BNPP	Mean	0.8627	0.0556	0.0760	0.0819
	SD	0.3923	0.0231	0.0192	0.0219
	Median	0.7438	0.0508	0.0724	0.0772
Sanofi	Mean	0.6704	0.0453	0.0747	0.0806
	SD	0.1873	0.0213	0.0212	0.0240
	Median	0.6087	0.0414	0.0700	0.0758

Note. Calibration results for two years of data. We calibrate daily a Hawkes process to the trade arrival times for each symbol. For comparison, we put the Poisson equivalent λ, defined as the mean number of trades per second for every day. For every measure, we put mean, standard deviation, and median values for the calibrated parameters.

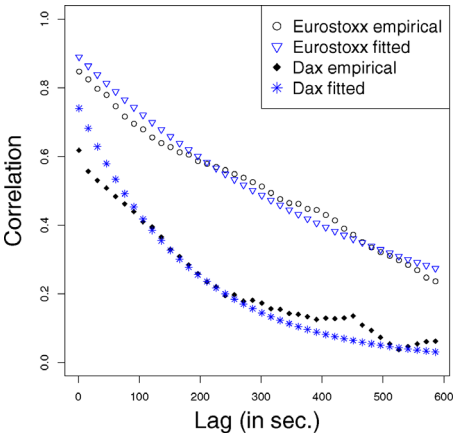

FIGURE 6

Empirical autocorrelation of the time series of the number of trades occurring during $\tau = 60$ seconds versus the theoretically fitted one. For the Dax, estimated parameters are : $\lambda_\infty = 0.0326806$, $\alpha = 0.0431643$ and $\beta = 0.0486235$ and for Eurostoxx, estimated parameters are : $\lambda_\infty = 0.033282$, $\alpha = 0.04259$ and $\beta = 0.0446049$. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

3.3. Branching Structure of Trading Activity

An interesting property of the Hawkes process is its branching structure. Indeed, the occurrence of a jump increases the intensity of the process, thereby the probability to observe another jump. As pointed out by Hewlett (2006), this results in a direct and indirect impulse response of the process intensity to a jump event. Denoting the expected increase of the process intensity at time t as a response to a jump occurring at time 0 by $f(t)$, we have the following decomposition:

- Direct response: an increase of the intensity by α that will decay exponentially as time passes, thus leading to an increase of the intensity at time s of $\alpha\, e^{-\beta s}$.
- Indirect response: at any time s between 0 and t, the direct increase of the intensity by $\alpha\, e^{-\beta s}$ leads to an indirect increase of the expected number of jumps at time t, which equals to $\alpha\, e^{-\beta s} ds f(t-s)$; we then need to integrate over the range $[0\ t]$ to obtain the total indirect effect.

As a consequence, the expected direct and indirect increase of the intensity at time t caused by a jump at time 0 writes as:

$$f(t) = \alpha\, e^{-\beta t} + \int_0^t \alpha\, e^{-\beta s} f(t-s)\, ds.$$

The solution of this integral equation is given by:

$$f(t) = \alpha\, e^{-(\beta-\alpha)t}.$$

Therefore, the N_{response}, which is the expected number of jumps triggered by one jump occurring at time 0 if the process is observed indefinitely (the impulse response), is:

$$N_{\text{response}} = \int_0^\infty f(s)\, ds = \frac{\alpha}{\beta - \alpha}. \tag{30}$$

TABLE III
Calibration Results for Other Assets

Symbol	*Measure*	*Poisson* λ	λ_∞	α	β	$N_{response}$
Bund	Mean	1.2742	0.0671	0.0956	0.1013	19.683
	SD	0.4434	0.0244	0.0218	0.0232	8.3592
	Median	1.2063	0.0645	0.0934	0.0983	18.256
Bobl	Mean	0.6699	0.0546	0.0816	0.0894	13.236
	SD	0.1705	0.0216	0.0225	0.0257	6.4551
	Median	0.6173	0.0498	0.0790	0.0858	11.329
Schatz	Mean	0.6245	0.0473	0.0877	0.0952	14.896
	SD	0.1397	0.0209	0.0211	0.0236	7.4515
	Median	0.5712	0.0440	0.0835	0.0903	13.556
JPY	Mean	1.6023	0.0536	0.1130	0.1173	29.737
	SD	0.7144	0.0172	0.0207	0.0215	11.123
	Median	1.5510	0.0518	0.1133	0.1178	29.139
EURO	Mean	4.1955	0.0788	0.1220	0.1245	53.526
	SD	1.7419	0.0271	0.0192	0.0194	18.999
	Median	4.1956	0.0770	0.1258	0.1282	52.045
GOLD	Mean	2.3191	0.0852	0.1104	0.1149	27.716
	SD	0.8824	0.0290	0.0237	0.0244	10.472
	Median	2.1555	0.0815	0.1175	0.1212	25.743
Crude Oil Brent	Mean	2.0453	0.0550	0.1243	0.1279	37.528
	SD	0.7018	0.0154	0.0143	0.0147	12.877
	Median	1.9787	0.0535	0.1255	0.1302	36.86
Natural GAS	Mean	1.4524	0.0688	0.1177	0.1241	21.17
	SD	0.4362	0.0181	0.0156	0.0166	8.4653
	Median	1.3653	0.0680	0.1168	0.1246	18.532
Sugar	Mean	0.8082	0.0434	0.1213	0.1289	19.964
	SD	0.3214	0.0196	0.0174	0.0190	11.93
	Median	0.6869	0.0382	0.1272	0.1351	18.539
CORN	Mean	1.0338	0.0626	0.1069	0.1146	17.348
	SD	0.4563	0.0213	0.0332	0.0334	9.3997
	Median	0.9563	0.0537	0.1226	0.1312	17.333
WHEAT	Mean	1.1562	0.0639	0.1119	0.1193	18.807
	SD	0.4334	0.0242	0.0215	0.0227	8.9511
	Median	1.0926	0.0594	0.1182	0.1244	17.372

Note. Calibration results for 2 years of data (2010 and 2011) and for different asset classes (interest rates, foreign exchange, metal commodities) to assess the robustness of the model. Schatz, Bobl, and Bund are, respectively, the 2-year, 5-year, and 10-year futures on German government bonds. For each asset class we take for daily data the front maturing futures to perform the calibration. The last column N_{response} will be defined in Section 3.3.

We use our daily calibrations on real data of the Hawkes process to measure an average N_{response} for the studied assets and report the results in Table IV.

We clearly observe a difference between futures, given by the symbols Dax and Eurostoxx, and the stocks represented by BNPP and Sanofi. This points toward considering this number as an indicator of liquidity and trading activity. Formula (30) suggests the ratio $\frac{\beta}{\alpha}$ as the key quantity to evaluate the impulse response value and the numbers are consistent with the fact that futures are more actively traded than the stocks due to a stronger branching structure (controled by α and β).

A robustness check was also performed for other assets in Table III. According to this measure a trade on the Bund triggers more other trades than does a trade on the Bobl. The Euro currency, and to a lesser extent the JPY, seem to be more reactive markets than the

TABLE IV
Market Liquidity Indicator

Symbol	*Average* $N_{response}$
Dax	26
Eurostoxx	22
BNPP	14
Sanofi	10

Note. $N_{response}$ as a characteristic of market liquidity. Symbols are ranked from the most liquid to the less liquid.

others. Among the commodities, the Crude Oil Brent dominates the Natural Gas, Sugar, Corn, and the Wheat.

3.4. Diffusive Limit and Signature Plot

In all the preceding sections we dealt with the trading process from a microscopic point of view, that is to say at transaction level. In the classical high-frequency literature, mainly developed by econometricians, most of the studies are carried out at this time scale and aim at explaining the price formation process. Many models encompass the subtle interactions of the many components of the trading process such as order flow, order signs, volumes, and other quantities to achieve a microscopic foundation of the price process. For instance, the ACD models, as proposed in Engle and Russell (1998), fall in this trend of research. For an overview of models and techniques we refer to Hautsch (2012) and references therein.

