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

José Da Fonseca* Riadh Zaatour†

July 12, 2013

Abstract

The aim of this paper is to provide an explicit form for the moments and the autocorrelation function of the number of jumps over a given interval for the 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 optimization algorithm that can be solved almost instantaneously is developed. The estimation strategy is applied to trade arrival times for major stocks that show a clustering behaviour, a feature the Hawkes process is shown to handle effectively. As the calibration is fast the estimation is rolled to determine the stability of the estimated parameters. A forecasting test underlining the advantages of the Hawkes process as a modelling framework is performed. Lastly, thanks to the quantities computed the determination of the diffusive limit of a simple model for the price evolution based on the Hawkes process is made explicit. It allows the determination of the connection between the parameters driving the high frequency activity to the daily volatility.

JEL-Classification: C13; C32; C58.
Keywords: Hawkes process, Calibration, High-frequency data, Trade clustering, Diffusive limit.

*Auckland University of Technology, Business School, Department of Finance, Private Bag 92006, 1142 Auckland, New Zealand. Phone: ++64 9 9219999 extn 5063. Email: jose.dafonseca@aut.ac.nz

†Corresponding author: Chair of Quantitative Finance, Ecole Centrale Paris, Grande Voie des Vignes, 92290 Châtenay-Malabry, France. Email: zaatour_riadh@yahoo.fr

Introduction

Trading activity leads to time series of irregularly spaced points that show a clustering behaviour. 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 et al. (2013) 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 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 which 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 has several advantages from an analytical point of view, that we will develop in this paper, that allow for very interesting applications.

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, as high frequency trading activity requires a fast estimation procedure. The maximization of the likelihood function, that 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, we can easily perform some forecast analysis and specify the horizon beyond which the Hawkes process does not perform better than a very simple model. It also underlines the well-known fact that complicated models that lead to high likelihood function values can perform poorly in terms of forecast. Therefore, the apparent simplicity of the Hawkes process might be, after all, an advantage.

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

Third, 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.

Fourth, we can compute the diffusive limit for the process and therefore make explicit 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. To perform such analysis the analytical tractability of the Hawkes process turns out to be essential and underlines, once again, the advantages related to the simplicity of this process. With respect to that aspect we contribute to a new trend of the literature developed by Cont et al. (2010), Cont and De Larrard (2011), Cont and De Larrard (2012), Bacry et al. (2013), Abergel and Jedidi (2013), Bacry et al. (2012), Kirilenko et al. (2013) aiming at connecting these two scales (the high frequency quantities and the daily quantities).

The structure of the paper is as follows. In the first part, we describe the basic properties of the Hawkes process as well as the Dynkin formula that will be our main mathematical tool. In the second part, we present the computation of the moments and the autocorrelation function for the number of jumps over a given time interval. In the third part, we present the usual optimization algorithms used in the literature and the method of moments based on the analytical results. In the fourth part, we present the data, various estimation results and a forecasting experiment. In this part we also present the impulse response analysis allowed by the model estimated on different data sets as well as the diffusive limit of the model. The last part concludes the paper. Some technical results, tables and figures are gathered in the appendix.

1 Univariate Hawkes Process

1.1 Dynamics and properties

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

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

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

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

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 dynamics

$$d\lambda_t = \beta\left(\lambda_\infty - \lambda_t\right)dt + \alpha dN_t. \tag{4}$$

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-excited" 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 $e^{\beta t}\lambda_t$ yields :

$$\lambda_t = e^{-\beta t}\left(\lambda_0 - \lambda_\infty\right) + \lambda_\infty + \int_0^t \alpha e^{-\beta(t-s)} dN_s. \tag{5}$$

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 large the impact of λ_0, the initial value for the intensity vanishes, leaving us with :

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

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

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

The 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 et al. (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 :

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

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

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

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

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

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

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

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

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

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 heavily rely on this formula to compute some distributional properties of Hawkes process.

1.2 Affine structure and moment-generating function

As underlined in Errais et al. (2010) the process $X_t = (\lambda_t, N_t)$ is a Markov process which is *affine*. Hence a closed-form solution for the moment-generating function is available that we now recall.

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 $f(t, X_t) = \mathbb{E}_t^x\left[e^{u^\top X_T}\right] = \mathbb{E}_t^x\left[e^{u_1\lambda_T + u_2 N_T}\right]$. Clearly, $f(t, \lambda_t, N_t)$ 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 $f(T, X_T) = e^{u^\top X_T}$. As $X_t = (\lambda_t, N_t)$ is a Markov affine point process we guess that the solution of (9) is an exponential affine form of the state variable, that is to say :

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

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

$$\begin{aligned}\frac{\partial a}{\partial t} &= -\beta\lambda_\infty b(t)\\ \frac{\partial b}{\partial t} &= \beta b(t) + 1 - e^{\alpha b(t)+u_2}\\ \frac{\partial c}{\partial t} &= 0\end{aligned} \tag{10}$$

with terminal conditions $a(T) = 0$, $b(T) = u_1$ and $c(T) = u_2$. From the last equation we deduce that $c(t) = u_2$. Choosing $u_1 = 0$ in the above equations yields the moment-generating function for N_t which writes as (we take $u_2 = u$):