In our study, we consider a simpler framework focusing only on the order flow. Obviously, this simplification comes at the cost of neglecting important aspects of the price formation process, such as the volume for example, but it allows us to address the important question of connecting the microscopic price formation process observed at transaction level to its macroscopic properties at a coarser time scale. In other words, we connect the stochastic differential equations used to model an asset price evolution at a daily frequency, such as in the Black–Scholes model, which relies mainly on the continuous Brownian motion, to the discontinuous point process describing individual transactions. The Hawkes process, thanks to its strong analytical tractability, enables us to relate these two time scales.

In recent years, many authors developed this bottom-up point of view in price modeling, establishing connections between order-book level price formation mechanisms and statistical macroscopic price properties. Among other references let us mention Abergel and Jedidi (2013) whose model for the order book is based on a multidimensional Markov chain with independent Poissonian order arrival times. Within their framework they prove the convergence of the price process to a Brownian motion, thus bridging the gap between high-frequency and low-frequency quantities. In Cont and De Larrard (2011, 2012), the order book is described as a Markovian queuing system for which the authors establish a diffusive limit and calculate some quantities of interest such as volatility. In Kirilenko et al. (2013), the authors use this same idea of different time scales and relate microscopic causes to macroscopic effects, they study the influence of high-frequency traders on asset volatility. Bacry et al. (2013a) introduces a model for microstructure price evolution based on mutually exciting Hawkes processes. They connect the signature plot of volatility and Epps effect of asset correlations to the model parameters driving the price process. Also of great interest is Bacry et al. (2013b)where the authors establish diffusive limits for such kind of models.

In this section, we consider the modeling framework proposed by Bacry et al. (2013a)[7] and develop a toy model for the movements of the *mid* price of a traded asset using the Hawkes processes presented in the analytical part. The *mid* price is the mean of the best ask price and the best bid price in the order book. As the best ask price or the best bid price moves up (down) by one tick, the mid price will move (down) by *half* a tick. Despite its simplicity, the model captures the essential features of the price process.[8] The model writes:

$$S_t = S_0 + (N_t^{\text{up}} - N_t^{\text{down}})\frac{\delta}{2}, \tag{31}$$

where δ is the tick value. The N_t^{up} and N_t^{down} are Hawkes processes capturing the up and down jumps of the mid price. Both of them follow a dynamic of the form (4). We consider them independent and with the same parameters in order to avoid price explosion. In the stationary regime their intensities are given by:

$$\lambda_t^{\text{up}} = \lambda_\infty + \int_0^t \alpha\, \mathrm{e}^{-\beta(t-s)}\, \mathrm{dN}_s^{\text{up}}, \tag{32}$$

$$\lambda_t^{\text{down}} = \lambda_\infty + \int_0^t \alpha\, \mathrm{e}^{-\beta(t-s)}\, \mathrm{dN}_s^{\text{down}}. \tag{33}$$

Later, we provide the specification of Bacry et al. (2013a)'s model and explain the differences. In order to relate this high-frequency description for the price to its low-frequency description behavior we need a limit theorem. In Bacry et al. (2013b), the authors rely on the martingale theory and limit theorems for semi-martingales to prove stability and convergence results for a general model with mutually exciting processes and a generic kernel.[9] In our case, as the kernel is exponential the process $X_t = (S_t, N_t^{\text{up}}, \lambda_t^{\text{up}}, N_t^{\text{down}}, \lambda_t^{\text{down}})$ is a Markov process. Its infinitesimal generator writes:

$$\begin{aligned}
\mathcal{L}f(x) = {} & \beta(\lambda_\infty - \lambda_t^{\text{up}})\frac{\partial f}{\partial \lambda^{\text{up}}}(x) + \beta(\lambda_\infty - \lambda_t^{\text{down}})\frac{\partial f}{\partial \lambda^{\text{down}}}(x) \\
& + \lambda_t^{\text{up}}\left[f\left(S_t + \frac{\delta}{2}, N_t^{\text{up}} + 1, \lambda_t^{\text{up}} + \alpha, N_t^{\text{down}}, \lambda_t^{\text{down}}\right) - f(x)\right] \\
& + \lambda_t^{\text{down}}\left[f\left(S_t - \frac{\delta}{2}, N_t^{\text{up}}, \lambda_t^{\text{up}}, N_t^{\text{down}} + 1, \lambda_t^{\text{down}} + \alpha\right) - f(x)\right].
\end{aligned}$$

The explicit form of the infinitesimal generator allows us to apply Foster–Lyapounov techniques in order to establish stability results. We refer to Meyn and Tweedie (2009) for a detailed exposition on such techniques as well as stochastic stability concepts. For instance, ergodicity of the process X_t, that is to say its convergence to a stationary regime, can be easily established thanks to the Foster–Lyapounov test function criterion. In our case, we define

[7] We might use the symbol "BDHM," which stands for Bacry, Delattre, Hoffmann, and Muzy.

[8] For instance, the so-called trade-throughs Pomponio and Abergel (2013) (i.e., trades consuming many successive limits and then moving the best quote by more than one tick) can be regarded in this model as successive one tick movements occurring very closely in time.

[9] The function $g(t) = \alpha\, \mathrm{e}^{-\beta t}$ is called the kernel of the Hawkes process. Other forms are possible but this choice leads to the most tractable Hawkes process.

the function $V(x) = \frac{\lambda^{\text{up}}+\lambda^{\text{down}}}{2\lambda_\infty}$, then a simple calculation yields the *geometric drift condition*:

$$\mathcal{L}V(x) \leq (\alpha - \beta)V(x) + \beta. \tag{34}$$

As $\alpha < \beta$ this condition ensures the mean reversion of the process and thanks to Theorems 6.1 and 7.1 in Meyn and Tweedie (1993) (and especially (CD3)), the V-uniform ergodicity of the process X_t.

Let us then write unit-time price increments:

$$\eta_i = [(N_i^{\text{up}} - N_{i-1}^{\text{up}}) - (N_i^{\text{down}} - N_{i-1}^{\text{down}})] \times \frac{\delta}{2}$$

and consider the random sums:

$$S_n = \sum_{i=1}^{n} \eta_i$$

with $\{\eta_i; i = 1, \ldots, n\}$ being the price increments (note that $\mathbb{E}[\eta_i] = 0$). We focus on the asymptotic behavior of the rescaled price process:

$$\bar{S}_t^n = \frac{S_{\lfloor nt \rfloor}}{\sqrt{n}}.$$

The V-uniform ergodicity and Theorem 16.1.5 in Meyn and Tweedie (2009) allow us to conclude that the increments are geometrically mixing, and Theorem 19.3 of Billingsley (1999) proves that $\bar{S}_t^n$ converges to a Brownian motion in the sense of Skorokhod topology:

$$\bar{S}_t^n \to \sigma W_t.$$

Moreover, calculations done before for the moments of the Hawkes process increments lead to a very simple expression for the volatility. In fact, we have:

$$\begin{aligned}
\sigma^2 &= \lim_{n\to\infty} \frac{\text{Var}(S_n)}{n} \\
&= 2\frac{\delta^2}{4}(\mathbb{E}[(N_1^{\text{up}} - N_0^{\text{up}})^2] - \mathbb{E}[N_1^{\text{up}} - N_0^{\text{up}}]^2) \\
&\quad + 4\frac{\delta^2}{4}\sum_{i=1}^{\infty} \mathbb{E}[(N_1^{\text{up}} - N_0^{\text{up}})(N_{1+i}^{\text{up}} - N_i^{\text{up}})] - \mathbb{E}[N_1^{\text{up}} - N_0^{\text{up}}]\mathbb{E}[N_{1+i}^{\text{up}} - N_i^{\text{up}}] \\
&= \frac{\delta^2}{2}\left(V(1) + 2\sum_{k=0}^{\infty} \text{Cov}(1,k)\right),
\end{aligned}$$

where thanks to (22) $\text{Cov}(1,k)$ writes:

$$\text{Cov}(1,k) = \frac{\lambda_\infty \beta\alpha(2\beta - \alpha)(\mathrm{e}^{(\alpha-\beta)} - 1)^2}{2(\alpha - \beta)^4}\,\mathrm{e}^{(\alpha-\beta)k}.$$

Then, summing up with the expression of the variance and after some simplifications we obtain:

$$\sigma^2 = \frac{\delta^2}{2} \frac{\lambda_\infty}{\left(1 - \frac{\alpha}{\beta}\right)^3}. \tag{35}$$

Note the dependence of the volatility on the ratio $\frac{\alpha}{\beta}$. The larger this ratio is, the larger is the volatility is (under the hypothesis of stability $\alpha < \beta$). An upward (downward) chock is likely to trigger another upward (downward) chock if this ratio is large, and therefore it induces a positive autocorrelation for the mid price and a more persistent path with the effect of increasing asset's volatility.

Besides giving a framework that allows the connection of the microscopic price formation mechanism to its macroscopic behavior, as shown above, the Hawkes process can reproduce some stylized facts across time scales. Among these stylized facts is the volatility signature plot, which depends on the realized variance over a period T calculated by sampling the data by time intervals of length τ. Within the toy model (31) we have:

$$\begin{aligned}
\hat{C}(\tau) &= \frac{1}{T} \sum_{n=0}^{T/\tau-1} (S_{(n+1)\tau} - S_{n\tau})^2 \\
&= \frac{1}{T} \sum_{n=0}^{T/\tau-1} ((N^{\text{up}}_{(n+1)\tau} - N^{\text{up}}_{n\tau}) - (N^{\text{down}}_{(n+1)\tau} - N^{\text{down}}_{n\tau}))^2 \frac{\delta^2}{4} \\
&= \frac{1}{T} \sum_{n=0}^{T/\tau-1} (N^{\text{up}}_{(n+1)\tau} - N^{\text{up}}_{n\tau})^2 \frac{\delta^2}{4} + \frac{1}{T} \sum_{n=0}^{T/\tau-1} (N^{\text{down}}_{(n+1)\tau} - N^{\text{down}}_{n\tau})^2 \frac{\delta^2}{4} \\
&\quad - 2\frac{1}{T} \sum_{n=0}^{T/\tau-1} (N^{\text{up}}_{(n+1)\tau} - N^{\text{up}}_{n\tau})(N^{\text{down}}_{(n+1)\tau} - N^{\text{down}}_{n\tau}) \frac{\delta^2}{4}.
\end{aligned}$$

By definition the mean signature plot is the expectation of the above quantity and can be computed explicitly as we have (we omit the proof as it is straightforward):

Proposition 3. *The mean signature plot, defined by* $C(\tau) = \mathbb{E}[\hat{C}(\tau)]$, *is given by:*

$$\begin{aligned}
C(\tau) &= \mathbb{E}[\hat{C}(\tau)] \\
&= \frac{\delta^2}{2\tau} V(\tau) \\
&= \frac{\delta^2}{2} \Lambda \left(\kappa_-^2 + (1 - \kappa_-^2) \frac{(1 - \mathrm{e}^{-\tau\gamma_-})}{\tau\gamma_-} \right),
\end{aligned}$$

where

$$\Lambda = \frac{\lambda_\infty}{1 - \alpha/\beta}, \quad \kappa_- = \frac{1}{1 - \alpha/\beta}, \quad \textit{and} \quad \gamma_- = \beta - \alpha.$$

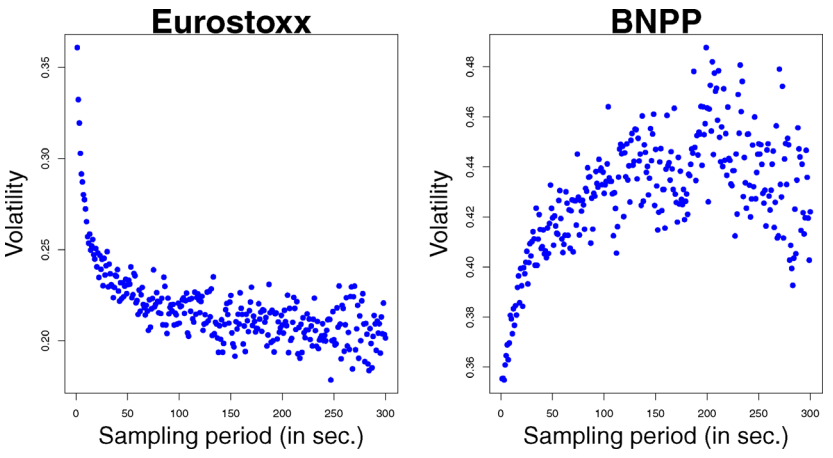

FIGURE 7
Signature Plot for Eurostoxx futures and BNPP stock on April 01, 2011, computed on mid prices to eliminate bid–ask bounce.

We eventually refer to the signature plot instead of the mean signature plot. Notice that when τ becomes larger the above expression converges to the asymptotic diffusive variance of the model calculated in (35). The mean signature plot is an increasing function with respect to τ (or equivalently the signature plot is decreasing with the sampling frequency) and this is due to the positive serial autocorrelation of the returns. This captures situations as the one observed in Figure 7.

Lastly, within our simple toy model we can determine the autocorrelation function of the price increments computed over intervals of size τ and lagged by δ; it is given by:

$$CorrStock(\tau, \delta) = \frac{\mathbb{E}[(S_{t+\tau} - S_t)(S_{t+2\tau+\delta} - S_{t+\tau+\delta})]}{\sqrt{\mathbb{E}[(S_{t+\tau} - S_t)]\mathbb{E}[(S_{t+2\tau+\delta} - S_{t+\tau+\delta})]}}, \tag{36}$$

which thanks to the closed form formula obtained in the analytical section leads to:

$$\frac{e^{-2\beta\tau+\delta(\alpha-\beta)}(e^{\tau\beta} - e^{\tau\alpha})^2\alpha(2\beta - \alpha)}{2\beta^2(\beta - \alpha)}. \tag{37}$$

The stability condition (i.e., $\alpha < \beta$) ensures the positiveness of this serial autocorrelation and it confirms the intuition developed above. Moreover, the decay of the autocorrelation as a function of the lag depends on the parameter α, which tends to reduce the decaying behavior of this function.

In order to compare our results to Bacry et al. (2013a) let us briefly recapitulate their main findings. In Bacry et al. (2013a), the authors propose a model similar to (31) but with Hawkes processes that are *mutually excited* and not *self-excited* as in our case. To be more precise the dynamics for the intensities are given by:

$$\lambda_t^{\text{up}} = \lambda_\infty + \int_0^t \alpha\, e^{-\beta(t-s)}\, dN_s^{\text{down}}, \tag{38}$$

$$\lambda_t^{\text{down}} = \lambda_\infty + \int_0^t \alpha\, e^{-\beta(t-s)}\, dN_s^{\text{up}}. \tag{39}$$

Notice that an up jump increases the down intensity that increases the probability of a down jump and if this one occurs it will increase the up jump intensity. The process is purely mutually excited and possesses a mean reversion behavior. These dynamics are significantly different from those driving the toy model. The diffusive limit for this model is:

$$\sigma^2_{\text{BDHM}} = \frac{\delta^2}{2} \frac{\lambda_\infty}{\left(1 - \frac{\alpha}{\beta}\right)\left(1 + \frac{\alpha}{\beta}\right)^2} \tag{40}$$

and the mean signature plot is:

$$C(\tau) = \frac{\delta^2}{2} \Lambda \left(\kappa_+^2 + (1 - \kappa_+^2) \frac{1 - \mathrm{e}^{-\tau\gamma_+}}{\tau\gamma_+} \right)$$

with:

$$\Lambda = \frac{\lambda_\infty}{1 - \alpha/\beta}, \quad \kappa_+ = \frac{1}{1 + \alpha/\beta}, \quad \text{and} \quad \gamma_+ = \alpha + \beta.$$