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

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

$$\mathbb{E}_t^x \left[e^{u(N_{t+\tau}-N_t)}\right] = e^{a(\tau)+\lambda_t b(\tau)} \tag{11}$$

with

$$\frac{\partial a}{\partial \tau} = \beta\lambda_\infty b(\tau), \tag{12}$$

$$\frac{\partial b}{\partial \tau} = -\beta b(\tau) - 1 + e^{\alpha b(\tau)+u} \tag{13}$$

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

The above system of ODEs fully characterizes the moment-generating function and the Laplace transform of the process which completely determines its distribution. However, an explicit solution for the equation (13) is most often not available. From the moment-generating function (11) 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 ODE for $b_u = \frac{db}{du}$ and $a_u = \frac{da}{du}$ of the form:

$$\frac{\partial a_u}{\partial \tau} = \beta\lambda_\infty b_u(\tau), \tag{14}$$

$$\frac{\partial b_u}{\partial \tau} = -\beta b_u(\tau) - 1 + \alpha b_u(\tau) e^{\alpha b(\tau)+u} \tag{15}$$

with $a_u(0) = b_u(0) = 0$ that can be explicitly solved in conjunction with equations (12) and (13) when $u = 0$. Therefore, we can compute the different moments for 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 quickly become tedious. Also, whenever we wish to compute moments for the intensity process λ_t then the resolution of (10) is not possible. What is more, the computation of the autocovariance function of the number of jumps increments, that is to say

$$\mathbb{E}_t^x \left[(N_{t_4} - N_{t_3})(N_{t_2} - N_{t_1}) \right] \tag{16}$$

with $t < t_1 < t_2 < t_3 < t_4$ can be obtained from (11) by performing successive conditioning. More precisely, we need to compute the function

$$\mathbb{E}_t^x \left[e^{u_2(N_{t_4} - N_{t_3}) + u_1(N_{t_2} - N_{t_1})} \right].$$

In that case it will introduce the intensity process, which appears on the right hand side of (11), 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 becomes more complicated. As the quantity (16) is essential, because it carries the clustering property of the Hawkes process, we need to develop a simpler approach to compute these quantities.

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 Computing the Moments and the Autocovariance

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, we apply Dynkin's formula to $f \equiv N_t$ and taking into account the fact that:

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

we obtain

$$\mathbb{E}\left[N_t\right] = N_0 + \mathbb{E}\left[\int_0^t \lambda_s ds\right].$$

Thanks to Fubini-Tonnelli's theorem we have:

$$\mathbb{E}\left[N_t\right] = N_0 + \int_0^t \mathbb{E}\left[\lambda_s\right] ds. \tag{17}$$

This equation could have been obtained by recalling that $N_t - \int_0^t \lambda_s ds$ 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.

Indeed, in order to calculate the r.h.s. of the equation (17), we rely again on Dynkin's formula. Following Errais et al. (2010), let $f \equiv \lambda_t$ in (7) then we have:

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

Dynkin's formula leads to:

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

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 of the process

$$d\mathbb{E}\left[\lambda_t\right] = \left(\beta\lambda_\infty + \left(\alpha - \beta\right)\mathbb{E}\left[\lambda_t\right]\right) dt.$$

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

$$\mathbb{E}\left[\lambda_t\right] = \lambda_\infty\beta\frac{e^{(\alpha-\beta)t} - 1}{\alpha - \beta} + e^{(\alpha-\beta)t}\lambda_0. \tag{18}$$

From the above equation a stability condition is given by $\frac{\alpha}{\beta} < 1$. Using the above result in equation (17) yields the expression for the mean number of jumps:

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

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

$$\mathbb{E}\left[N_{t+\tau}-N_t\right]=\frac{-\lambda_\infty\beta\tau}{\alpha-\beta}+e^{t(\alpha-\beta)}\frac{\left(-\lambda_\infty\beta+e^{(\alpha-\beta)\tau}\lambda_\infty\beta-\alpha\lambda_0+e^{(\alpha-\beta)\tau}\alpha\lambda_0+\beta\lambda_0-e^{(\alpha-\beta)\tau}\beta\lambda_0\right)}{(\alpha-\beta)^2}. \tag{19}$$

The equation (19) 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

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

where $\Lambda=\frac{\lambda_\infty}{1-\alpha/\beta}$ stands for stationary regime expected intensity, giving the long run expected value of the number of jumps during a time interval of length τ. The strategy of taking the limit to simplify the dependency of the results on the initial value of the process, which is unknown, is borrowed from Aït-Sahalia et al. (2010) who used the Hawkes process for modeling contagion effects between stocks. It puts the process in its long-run, or, said differently, stationary-regime. Having determined the first moments we pursue our approach and focus on higher moments.

The computation of the second moment of the number of jumps during an interval, namely

$$I=\mathbb{E}\left[(N_{t_2}-N_{t_1})^2\right]=\mathbb{E}\left[\mathbb{E}_{t_1}\left[N_{t_2}^2\right]-2N_{t_1}\mathbb{E}_{t_1}\left[N_{t_2}\right]+N_{t_1}^2\right] \tag{21}$$

leads to apply Dynkin's formula to the function $f\equiv N^2$. As we have

$$\mathcal{L}f\left(X_t\right)=2\lambda_tN_t+\lambda_t,$$

it results that

$$\mathbb{E}_{t_1}[N_{t_2}^2]=N_{t_1}^2+2\int_{t_1}^{t_2}\mathbb{E}_{t_1}[\lambda_uN_u]du+\int_{t_1}^{t_2}\mathbb{E}_{t_1}[\lambda_u]du, \tag{22}$$

and when inserted in the previous equation leads to:

$$I=2\int_{t_1}^{t_2}\mathbb{E}[\lambda_uN_u]du+\int_{t_1}^{t_2}\mathbb{E}[\lambda_u]du-2\mathbb{E}\left[N_{t_1}\int_{t_1}^{t_2}\mathbb{E}_{t_1}[\lambda_u]du\right]. \tag{23}$$