In Bacry et al. (2013a)'s model an upward chock will increase the down intensity and trigger a downward chock on the mid price, thereby leading to a mean reverting behavior for the mid price. Therefore, as a function of the sampling period the signature plot is decreasing with respect to τ (or equivalently the signature plot is increasing with the sampling frequency) because of this negative serial autocorrelation of the returns. In Andersen et al. (1999) the authors mention "the patterns of bias injected in realized volatility as underlying returns are sampled progressively more frequently." This bias can lead to a decreasing or increasing volatility as a function of the sampling period. The decreasing pattern is the most frequent but inverted pattern is also possible and can be found in data, even for liquid stocks, as illustrated in Figure 7. Bacry et al. model is compatible with a decreasing pattern, whereas the toy model is compatible with an increasing pattern.[10] Due to the positive (negative) autocorrelation of the returns in the toy (Bacry et al.) model we have, for a given pair (α, β), the inequality $\sigma > \sigma_{\text{BDHM}}$.

Within the specification (38) and (39) the autocorrelation function (36) can be computed and is given by:

$$-\frac{\mathrm{e}^{-(\delta+2\tau)(\beta+\alpha)}(\mathrm{e}^{\tau(\beta+\alpha)} - 1)^2 \alpha(2\beta + \alpha)}{2\beta^2(\beta + \alpha)}. \tag{41}$$

This quantity is negative and this confirms the intuition. Also, the decay of the autocorrelation as a function of the lag depends on parameter α, now controling the mutual excitation property of the process, and this parameter increases the decaying behavior of the function.

Notice that both models produce different asymptotic volatility formulas. In order to assess their plausibility, we calibrate a Hawkes process to the mid price up-jumps and calculate the asymptotic volatilities for the two models; the toy model given by (35) and Bacry et al. (2013a)'s specification (40). For the toy model the parameter estimation is performed as in the previous section. We use only the autocorrelation function with $\tau = 60$ seconds and δ

[10]Let us stress that these figures are based on mid price sampling so that any noise due to bid–ask bounce is eliminated. Also, whether the signature plot is an increasing or decreasing function of the sampling frequency depends on the market conditions. Some liquid stocks (or indexes) can display both; BNPP is an example, and understanding the determinants of this functional dependency remains an open question.

TABLE V
Asymptotic Volatility Values

Symbol	λ_∞	α	β	σ_{BDHM} (%)	*Empirical* σ (%)	*Toy model* σ (%)
Eurostoxx	0.0184	0.0160	0.0219	12.79	40.77	80.64
Dax	0.0429	0.0226	0.0259	4.80	25.75	71.36
BNPP	0.0379	0.0481	0.0569	5.36	41.86	72.24
Sanofi	0.0279	0.0488	0.0587	4.00	30.30	45.37
Bund	0.0267	0.0180	0.0261	2.70	8.25	14.47
Bobl	0.0228	0.0187	0.0288	2.45	6.92	11.19
Schatz	0.0257	0.0223	0.0372	1.38	2.63	5.39
JPY	0.0313	0.0659	0.0764	1.98	9.89	29.88
EURO	0.0474	0.0648	0.0725	4.87	16.76	112.25
GOLD	0.0728	0.0775	0.0868	3.77	24.48	71.64
Crude Oil Brent	0.0474	0.0472	0.0528	6.22	41.25	126.29
Natural GAS	0.0548	0.0931	0.1090	11.24	58.65	150.30
Sugar	0.0410	0.0556	0.0758	10.58	51.37	73.30
CORN	0.0419	0.0552	0.0694	9.34	43.73	75.94
WHEAT	0.0451	0.0626	0.0763	10.59	57.17	94.53

Note. Values for the asymptotic volatilities as given by our toy model (35) and the Bacry et al. (2013a) model (40) as well as the realized volatility of the day. We also put median estimated values for the Hawkes processes in the toy model (31), (32) and (33). Both models give plausible values. σ_{BDHM} volatility underestimates systematically the realized volatility, whereas our toy model systematically overestimates it.

ranging from 0 to 600 seconds by step of 60 seconds and λ_∞ is deduced thanks to (20) with $\tau = 60$ seconds (it corresponds to the "Fast calibration II"). For Bacry et al. (2013a)'s model we use the MLE algorithm.[11]

Results are reported in Table V (apart from the asymptotic volatilities we only report the estimated parameters for the toy model). We rescaled the obtained volatilities by the spot value in order to obtain the more usual Black–Scholes volatility corresponding to a lognormal model. One can clearly see that our model systematically overestimates volatility whereas Bacry et al. (2013a)'s model systematically underestimates it. Qualitatively, this is because our model magnifies up and down movements, thus inflating the realized volatility, whereas Bacry et al. (2013a)'s model moderates up and down movements thanks to its mean reversion mechanism, thus underestimating the volatility.

Our results suggest a more general specification allowing for both self- and mutual-excitations. The particularly simple approach we adopted in our calculations based on the infinitesimal generator and Dynkin's formula may be generalized to a multidimensional setting and may lead to tractable results. It would allow for both effects, which were underlined in the empirical literature (see Hautsch, 2012), on the price dynamic and provide a deeper understanding of how the volatility measured at a macroscopic level (at a daily or low frequency) depends on the trading activity observed at a microscopic level (at high frequency).

4. CONCLUSION

In this study, we explicitly compute the moments and the autocorrelation function of the number of jumps over an interval for the Hawkes process. Using these quantities we develop a method of moments estimation strategy that is extremely fast compared with the usual maximum likelihood estimation strategy. This aspect is essential as we are interested in the trade clustering activity observed in high-frequency data or if we wish to apply the model in

[11] For this model we only report the volatility value σ_{BDHM}, the parameter values are available upon request.

real time. We use our estimation framework to calibrate the Hawkes process on trades for four stocks over a 2-year sample. The Hawkes process can cope with the trade clustering effect thanks to its autocorrelation structure. As our calibration is fast, we roll the daily estimation over 2 years to analyze the parameters stability, and they are found to be reasonably stable. We perform a robustness check on other assets and obtain similar results.

Thanks to the analytical tractability of the Hawkes process we explicitly compute the impulse response associated with the process, which determines the market impact of a trade. We use this as a measure of liquidity and the estimated parameters lead to reasonable conclusions.

Lastly, within a simple model based on the Hawkes process we explicitly compute the diffusive limit for the price process. This allows us to connect the microscopic dynamic, that is to say the high-frequency dynamic, to the macroscopic dynamic, the volatility computed at a daily frequency (with the Black–Scholes volatility being the most well-known quantity).

Our work points toward several extensions. First, we computed the diffusive limit under the restrictive hypothesis that the Hawkes processes are only self-excited whereas in Bacry et al. (2013a), on which we heavily rely, the Hawkes processes are only mutually excited. The reality should lie between the two and requires Hawkes processes that are both self- and mutually excited. To this end we would need to perform the computations in the multidimensional case. We suspect that they can be performed for that case.

Another aspect of interest is the diffusive limit concept. In this work we connect the dynamic driving the trade process, using a Hawkes process, to the daily volatility. It would be of interest to go further at the microscopic level by modeling, for example, the level I quotes. The Hawkes process provides a natural modeling framework and would extend the interesting existing models based on the Poisson process. To compute the diffusive limit for a model based on the Hawkes process the moments as well as the autocorrelation are needed and they can be obtained using the computation strategy developed in this work. These interesting problems are left for future investigation.

APPENDIX A

COMPUTING THE FIRST MOMENT OF N_T BY DERIVATION OF THE MOMENT-GENERATING FUNCTION

Choosing $u_1 = 0$ in the system of Equations (11)–(13) yields the moment-generating function for N_t, which writes if we take $u_2 = u$:

$$\mathbb{E}_t^x[e^{uN_T}] = e^{a(t)+b(t)\lambda_t+uN_t}.$$

Expressed in terms of $\tau = T - t$ the expectation and the system of ordinary differential equations are given by:

$$\mathbb{E}_t^x[e^{u(N_{t+\tau}-N_t)}] = e^{a(\tau)+\lambda_t b(\tau)} \tag{A.1}$$

with:

$$\frac{\partial a}{\partial \tau} = \beta\lambda_\infty b(\tau), \tag{A.2}$$

$$\frac{\partial b}{\partial \tau} = -\beta b(\tau) - 1 + e^{\alpha b(\tau)+u} \tag{A.3}$$

and $a(0) = b(0) = 0$.

The above system of ODE fully characterizes the moment-generating function and the Laplace transform of the process, which completely determines its distribution. However, an explicit solution for Equation (A.3) is most often not available. From the moment-generating function (A.1) we can retrieve the moments of the process *after* differentiating with respect to u and evaluating for $u = 0$ the function. This leads to differentiate with respect to u the above system. For instance, for the expected number of jumps this yields to a system of ODEs for $b_u = \frac{db}{du}$ and $a_u = \frac{\mathrm{d}a}{\mathrm{d}u}$ of the form:

$$\frac{\partial a_u}{\partial \tau} = \beta\lambda_\infty b_u(\tau), \tag{A.4}$$

$$\frac{\partial b_u}{\partial \tau} = -\beta b_u(\tau) - 1 + \alpha b_u(\tau)\,\mathrm{e}^{\alpha b(\tau)+u} \tag{A.5}$$

with $a_u(0) = b_u(0) = 0$ that can be explicitly solved in conjunction with Equations (A.2) and (A.3) when $u = 0$. Therefore, we can compute the first moment. Iterating this procedure for higher moments leads to tedious computations, and whenever the moments of $\lambda_t N_t$ are needed, the computations get out of hand. Needless to say that difficulties significantly increase when the dimension of the process increases.

Proof of Lemma 1

We apply Dynkin's formula (8) to $f \equiv N_t$ and taking into account the fact that:

$$\mathcal{L}f(X_t) = \lambda_t$$

we obtain:

$$\mathbb{E}[N_t] = N_0 + \mathbb{E}\left[\int_0^t \lambda_s \,\mathrm{d}s\right].$$

Using Fubini–Tonnelli's theorem we have:

$$\mathbb{E}[N_t] = N_0 + \int_0^t \mathbb{E}[\lambda_s]\,\mathrm{d}s. \tag{A.6}$$

Differentiating this integral equation gives (15). This equation could have been obtained by recalling that $N_t - \int_0^t \lambda_s \,\mathrm{d}s$ is a martingale, by definition of the intensity of a point process, as explained in Brémaud (1981). We nevertheless quote the Dynkin formula method as the same reasoning will prove useful for other functions as well.

To obtain the ODE (16) we rely again on Dynkin's formula. Following Errais et al. (2010), let $f \equiv \lambda_t$ in (7) then as we have:

$$\mathcal{L}f(X_t) = \beta(\lambda_\infty - \lambda_t) + \alpha\lambda_t.$$

Dynkin's formula leads to:

$$\begin{aligned}\mathbb{E}[\lambda_t] &= \lambda_0 + \mathbb{E}\left[\int_0^t (\beta(\lambda_\infty - \lambda_s) + \alpha\lambda_s)\,\mathrm{d}s\right] \\ &= \lambda_0 + \beta\lambda_\infty t + (\alpha - \beta)\int_0^t \mathbb{E}[\lambda_s]\,\mathrm{d}s,\end{aligned}$$

where as before, we used Fubini–Tonnelli's theorem to swap the integration and expectation operators. Taking the differential with respect to t yields the ordinary differential equation satisfied by the expected intensity (16). ■

Proof of Lemma 2

For function $f \equiv N^2$ Equation (7) gives:

$$\mathcal{L}f(X_t) = 2\lambda_t N_t + \lambda_t$$

and Dynkin's formula (8) results in:

$$\mathbb{E}[N_t^2] = N_0^2 + 2\int_0^t \mathbb{E}[\lambda_u N_u]\,\mathrm{d}u + \int_0^t \mathbb{E}[\lambda_u]\,\mathrm{d}u \tag{A.7}$$

and differentiating this equation with respect to t leads to (17). ■

Following the same procedure for $f \equiv \lambda N$ and $f \equiv \lambda^2$ give the ODE (18) and (19), respectively.

Proof of Proposition 1

Taking into account the initial condition $\mathbb{E}[\lambda_0] = \lambda_0$ the solution is found to be:

$$\mathbb{E}[\lambda_t] = \lambda_\infty \beta \frac{\mathrm{e}^{(\alpha-\beta)t} - 1}{\alpha - \beta} + \mathrm{e}^{(\alpha-\beta)t}\lambda_0. \tag{A.8}$$

From the above equation a stability condition is given by $\frac{\alpha}{\beta} < 1$. Using the above result in Equation (A.6) yields the expression for the mean number of jumps:

$$\mathbb{E}[N_t] = N_0 + \frac{\lambda_\infty \beta(-1 + \mathrm{e}^{(\alpha-\beta)t} - (\alpha-\beta)t)}{(\alpha-\beta)^2} + \frac{(-1 + \mathrm{e}^{(\alpha-\beta)t})}{\alpha-\beta}\lambda_0.$$

We are interested in the expected number of jumps during an interval of length τ. Using the previous computation we conclude that it is given by:

$$\begin{aligned}&\mathbb{E}[N_{t+\tau} - N_t] \\ &\quad = \frac{-\lambda_\infty \beta \tau}{\alpha-\beta} + \mathrm{e}^{t(\alpha-\beta)}\frac{(-\lambda_\infty\beta + \mathrm{e}^{(\alpha-\beta)\tau}\lambda_\infty\beta - \alpha\lambda_0 + \mathrm{e}^{(\alpha-\beta)\tau}\alpha\lambda_0 + \beta\lambda_0 - \mathrm{e}^{(\alpha-\beta)\tau}\beta\lambda_0)}{(\alpha-\beta)^2}.\end{aligned} \tag{A.9}$$

Equation (A.9) depends on λ_0 the initial value for the intensity, which is unknown. To eliminate this value we take the limit $t \to \infty$, and under the stability condition $\frac{\alpha}{\beta} < 1$ we obtain (20).

We need to compute the second moment of the number of jumps during a given interval, namely:

$$I = \mathbb{E}[(N_{t_2} - N_{t_1})^2] = \mathbb{E}[\mathbb{E}_{t_1}[N_{t_2}^2] - 2N_{t_1}\mathbb{E}_{t_1}[N_{t_2}] + N_{t_1}^2]. \tag{A.10}$$

Using the ODE (17) it results that:

$$\mathbb{E}_{t_1}[N_{t_2}^2] = N_{t_1}^2 + 2\int_{t_1}^{t_2} \mathbb{E}_{t_1}[\lambda_u N_u]\,\mathrm{d}u + \int_{t_1}^{t_2} \mathbb{E}_{t_1}[\lambda_u]\,\mathrm{d}u \tag{A.11}$$

and when inserted in the previous equation leads to:

$$I = 2\int_{t_1}^{t_2} \mathbb{E}[\lambda_u N_u]\,\mathrm{d}u + \int_{t_1}^{t_2} \mathbb{E}[\lambda_u]\,\mathrm{d}u - 2\mathbb{E}\left[N_{t_1}\int_{t_1}^{t_2} \mathbb{E}_{t_1}[\lambda_u]\,\mathrm{d}u\right]. \tag{A.12}$$