The infinitesimal generator of the Hawkes process applied respectively to $f\equiv\lambda^2$ and $f\equiv\lambda N$ gives after proceeding as for the previous functions, to the following set of ordinary differential equations

$$d\mathbb{E}[\lambda_t^2]=(\alpha^2+2\beta\lambda_\infty)\mathbb{E}[\lambda_t]dt+2(\alpha-\beta)\mathbb{E}[\lambda_t^2]dt \tag{24}$$

$$d\mathbb{E}[\lambda_tN_t]=\beta\lambda_\infty\mathbb{E}[N_t]dt+(\alpha-\beta)\mathbb{E}[\lambda_tN_t]dt+\mathbb{E}[\lambda_t^2]dt+\alpha\mathbb{E}[\lambda_t]dt \tag{25}$$

$$d\mathbb{E}[N_t^2]=2\mathbb{E}[\lambda_tN_t]dt+\mathbb{E}[\lambda_t]dt, \tag{26}$$

where the last equation (26) is in fact (22). To integrate equation (25) we need to know the solutions of (24), (18) and (17). Once these ODE are integrated the computation of I is complete. Indeed, using the solution of the EDO satisfied by $\mathbb{E}[\lambda_t N_t]$ we obtain:

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

and

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

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

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

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 $\mathbb{E}_t\left[(N_{t+\tau} - N_t)^2\right]$, 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 consequence, we have the second moment of the jump number over a time interval of length τ.

To specify further the result, we note $\lim_{t\to\infty} \mathbb{E}[\lambda_t] = \Lambda$, $\lim_{t\to\infty} \mathbb{E}[\lambda_t^2] = \Lambda_2$, $\tau = t_2 - t_1$ and taking the limit on the expression for I we reach:

$$\begin{aligned} I = \tau\Lambda &+ 2 \int_{t_1}^{t_2} \int_{t_1}^{u} e^{(\alpha-\beta)(u-s)} \int_{t_1}^{s} dr ds du \beta\lambda_\infty \Lambda \\ &+ 2 \int_{t_1}^{t_2} \int_{t_1}^{u} e^{(\alpha-\beta)(u-s)} ds du \{\Lambda_2 + \alpha\Lambda\}, \end{aligned}$$

where the integral expressions are given by:

$$\begin{aligned} \int_{t_1}^{t_2} \int_{t_1}^{u} e^{(\alpha-\beta)(u-s)} \int_{t_1}^{s} dr ds du &= -(\alpha-\beta)^{-1} \frac{\tau^2}{2} - (\alpha-\beta)^{-2}\tau + (\alpha-\beta)^{-3}(e^{(\alpha-\beta)\tau} - 1), \\ \int_{t_1}^{t_2} \int_{t_1}^{u} e^{(\alpha-\beta)(u-s)} ds du &= -(\alpha-\beta)^{-1}\tau + (\alpha-\beta)^{-2}(e^{(\alpha-\beta)\tau} - 1). \end{aligned}$$

Recalling the results of the previous subsection, we are then able to obtain the long run variance of the number of jumps during a time interval of length τ:

$$\begin{aligned} \mathrm{V}(\tau) &= \lim_{t\to\infty} \mathbb{E}\left[(N_{t+\tau} - N_t)^2\right] - \mathbb{E}\left[N_{t+\tau} - N_t\right]^2 \\ &= \Lambda\left(\tau\kappa_-^2 + \left(1-\kappa_-^2\right)\frac{\left(1-e^{-\tau\gamma_-}\right)}{\gamma_-}\right), \end{aligned} \tag{27}$$

where

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

We have the first and second moments for the process $(N_{t+\tau} - N_t)_{t\geq 0}$ for a given τ and following the above procedure it is possible to compute higher order moments as well. However, we are interested by the autocovariance function for this process as it contains the information regarding the self-exciting or clustering property of the Hawkes process. We now turn our attention to the computation of this important quantity.

To compute the autocorrelation function of the number of jumps during different time intervals we need to determine $\mathbb{E}_t\left[\left(N_{t_1} - N_t\right)\left(N_{t_3} - N_{t_2}\right)\right]$, 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 $\delta = t_2 - t_1$. We perform the successive conditioning

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

The innermost conditional expectation is

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

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

$$\mathbb{E}_{t_1}\left[\lambda_{t_2}\right] = \lambda_\infty\beta\frac{e^{(\alpha-\beta)\delta} - 1}{\alpha-\beta} + e^{(\alpha-\beta)\delta}\lambda_{t_1}. \tag{28}$$

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 :