The first integral of (A.12) can be computed thanks to (18) and it gives:

$$\begin{aligned} I_1 &= \int_{t_1}^{t_2} \mathbb{E}[\lambda_u N_u]\,\mathrm{d}u \\ &= \int_{t_1}^{t_2} \mathrm{e}^{(\alpha-\beta)(u-t_1)}\mathbb{E}[\lambda_{t_1} N_{t_1}]\,\mathrm{d}u + \int_{t_1}^{t_2}\int_{t_1}^{u} \mathrm{e}^{(\alpha-\beta)(u-s)}\{\beta\lambda_\infty \mathbb{E}[N_s] + \mathbb{E}[\lambda_s^2] + \alpha\mathbb{E}[\lambda_s]\}\,\mathrm{d}s\,\mathrm{d}u, \end{aligned}$$

whereas for the third term of (A.12) is:

$$\begin{aligned} I_2 &= \mathbb{E}[N_{t_1}\int_{t_1}^{t_2} \mathbb{E}_{t_1}[\lambda_u]\,\mathrm{d}u] \\ &= \mathbb{E}\left[N_{t_1}\left(\int_{t_1}^{t_2} \mathrm{e}^{(\alpha-\beta)(u-t_1)}\lambda_{t_1}\,\mathrm{d}u + \int_{t_1}^{t_2}\int_{t_1}^{u} \mathrm{e}^{(\alpha-\beta)(u-r)}\beta\lambda_\infty\,\mathrm{d}r\,\mathrm{d}u\right)\right] \\ &= \int_{t_1}^{t_2} \mathrm{e}^{(\alpha-\beta)(u-t_1)}\,\mathrm{d}u\mathbb{E}[N_{t_1}\lambda_{t_1}] + \int_{t_1}^{t_2}\int_{t_1}^{u} \mathrm{e}^{(\alpha-\beta)(u-r)}\,\mathrm{d}r\,\mathrm{d}u\beta\lambda_\infty\mathbb{E}[N_{t_1}]. \end{aligned}$$

As we have $\mathbb{E}[N_s] = \mathbb{E}[N_{t_1}] + \int_{t_1}^{s} \mathbb{E}[\lambda_r]\,\mathrm{d}r$ we arrive after substitution and simplification to:

$$I = \int_{t_1}^{t_2} \mathbb{E}[\lambda_u]\,\mathrm{d}u + 2\int_{t_1}^{t_2}\int_{t_1}^{u} \mathrm{e}^{(\alpha-\beta)(u-s)}\left\{\beta\lambda_\infty\int_{t_1}^{s} \mathbb{E}[\lambda_r]\,\mathrm{d}r + \mathbb{E}[\lambda_s^2] + \alpha\mathbb{E}[\lambda_s]\right\}\mathrm{d}s\,\mathrm{d}u.$$

We can therefore calculate the time t expected second moment of the number of jumps occurring during an interval of length τ, by first conditioning on $\mathcal{F}_t$, obtaining $\mathbb{E}_t[(N_{t+\tau} - N_t)^2]$, which is an expression depending only on the expectations $\mathbb{E}[\lambda_t]$ and $\mathbb{E}[\lambda_t^2]$. These last two terms depend on λ_0 but by letting $t \to \infty$ we obtain its stationary regime value and get an expression independent of the initial intensity. As a result we have the second moment of the number of jumps over a time interval of length τ.

To specify further the result, we note $\lim_{t\to\infty} \mathbb{E}[\lambda_t] = \Lambda$, $\lim_{t\to\infty} \mathbb{E}[\lambda_t^2] = \Lambda_2$ and taking the limit on the expression for I we reach:

$$\begin{aligned} \lim_{t\to+\infty} \mathbb{E}[(N_{t+\tau} - N_t)^2] = \lim_{t\to+\infty} \tau\Lambda + 2\int_t^{t+\tau}\int_t^u \mathrm{e}^{(\alpha-\beta)(u-s)}\int_t^s \mathrm{d}r\,\mathrm{d}s\,\mathrm{d}u\beta\lambda_\infty\Lambda \\ + 2\int_t^{t+\tau}\int^u \mathrm{e}^{(\alpha-\beta)(u-s)}\,\mathrm{d}s\,\mathrm{d}u\{\Lambda_2 + \alpha\Lambda\}, \end{aligned}$$

where the integral expressions are given by:

$$\int_t^{t+\tau}\int_t^u \mathrm{e}^{(\alpha-\beta)(u-s)}\int_t^s \mathrm{d}r\,\mathrm{d}s\,\mathrm{d}u = -(\alpha-\beta)^{-1}\frac{\tau^2}{2} - (\alpha-\beta)^{-2}\tau + (\alpha-\beta)^{-3}(\mathrm{e}^{(\alpha-\beta)\tau}-1)$$

$$\int_t^{t+\tau}\int_t^u \mathrm{e}^{(\alpha-\beta)(u-s)}\,\mathrm{d}s\,\mathrm{d}u = -(\alpha-\beta)^{-1}\tau + (\alpha-\beta)^{-2}(\mathrm{e}^{(\alpha-\beta)\tau}-1).$$

Equation (21) is deduced using the above equation and (20).

To compute the autocovariance function of the number of jumps during different time intervals we need to determine $\mathbb{E}_t[(N_{t_1}-N_t)(N_{t_3}-N_{t_2})]$, where $t<t_1<t_2<t_3$. In order to simplify notations we consider the variables $\Delta_1 = t_1 - t$, $\Delta_2 = t_3 - t_2$, and $\delta = t_2 - t_1$. By performing successive conditionings we get:

$$\mathbb{E}[(N_{t_1}-N_t)(N_{t_3}-N_{t_2})] = \mathbb{E}[\mathbb{E}_t[\mathbb{E}_{t_1}[\mathbb{E}_{t_2}[(N_{t_1}-N_t)(N_{t_3}-N_{t_2})]]]].$$

The innermost conditional expectation is:

$$\mathbb{E}_{t_2}[(N_{t_1}-N_t)(N_{t_3}-N_{t_2})] = (N_{t_1}-N_t) \times\left[\frac{\lambda_\infty\beta(-1+\mathrm{e}^{(\alpha-\beta)\Delta_2}-(\alpha-\beta)\Delta_2)}{(\alpha-\beta)^2} + \frac{(-1+\mathrm{e}^{(\alpha-\beta)\Delta_2})}{\alpha-\beta}\lambda_{t_2}\right],$$

thanks to calculations done for the first moment. Then, conditioning down by $\mathcal{F}_{t_1}$, one has to compute:

$$\mathbb{E}_{t_1}[\lambda_{t_2}] = \lambda_\infty\beta\frac{\mathrm{e}^{(\alpha-\beta)\delta}-1}{\alpha-\beta} + \mathrm{e}^{(\alpha-\beta)\delta}\lambda_{t_1}. \tag{A.13}$$

This results in an expression depending on $N_{t_1}\lambda_{t_1}$ and $N_t\lambda_{t_1}$. Further conditioning down with respect to $\mathcal{F}_t$, one has to calculate :

$$\mathbb{E}_t[\lambda_{t_1}] = \lambda_\infty\beta\frac{\mathrm{e}^{(\alpha-\beta)\Delta_1}-1}{\alpha-\beta} + \mathrm{e}^{(\alpha-\beta)\Delta_1}\lambda_t.$$

Lastly, the quantity $\mathbb{E}_t[\lambda_{t_1}N_{t_1}]$ is already known from the previous computations. Collecting all results together we determine the autocovariance of the process and to simplify the final expression we suppose $\Delta_1 = \Delta_2 = \tau$, so that we obtain:

$$\lim_{t\to\infty}\mathbb{E}[(N_{t+\tau}-N_t)(N_{t+2\tau+\delta}-N_{t+\tau+\delta})] = \frac{\lambda_\infty\beta\alpha(2\beta-\alpha)(\mathrm{e}^{(\alpha-\beta)\tau}-1)^2}{2(\alpha-\beta)^4}\mathrm{e}^{(\alpha-\beta)\delta} + \frac{\lambda_\infty^2\beta^2}{(\alpha-\beta)^2}\tau^2. \tag{A.14}$$