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

Lastly, the quantity $\mathbb{E}_t\left[\lambda_{t_1}N_{t_1}\right]$ is already known from the previous computations. Collecting all results together, we determine the autocovariance of the process, where, to simplify the final expression, we

suppose $\Delta_1 = \Delta_2 = \tau$:

$$\begin{aligned}\mathrm{Cov}\,(\tau,\delta) &= \lim_{t\to\infty} \mathbb{E}\left[(N_{t+\tau}-N_t)(N_{t+2\tau+\delta}-N_{t+\tau+\delta})\right] \\ &= \frac{\lambda_\infty \beta\alpha(2\beta-\alpha)\left(e^{(\alpha-\beta)\tau}-1\right)^2}{2(\alpha-\beta)^4} e^{(\alpha-\beta)\delta} + \frac{\lambda_\infty^2\beta^2}{(\alpha-\beta)^2}\tau^2. \end{aligned} \tag{29}$$

From the autocovariance we deduce the autocorrelation function of the number of jumps over an interval of length τ with a lag δ :

$$\begin{aligned}\mathrm{Acf}\,(\tau,\delta) &= \lim_{t\to\infty} \frac{\mathbb{E}[(N_{t+\tau}-N_t)(N_{t+2\tau+\delta}-N_{t+\tau+\delta})] - \mathbb{E}[(N_{t+\tau}-N_t)]\mathbb{E}[(N_{t+2\tau+\delta}-N_{t+2\tau+\delta})]}{\sqrt{\mathrm{var}(N_{t+\tau}-N_t)\mathrm{var}(N_{t+2\tau+\delta}-N_{t+\tau+\delta})}} \\ &= \frac{e^{-2\beta\tau}\left(e^{\alpha\tau}-e^{\beta\tau}\right)^2\alpha(\alpha-2\beta)}{2\left(\alpha(\alpha-2\beta)\left(e^{(\alpha-\beta)\tau}-1\right)+\beta^2\tau(\alpha-\beta)\right)} e^{(\alpha-\beta)\delta}. \end{aligned} \tag{30}$$

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 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 in fact 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. $\sum_{i\leq nj\leq m} a_{ij}x^i x^j$) 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 et al. (2013) and Filipović et al. (2013). To be more precise, if we denote $Z_t = \left(\mathbb{E}[\lambda_t], \mathbb{E}[N_t], \mathbb{E}[\lambda_t^2], \mathbb{E}[\lambda_t N_t], \mathbb{E}[N_t^2]\right)^\top$ then using the ODEs obtained above the vector Z_t satisfies the ODE

$$\frac{dZ_t}{dt} = AZ_t + B \tag{31}$$

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

$$Z_t = e^{At} Z_0 + \int_0^t e^{A(t-s)} B ds \tag{32}$$

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

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 computed in the previous section we develop a GMM estimation strategy whose computational speed appears to be very fast compared to the existing alternatives.

3.1 Maximum likelihood estimation

Let $(X_t)_{t\geq 0}$ be a simple point process on $[0,T]$ and $t_1 ... t_{N_T}$ 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), among other references, the log-likelihood of $(X_t)_{t\geq 0}$ is of the form :

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

In the case of a Hawkes process we have

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

and simplifying the above integral

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

we end with

$$L = T - T\lambda_\infty - \sum_{i=1}^{N_T} \frac{\alpha}{\beta}\left(1 - e^{-\beta(T-t_i)}\right) + \sum_{i=1}^{N_T} \ln(\lambda_\infty + \alpha A(i)), \tag{33}$$

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

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

Some authors, as Ozaki (1979), pointed out that $A(i)$ used in (33) satisfies a recursive relation. Indeed, defining $A(1) = 0$ then for $i > 1$, $A(i+1) = e^{-\beta(t_{i+1} - t_i)} \times (1 + A(i))$, this property 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 is therefore of interest.

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 method of moments. The calibration procedure will then consist in a least squares fit of theoretical moments with the empirically observed ones. The problem writes as

$$\hat{\theta} = \text{argmin} \sum_{i=1}^{3} \left[M^i - f_i(\theta) \right]^2, \tag{34}$$

where the M^i are some empirically estimated moments of the considered sample, and $f_i(\theta)$ are the corresponding theoretical moments. As we have 3 parameters we use the same number of moments to build the function (34). The optimization problem can then be solved very quickly by Levenberg-Marquardt algorithm (we use this implementation by Lourakis (Jul. 2004)).

As we have more than three moments in closed form at our disposal and also the autocorrelation function we can rely on the Generalized Method of Moments (GMM), see Hall (2004) for an exhaustive treatment. The inference problem now writes as

$$\hat{\theta} = \text{argmin} \left\{ (M - f(\theta))^\top W (M - f(\theta)) \right\},$$

where M is now the vector of empirical moments (with eventually the autorcorrelation condition), and $f(\theta)$ is the vector of corresponding theoretical moments. W is a symmetric positive definite weight matrix so it is a quadratic form. As usual, if W is the identity matrix, and $\dim(W) = 3$, we recover our original minimization problem as written in equation (34).

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. Intuitively, the optimal weighting matrix will attribute less weight to the noisier moments, and more weight to moments that are easiest to estimate with better accuracy. It turns out indeed that the optimal weight matrix is the inverse variance-covariance matrix of the error terms, so in our notation $W_{opt} = \mathbb{E}\left[\left(M - f\left(\hat{\theta}\right)\right) \times \left(M - f\left(\hat{\theta}\right)\right)^{\top}\right]^{-1}$, with $\hat{\theta}$ being the true parameter value. As this true parameter value is unknown we rely on Hansen's two-step method Hansen (1982) to obtain the optimal weighting matrix:

- Let W be the identity matrix and estimate the parameters. This gives a consistent estimate θ_0.
- Update the weight matrix as : $W_{opt} = \mathbb{E}\left[(M - f(\theta_0)) \times (M - f(\theta_0))^{\top}\right]^{-1}$, finding θ_1, and iterate this step until convergence, i.e $\theta_{k+1} \approx \theta_k$.