If we subtract the mean value, we obtain the expression (22). ■

EXPRESSION FOR THE SKEWNESS

By application of the infinitesimal generator operator to adequate functions, one has the following ordinary differential equations:

$$
\begin{aligned}
d\mathbb{E}[N^3] &= \mathbb{E}[\lambda_t]\,\mathrm{d}t + 3\mathbb{E}[\lambda_t N_t]\,\mathrm{d}t + 3\mathbb{E}[\lambda_t N_t^2]\,\mathrm{d}t,\\
d\mathbb{E}[\lambda_t N_t^2] &= \mathbb{E}[\lambda_t^2]\,\mathrm{d}t + 2\mathbb{E}[\lambda_t^2 N_t]\,\mathrm{d}t + \alpha\mathbb{E}[\lambda_t]\,\mathrm{d}t + 2\alpha\mathbb{E}[\lambda_t N_t]\,\mathrm{d}t + (\alpha-\beta)\mathbb{E}[\lambda_t N_t^2]\,\mathrm{d}t\\
&\quad + \beta\lambda_\infty\mathbb{E}[N_t^2]\,\mathrm{d}t,\\
d\mathbb{E}[\lambda_t^2 N_t] &= \mathbb{E}[\lambda_t^3]\,\mathrm{d}t + 2\alpha\mathbb{E}[\lambda_t^2]\,\mathrm{d}t + 2(\alpha-\beta)\mathbb{E}[\lambda_t^2 N_t]\,\mathrm{d}t + \alpha^2\mathbb{E}[\lambda_t]\,\mathrm{d}t\\
&\quad + (\alpha^2 + 2\lambda_\infty\beta)\mathbb{E}[\lambda_t N_t]\,\mathrm{d}t,\\
d\mathbb{E}[\lambda_t^3] &= 3(\alpha-\beta)\mathbb{E}[\lambda_t^3]\,\mathrm{d}t + 3(\alpha^2+\lambda_\infty\beta)\mathbb{E}[\lambda_t^2]\,\mathrm{d}t + \alpha^3\mathbb{E}[\lambda_t]\,\mathrm{d}t.
\end{aligned}
$$

The stationary regime third moment then writes:

$$
\begin{aligned}
\lim_{t\to\infty}\mathbb{E}[(N_{t+\tau}-N_t)^3] &= \frac{1}{2(\alpha-\beta)^6}\lambda_\infty\beta\\
&\quad\times[-\mathrm{e}^{2(\alpha-\beta)\tau}\alpha^2(2\alpha-3\beta)(\alpha-\beta)\\
&\quad+2\mathrm{e}^{(\alpha-\beta)\tau}\alpha(\alpha^3-4\alpha^2\beta+3\alpha\beta^2+6\beta^3+3(\lambda_\infty+\alpha)(\alpha-2\beta)(\alpha-\beta)\beta\tau)\\
&\quad+\beta(3\alpha(\alpha^2-\alpha\beta-4\beta^2)\\
&\quad+2(-\alpha+\beta)(3\lambda_\infty\alpha(\alpha-2\beta)+\beta^2(2\alpha+\beta))\tau\\
&\quad+6\lambda_\infty(\alpha-\beta)^2\beta^2\tau^2+2\lambda_\infty^2\beta(-\alpha+\beta)^3\tau^3)].
\end{aligned}
$$

EXPRESSION FOR THE KURTOSIS

Similarly to the preceding paragraph, we have the following ordinary differential equations:

$$
\begin{aligned}
d\mathbb{E}[N^4] &= \mathbb{E}[\lambda_t]\,\mathrm{d}t + 4\mathbb{E}[\lambda_t N_t]\,\mathrm{d}t + 6\mathbb{E}[\lambda_t N_t^2]\,\mathrm{d}t + 4\mathbb{E}[\lambda_t N_t^3]\,\mathrm{d}t,\\
d\mathbb{E}[\lambda_t N_t^3] &= \mathbb{E}[\lambda_t^2]\,\mathrm{d}t + 3\mathbb{E}[\lambda_t^2 N_t]\,\mathrm{d}t + 3\mathbb{E}[\lambda_t^2 N_t^2]\,\mathrm{d}t + \alpha\mathbb{E}[\lambda_t]\,\mathrm{d}t + 3\alpha\mathbb{E}[\lambda_t N_t]\,\mathrm{d}t\\
&\quad + 3\alpha\mathbb{E}[\lambda_t N_t^2]\,\mathrm{d}t + (\alpha-\beta)\mathbb{E}[\lambda_t N_t^3]\,\mathrm{d}t + \lambda_\infty\beta\mathbb{E}[N_t^3]\,\mathrm{d}t,\\
d\mathbb{E}[\lambda_t^2 N_t^2] &= \mathbb{E}[\lambda_t^3]\,\mathrm{d}t + 2\mathbb{E}[\lambda_t^3 N_t]\,\mathrm{d}t + 2\alpha\mathbb{E}[\lambda_t^2] + 4\alpha\mathbb{E}[\lambda_t^2 N_t]\,\mathrm{d}t + 2\alpha\mathbb{E}[\lambda_t^2 N_t^2]\,\mathrm{d}t + \alpha^2\mathbb{E}[\lambda_t]\,\mathrm{d}t\\
&\quad + 2\alpha^2\mathbb{E}[\lambda_t N_t]\,\mathrm{d}t + (\alpha^2+2\lambda_\infty\beta)\mathbb{E}\lambda_t N_t^2]\,\mathrm{d}t - 2\beta\mathbb{E}\lambda_t^2 N_t^2]\,\mathrm{d}t,\\
d\mathbb{E}[\lambda_t^3 N_t] &= \mathbb{E}[\lambda_t^4]\,\mathrm{d}t + 3\alpha\mathbb{E}[\lambda_t^3]\,\mathrm{d}t + 3(\alpha-\beta)\mathbb{E}[\lambda_t^3 N_t]\,\mathrm{d}t + 3\alpha^2\mathbb{E}[\lambda_t^2]\,\mathrm{d}t\\
&\quad + 3(\alpha^2+\lambda_\infty\beta)\mathbb{E}[\lambda_t^2 N_t]\,\mathrm{d}t + \alpha^3\mathbb{E}[\lambda_t]\,\mathrm{d}t + \alpha^3\mathbb{E}[\lambda_t N_t]\,\mathrm{d}t,\\
d\mathbb{E}[\lambda_t^4] &= (4\alpha-4\beta)\mathbb{E}[\lambda_t^4]\,\mathrm{d}t + (6\alpha^2+4\lambda_\infty\beta)\mathbb{E}[\lambda_t^3]\,\mathrm{d}t + 4\alpha^3\mathbb{E}[\lambda_t^2]\,\mathrm{d}t + \alpha^4\mathbb{E}[\lambda_t]\,\mathrm{d}t.
\end{aligned}
$$

The stationary regime fourth moment of the process writes:

$$\begin{aligned}
\lim_{t\to\infty} \mathbb{E}[(N_{t+\tau} - N_t)^4] = \frac{1}{6(\alpha-\beta)^8} \\
&\times \lambda_\infty \beta[-2e^{3(\alpha-\beta)\tau}\alpha^3(3\alpha-4\beta)(\alpha-\beta)(2\alpha-\beta) \\
&+ 3e^{2(\alpha-\beta)\tau}\alpha^2(6\alpha^4 + 6(\lambda_\infty - 3\alpha)\alpha^2\beta + 3\alpha(-8\lambda_\infty + \alpha)\beta^2 \\
&+ 6(4\lambda_\infty + 5\alpha)\beta^3 - 18\beta^4 + 4(\lambda_\infty + 2\alpha)(2\alpha - 3\beta)(\alpha-\beta)^2\beta\tau) \\
&- 6e^{(\alpha-\beta)\tau}\alpha(\alpha^5 + 6\lambda_\infty\alpha^3\beta - 3\alpha^4\beta - 24\lambda_\infty\alpha^2\beta^2 - \alpha^3\beta^2 \\
&+ 24\lambda_\infty\alpha\beta^3 + 20\alpha^2\beta^3 - 45\alpha\beta^4 - 14\beta^5 \\
&+ 2(\lambda_\infty + \alpha)(\alpha-\beta)\beta(2\alpha^3 - 8\alpha^2\beta + 3\alpha\beta^2 + 18\beta^3)\tau \\
&+ 6(\lambda_\infty + \alpha)^2(\alpha - 2\beta)(\alpha-\beta)^2\beta^2\tau^2) \\
&+ \beta(\alpha(2\alpha^4 + 18\lambda_\infty\alpha(\alpha-2\beta)^2 + 15\alpha^3\beta + 22\alpha^2\beta^2 - 216\alpha\beta^3 - 84\beta^4) \\
&+ 6\beta(-\alpha+\beta)(6\lambda_\infty\alpha^3 + 6\alpha(-6\lambda_\infty + \alpha)\beta^2 + 8\alpha\beta^3 + \beta^4)\tau \\
&+ 6\lambda_\infty(\alpha-\beta)^2\beta(6\lambda_\infty\alpha(\alpha-2\beta) + \beta^2(8\alpha+7\beta))\tau^2 \\
&+ 36\lambda_\infty^2\beta^3(-\alpha+\beta)^3\tau^3 + 6\lambda_\infty^3(\alpha-\beta)^4\beta^2\tau^4)].
\end{aligned}$$

REFERENCES

Abergel, F., & Jedidi, A. (2013). A mathematical approach to order book modeling. International Journal of Theoretical and Applied Finance, 16.

Aït-Sahalia, Y., Cacho-Diaz, J., & Laeven, R. J. A. (2010). Modeling financial contagion using mutually exciting jump processes. Working paper, National Bureau of Economic Research.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (October 1999). (Understanding, optimizing, using and forecasting) realized volatility and correlation. Working Paper Seires 99-061, New York University, Leonard N. Stern School of Business.

Bacry, E., Delattre, S., Hoffmann, M., & Muzy, J.-F.. (January 2013a). Modelling microstructure noise with mutually exciting point processes. Quantitative Finance, 13, 65–77.

Bacry, E., Delattre, S., Hoffmann, M., & Muzy, J. F. (2013b). Scaling limits for hawkes processes and application to financial statistics. Stochastic Processes and Applications, 123, 2475–2499.

Billingsley, P. (1999). Convergence of probability measures. Wiley Series in probability and statistics: Probability and statistics. New York: Wiley.

Bollerslev, T., & Zhou, H.(2002). Estimating stochastic volatility diffusion using conditional moments of integrated volatility. Journal of Econometrics, 109, 33–65.

Bollerslev, T., & Zhou, H. (2004). Corrigendum to "Estimating stochastic volatility diffusion using conditional moments of integrated volatility." Journal of Econometrics, 119, 221–222.

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

Brémaud, P. (1981). Point processes and queues, martingale dynamics. Berlin-Heidelberg-New York: Springer.

Brémaud, P., & Massoulié, L. (1994). Imbedded construction of stationary sequences and point processes with a random memory. Queueing Systems, 17, 213–234.

Cont, R., & De Larrard, A. (January 2011). Price dynamics in a Markovian limit order market. Social science research network working paper series.

Cont, R., & De Larrard, A. (February 2012). Order book dynamics in liquid markets: Limit theorems and diffusion approximations. Social science research network working paper series.

Cont, R., Stoikov, S., & Talreja, R. (2010). A stochastic model for order book dynamics. Operations Research, 58, 549–563.

Cuchiero, C., Keller-Ressel, M., & Teichmann, J. (2012). Polynomial processes and their applications to mathematical finance. Finance and Stochastics, 16, 711–740.

Daley, D. J., & Jones, D. V. (2002). An introduction to the theory of point processes, Vol. 1, Elementary theory and methods (2nd ed.). Berlin-Heidelberg-New York: Springer.

Daley, D. J., & Jones, D. V. (2008). An introduction to the theory of point processes, Vol. 2, General theory and structure (2nd ed.). Berlin-Heidelberg-New York: Springer.

Duffie, D., & Kan, R. (1996). A yield-factor model of interest rates. Mathematical Finance, 6, 379–406.

Engle, R. F., & Russell, J. R. (September 1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. Econometrica, 66, 1127–1162.

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

Filipović, D., Mayerhofer, E., & Schneider, P. (2013). Density approximations for multivariate affine jump-diffusion processes. Journal of Econometrics, 176, 93–111.

Golub, G. H., & Van Loan, C. F. (1996). Matrix computations. Johns Hopkins Studies in Mathematical Sciences. Baltimore: The Johns Hopkins University Press.

Hall, A. R. (2004). Generalized method of moments. Advanced texts in econometrics. Oxford: Oxford University Press.

Hautsch, N. (2012). Econometrics of financial high-frequency data. Berlin, Heidelberg: Springer.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. Biometrika, 58, 83–90.

Hewlett, P. (2006). Clustering of order arrivals, price impact and trade path optimisation. Working paper.

Kirilenko, A., Sowers, R. B., & Meng, X. (February 2013). A multiscale model of high-frequency trading. Algorithmic Finance, 2, 59–98.

Large, J. (February 2007). Measuring the resiliency of an electronic limit order book. Journal of Financial Markets, 10, 1–25.

Lewis, E., & Mohler, G. (2011). A nonparametric em algorithm for multiscale hawkes processes Working paper.

Lourakis, M. (2004). levmar: Levenberg–Marquardt nonlinear least squares algorithms in C/C++. URL `http://www.ics.forth.gr/~lourakis/levmar/`. Accessed on January 31, 2005.

Meddahi, N. (2003). Arma representation of integrated and realized variances. Econometrics Journal, 6, 335–356.

Meyn, S., & Tweedie, R. L. (2009). Markov chains and stochastic stability (2nd ed.). Cambridge: Cambridge University Press.

Meyn, S. P., & Tweedie, R. L. (1993). Stability of Markovian processes III: Foster–Lyapunov criteria for continuous-time processes. Advances in Applied Probability, 25, 518–548.

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E. (2011). Self-exciting point process modeling of crime. Journal of the American Statistical Association, 106, 100–108.

Muni Toke, I., & Pomponio, F. (2011). Modelling trades-through in a limited order book using hawkes processes. Economics discussion paper 2011-32, Kiel Institute for the World Economy. URL http://www.economics-ejournal.org/economics/discussionpapers/2011-32

Ogata, Y. (1981). On lewis simulation method for point processes. IEEE Transactions on Information Theory, 27, 23–31.

Ozaki, T. (1979). Maximum likelihood estimation of hawkes' self-exciting point processes. Annals of the Institute of Statistical Mathematics, 31, 145–155.

Pomponio, F., & Abergel, F. (2013). Multiple-limit trades: empirical facts and application to lead-lag measures. Quantitative Finance, 13, 783–793.

Revuz, D., & Yor, M. (1999). Continuous martingales and Brownian motion (3rd ed.). Berlin-Heidelberg-New York: Springer.

Sarkar, A., & Schwartz, R. A. (2006). Two-sided markets and intertemporal trade clustering: Insights into trading motives. Staff Reports 246, Federal Reserve Bank of New York.

Veen, A., & Schoenberg, F. P. (2008). Estimation of spacetime branching process models in seismology using an emtype algorithm. Journal of the American Statistical Association, 103, 614–624.

Vere-Jones, D. (1970). Stochastic models for earthquake occurrence. Journal of the Royal Statistical Society Series B, 32, 1–62.