Some numerical experiments lead us to the following conclusions: the optimization problem (34) based on the mean and variance of number jumps during an interval τ (i.e. equations (20) and (27)), and autocorrelation function (30) gives the best results if calibration quality and speed are taken into account. Even when there are more moments than the number of parameters, calculation time remains negligible and far more rapid than the MLE. Also we found it very important to take the autocorrelation function into account in the objective function.

Another numerical problem that we met is that the order of magnitude of the different moments is not the same. For example, while the autocorrelation is always less than 1 in absolute value, the mean or variance of the jump numbers grow linearly with the length of the estimation intervals, and can be typically of the order of hundreds for the applications considered. To deal with this problem, we slightly modify the optimization problem:

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

where the components of the vector $(1 - \frac{f(\theta)}{M})$ are $(1 - \frac{f_i(\theta)}{M_i})$ and involve the relative ratio of the

theoretical moment to its empirically estimated counterpart. This is made possible by the fact that all the considered moments are different from zero, and this solves the problem of orders of magnitude of the different moments[2].

Lastly and more importantly 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 execution, price impact analysis of a trade), but it has also the advantage of being robust to data pollution, an aspect which is very common in such data.

4 Applications

4.1 Data

We rely on tick-by-tick data from TRTH (Thomson Reuters Tick History). The data consist of trades and quotes files timestamped in milliseconds. We particularly studied two stocks: BNP Paribas and Sanofi, as well as two futures on indices : Eurostoxx and Dax. For each studied day we consider the front maturing futures, i.e. the futures with the shortest maturity date. The data cover the period between 2010/01/01 to 2011/12/31.

All the trading days begin at 9:00 and end at 17:30. We nevertheless 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; whereas in our moment based inference all, the trades will individually be taken into account in the moment calculation. This, among other

[2]Similar problem appears in Aït-Sahalia et al. (2010) where the authors use the matrix W to control for the discrepancies between the moments.

factors, will make the method of moments more robust to data imprecision, a fact typical to high frequency data.

4.2 Trade clustering

As outlined in the introduction, trading activity is not a completely random and memory-less process. If it was so, a Poisson process would have been a good candidate for 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.

[Insert Figure 1 here]

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 on Eurostoxx. The clustering is graphically clear.

[Insert Figure 2 here]

Numerous reasons can be mentioned to explain this clustering of trade arrivals, 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 the number of trades autocorrelation gives information about the degree of clustering. Needless to say that with a Poisson process generating trade arrival times this correlation would have been zero.

Precisely, we want to compute:

$$\mathrm{C}(\tau,\delta)=\frac{\mathbb{E}[(N_{t+\tau}-N_t)(N_{t+2\tau+\delta}-N_{t+\tau+\delta})]-\mathbb{E}[(N_{t+\tau}-N_t)]\mathbb{E}[(N_{t+2\tau+\delta}-N_{t+\tau+\delta})]}{\sqrt{\mathrm{var}(N_{t+\tau}-N_t)\mathrm{var}(N_{t+2\tau+\delta}-N_{t+\tau+\delta})}}. \tag{35}$$

For example, we count the number of trades occurring during two sliding non overlapping intervals of one $\tau = 1$ minute length separated by a certain time lag δ. We change this time lag from 1 second to 20 minutes by a step of 1 second. This gives 1200 number of trades autocorrelation points, allowing to plot the function $C(\tau, \delta)$. An example plot is given in figure 3, where we considered different time intervals τ for the same underlying symbol, namely 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.

[Insert Figure 3 here]

The same phenomenon is observed for all the symbols we tested. 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 tested. An illustration is given in figure 4.

[Insert Figure 4 here]

These stylized facts justify the use of the Hawkes process as modelling framework.

Thanks to the previous computations, we have an expression for the number of trades autocorrelation function (35) as in the case of the Hawkes process, namely equation (30). A least squares fit of this function to the empirical one can be performed to calibrate the model parameters. Note, however, that the autocorrelation function for the number of trades does not depend on λ_∞. Therefore, we rely on equation (20) to obtain λ_∞ from the other parameters[3]. Based on figure 5 we choose to fit the autocorrelation function (30) for $\tau = 1$ min and δ ranging from 0 to 600 seconds by steps of 60 seconds and figure 4 convinces us that this choice is suitable for all the analyzed stocks.

[Insert Figure 5 here]

We calibrated the model every day for a one-year data sample for the four considered symbols. Results are summarized in table 1. A striking fact is the small ratio of the λ_∞ to a Poisson equivalent λ that

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

we put in the table for reference: it is the intensity we obtain if we calibrate a Poisson process to the data. Indeed, λ_∞ is always between 5% to 10% of its Poisson equivalent, showing that the Hawkes process attributes only this fraction of events to a background process activity, while explaining the remaining part by the branching aspect of the process, i.e. the self-exciting property.

[Insert Table 1 here]

The best test for the calibration procedure is if the so calibrated model does reproduce the empirical stylized fact that we were interested in, namely the trade arrival time clustering. This is indeed the case. Examples of fit of the autocorrelation functions of the number of trades are shown in figure 6. The good quality of the fit advocates the Hawkes process as a good modeling tool for trade arrival times.

[Insert Figure 6 here]

For robustness, we also conducted 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 2.

[Insert Table 2 here]

In all our results, one can notice the moderate values of standard deviations as compared to mean parameter values, as well as the similarity of mean and median values, denoting the relative stability of the calibration and absence of outliers. Additionally, in all our calibrations we did not include the stability constraint $\alpha < \beta$, the calibration automatically filled this constraint. Indeed, it is a necessary condition to fit the decreasing shape of the autocorrelation function as it is clear from equation (30).

4.3 Forecasting

Once the model is calibrated, one is naturally interested in its forecasting power. In our case, as stressed in the preceding section, the model essentially captures the autocorrelation of the number of trades occurring in consecutive time intervals. The shape of this function, exponentially decreasing function to zero, naturally suggests to model the number of trades occurring in fixed time intervals as an autoregressive process.

Given a set of moments $t_0 < t_1 < t_2 < \ldots < t_N$ with $t_i - t_{i-1} = \tau$, we define $y_{t_i} = \frac{1}{\tau}(N_{t_i} - N_{t_{i-1}})$ and denote y_t with $t \in \{t_1, \ldots, t_N\}$. For simplicity, we work with the centered version of y_t and consider the autoregressive model:

$$y_t = a_1 y_{t-1} + u_t$$

where u_t is a random variable, independent of y_t with 0 mean. Then, in the case of a Hawkes model, the parameters of the AR model for the number of trades can be computed. Indeed, from the previous results we have

$$\mathbb{E}_{t-1}[y_t^2] = a_1 \mathbb{E}_{t-1}[y_t y_{t-1}] + \mathbb{E}_{t-1}[y_t u_t]$$

leading to

$$a_1 = \frac{\mathbb{E}_{t-1}[y_t y_{t-1}]}{\mathbb{E}_{t-1}[y_t^2]} = \frac{\mathrm{Cov}(\tau, 0)}{\mathrm{V}(\tau)} = \mathrm{Acf}(\tau, 0)$$

where the autocovariance and variance functions are known.

The same reasoning can be done if one considers an AR(p) model :

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + ... + a_p y_{t-p} + u_t$$

then for every $i \in \{1, .., p\}$, multiplying the above equation by y_{t-i} and taking the expectation yields p linear equations with coefficients a_i :

$$C(\tau, \delta = (i-1) \times \tau) = V(\tau) a_i + \sum_{j=1, j \neq i}^{p} a_j C(\tau, \delta = (|i-j| - 1) \times \tau)$$

therefore the $(a_1, \ldots, a_p)^\top$ is the solution of the linear system:

$$\begin{pmatrix} 1 & \mathrm{Acf}[\tau,0] & ... & ... & \mathrm{Acf}[\tau,(p-2)\tau] \\ \mathrm{Acf}[\tau,0] & 1 & \mathrm{Acf}[\tau,0] & & \mathrm{Acf}[\tau,(p-3)\tau] \\ ... & & & & \\ ... & & & & \\ \mathrm{Acf}[\tau,(p-2)\tau] & & & \mathrm{Acf}[\tau,0] & 1 \end{pmatrix} \times \begin{pmatrix} a_1 \\ ... \\ ... \\ ... \\ a_p \end{pmatrix} = \begin{pmatrix} \mathrm{Acf}[\tau,0] \\ ... \\ ... \\ ... \\ \mathrm{Acf}[\tau,(p-1)\tau] \end{pmatrix}. \tag{36}$$

Solving this equation for a sufficiently large first guess of p allows the calculation of the a_i coefficients and to decide the length of the autoregressive process by discarding the small coefficients.

We conduct the forecast experiment by daily calibrating the Hawkes process and use the parameters of each calibrated day as forecast parameters for the following day. In order to assess the quality of this prediction, we calculate the residues of our forecast and compare them to the residues of a forecast that relies on a Poisson process. Results are summarized in Table 3. They show a superior forecasting power for the Hawkes process. Table 4 presents the distribution of the number of lags necessary for the prediction.

[Insert Table 3 here]

[Insert Table 4 here]

Nevertheless, both models give exactly the same results when the forecast horizon is sufficiently large (half an hour in our case). As an illustration, notice that in Figure 7, the standard deviation of Hawkes forecast residues converges to that of the Poisson residues when the forecast horizon increases. Moreover, as illustrated by figure 8, both distributions of residues converge when the forecast horizon is sufficiently large. This is an expected result as the Hawkes forecast method relies on the autocorrelation of the number of trades falling in consecutive intervals, which is significant for small time lags, but disappears after twenty to thirty minutes as previously shown in the paper.

[Insert Figure 7 here]

[Insert Figure 8 here]

4.4 Branching structure of trading activity

An interesting property of the Hawkes process is its branching structure. Indeed, the occurence of a jump increases the intensity of the process, and 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.

Therefore, 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

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

Expressed in terms of jump counts, this impulse response is equal to:

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

where $N_{response}$ denotes the expected number of jumps triggered by one jump occurring at time 0 if the process is observed indefinitely.

We use our daily calibrations on real data of the Hawkes process to measure an average $N_{response}$ for the studied assets. We obtain the following table:

[Insert Table 5 here]

We clearly observe a difference between futures, given by the symbols Dax and Eurostoxx, and the stocks represented by BNPP and Sanofi. This points towards considering this number as an indicator of liquidity and trading activity. For instance, despite the fact that the estimated parameters for the Dax and the Eurostoxx are close, the impulse response value for the former is larger and is consistent with the fact that futures on this index are more actively traded, due to a stronger branching structure.

A robustness check was also performed for other assets in the previously mentioned table 2. 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 others. Among the commodities, the Crude Oil Brent dominates the Natural Gas, Sugar, Corn and the Wheat.

4.5 Diffusive limit and signature plot

In all the preceding sections, we dealt with the trading process from a microscopic point of view i.e. at the transaction level. In the classical High Frequency literature mainly developed by econometricians,

most of the studies deal with this same level of granularity trying to uncover the intimate mechanisms of the price formation process. Many models try to encompass the subtle interactions of the many components of the trading process (order flow, order signs, quantities and so on...). For instance, ACD models as proposed in Engle and Russell (1998) fall in this trend of research. For an overview of models and techniques involved see Hautsch (2012) and references therein.

In our paper, we consider a simpler framework focusing only on the order flow. Certainly, this simplification comes at a cost of neglecting important aspects of the price formation process, such as the volume for example. Nevertheless, it allows us to address the important question of connecting the microscopic price formation process 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: Abergel and Jedidi (2013) model the order book as a multidimensional Markov chain with independent Poissonian order arrival times and prove the convergence of the price process to a Brownian motion. In Cont and De Larrard (2011) and Cont and De Larrard (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 these same ideas 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. (2013) introduce a model for microstructure price evolution based on mutually exciting Hawkes processes and connect the statistically observed signature plot of volatility and Epps effect of asset correlations to the microstructure price formation mechanism. They also establish diffusive limits of such models in Bacry et al. (2012).

In this section, we closely follow Bacry et al. (2013) in our setting of the univariate Hawkes process. We consider a toy model for the movements of the *mid* price of a traded asset. The *mid* price is the mean of the best ask price in the order book and the best bid price. 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.[4]. The model writes:

$$S_t = S_0 + \left(N_t^{up} - N_t^{down}\right)\frac{\delta}{2}, \tag{37}$$

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 a price explosion. In the stationary regime their intensities are given by:

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

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

In Bacry et al. (2012), 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[5]. In our case, as the kernel is exponential, the process $X_t = \left(S_t, N_t^{up}, \lambda_t^{up}, N_t^{down}, \lambda_t^{down}\right)$ is a Markov process. Its infinitesimal generator writes :

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

The availability of the infinitesimal generator allows the specifications of the conditions ensuring the stability results. For instance, ergodicity of the process X_t, that is roughly speaking, its convergence to a stationary regime, can be easily established thanks to the test of the function criterion based on Foster-Lyapounov inequalities. We refer to Meyn and Tweedie (2009) for a detailed exposition. In our case, we define the function $V(x) = \frac{\lambda^{up}+\lambda^{down}}{2\lambda_\infty}$, then a simple calculation yields the *geometric drift condition*:

$$\mathcal{L}V(x) \leq (\alpha - \beta)V(x) + \beta, \tag{40}$$

[4]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

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

which grants, thanks to (CD3) in Meyn and Tweedie (1993), the V-uniform ergodicity of the process X_t.

Let us then write unit-time price increments:

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

and consider the random sums

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

where $\{\bar{\eta_i}; i = 1 \dots n\}$ denote the centered price increments. We focus on the asymptotic behaviour of the rescaled (centered) 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) allows us to conclude that the increments are geometrically mixing and theorem 19.3 of Billingsley (1999) proves therefore, that $\bar{S_t^n}$ converges to a Brownian motion in the sense of Skorokhod topology.

$$\bar{S_t^n} \Rightarrow \sigma W_t.$$

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:

$$\begin{aligned}
\sigma^2 &= \lim_{n\to\infty} \frac{\mathrm{Var}(\bar{S_n})}{n} \\
&= 2\frac{\delta^2}{4}\left(\mathbb{E}[\bar{\eta_0}^2] + 2\sum_{n=0}^{\infty}\mathbb{E}[\bar{\eta_0}\bar{\eta_n}])\right) \\
&= \frac{\delta^2}{2}\left(V[1] + \sum_{n=0}^{\infty} 2\overline{\mathrm{Cov}}[1,n]\right),
\end{aligned}$$

where thanks to (29), the centered covariance $\overline{\mathrm{Cov}}[1,n]$ writes :

$$\overline{\mathrm{Cov}}[1,n] = \frac{\lambda_\infty \beta\alpha(2\beta-\alpha)\left(e^{(\alpha-\beta)\tau}-1\right)^2}{2(\alpha-\beta)^4} e^{(\alpha-\beta)n}.$$

Then, summing up with the expression of the variance, and simplifying, one obtains:

$$\sigma^2 = \frac{\delta^2}{2}\frac{\lambda_\infty \beta^3}{(\beta-\alpha)^3}. \tag{41}$$

Notice the importance of the denominator $\beta - \alpha$ in this formula. The smaller this difference is, the greater is the volatility. This is an expected feature because from equation (30) this difference characterizes the persistence of the autocorrelation of the Hawkes processes involved in the model, and

therefore the autocorrelation of mid price movements in either side inflates the process' variance.

Besides giving a framework that allows the connection of the the microscopic price formation mechanism to its macroscopic behaviour, as shown above, the Hawkes process can reproduce some stylized facts across time scales. Among these stylized facts is the volatility signature plot.

In Bacry et al. (2013), signature plot is presented as "an increase of the observed daily variance when one goes from large to small scale". 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 leads to a decreasing volatility as a function of the the sampling period. These patterns are the most frequent but inverted patterns are also possible and can be found in data, even for liquid stocks as illustrated in figure 9. Let us stress that these figures are realized based on mid price sampling, eliminating any noise due to bid-ask bounce.

[Insert Figure 9 here]

We can calculate the mean signature plot in our toy model. The realized variance over a period T calculated by sampling the data by time intervals of length τ can be written thanks to (37):

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

Therefore, the mean signature plot writes:

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

where

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

Notice that when τ becomes larger the above expression converges to the asymptotic diffusive variance of the model calculated in (41). The mean signature plot leads to an increasing shape of the volatility as the sampling period increases. This captures situations as the one observed in figure 9.

In Bacry et al. (2013), the authors propose a model similar to (37) 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:

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

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. The diffusive limit for this model is

$$\sigma^2 = \frac{\delta^2}{2} \frac{\lambda_\infty \beta^3}{(\beta-\alpha)(\beta+\alpha)^2}, \tag{42}$$

and the mean signature plot is

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

with

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

As a function of the sampling period the signature plot is decreasing (or equivalently the signature plot is increasing with the sampling frequency) and this is due to the serial negative autocorrelation of the returns. This corresponds to the more frequent shape observed in the market. However, an inverted shape can also happen on the market.

Notice that we also obtain 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 both in our model (41) and in Bacry et al. (2013)'s specification (42). Results are reported in table 6. 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. (2013)'s model systematically underestimates it. Qualitatively, this is due to the fact that our model expands up and down movements, thus inflating the realized volatility, whereas Bacry et al. (2013)'s model moderates up and down movements thanks to its mean reversion mechanism, thus underestimating the volatility.

[Insert Table 6 here]

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 yet keeping the computations tractable. It would allow us to take into account these effects, that were underlined in the empirical literature (see Hautsch (2012)), on the price dynamics and provide us with a deeper understanding of how the volatility measured at a macroscopic level (at a daily frequency) depends on the trading activity observed at a microscopic level (at high frequency).

5 Conclusion

In this paper we compute explicitly 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 which 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 in real time the model. We use our estimation framework to calibrate the Hawkes process on trades for four stocks over a one-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 two 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.

We perform a forecast analysis to determine the horizon beyond which the Hawkes process does not perform better than a simple model. This horizon is found to be smaller or equal to 10 minutes which is satisfactory as high frequency trading occurs within this time range. Thanks to the analytical tractability of the Hawkes process we compute explicitly the impulse response associated with the process which determines the market impact of a trade.

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

Our work points towards 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. (2013), 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 computation in the multidimensional case. As a matter of fact the computations performed here can be carried out for that case.

Another aspect of interest is the diffusive limit concept. In this work we connect the dynamics 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 modelling, for example, the level I quotes. The Hawkes process provides a natural modelling framework and would extend the interesting existing models based on the Poisson process. To perform such a limit the moments as well as the autocorrelation are needed and can be obtained using the the computation strategy developed in this work. These interesting problems are left for future work.

References

F. Abergel and A. Jedidi. A Mathematical Approach to Order Book Modeling. *International Journal of Theoretical and Applied Finance*, 2013.

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

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

E. Bacry, S. Delattre, M. Hoffmann, and J. F. Muzy. Scaling limits for hawkes processes and application to financial statistics. 2012.

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

P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. 1999.

C. G. Bowsher. Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, (2):876–912, 12 2007.

P. Brémaud. *Point Processes and Queues, Martingale Dynamics*. Springer-Verlag, 1981.

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

R. Cont and A. De Larrard. Price Dynamics in a Markovian Limit Order Market. *Social Science Research Network Working Paper Series*, Jan. 2011.

R. Cont and A. De Larrard. Order Book Dynamics in Liquid Markets: Limit Theorems and Diffusion Approximations. *Social Science Research Network Working Paper Series*, Feb. 2012.

R. Cont, S. Stoikov, and R. Talreja. A stochastic model for order book dynamics. *Operations Research*, 58(3):549–563, 2010.

C. Cuchiero, M. Keller-Ressel, and J. Teichmann. Polynomial processes and their applications to mathematical finance. *Finance and Stochastics, forthcoming*, 2013.

D. J. Daley and D. V. Jones. *An introduction to the theory of point processes, Vol 1, Elementary theory and methods*. Springer, 2. ed. edition, 2002.

D. J. Daley and D. V. Jones. *An introduction to the theory of point processes, Vol 2, General theory and structure*. Springer, 2. ed. edition, 2008.

D. Duffie and R. Kan. A Yield-Factor Model of Interest Rates. *Mathematical Finance*, 6(4):379–406, 1996.

R. F. Engle and J. R. Russell. Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica*, 66(5):1127–1162, Sept. 1998.

E. Errais, K. Giesecke, and L. R. Goldberg. Affine point processes and portfolio credit risk. *SIAM J. Financial Math.*, 1(1):642–665, 2010.

D. Filipović, E. Mayerhofer, and P. Schneider. Density approximations for multivariate affine jump-diffusion processes. *forthcoming Journal of Econometrics*, 2013.

G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins Studies in Mathematical Sciences. The Johns Hopkins University Press, 1996.

A. R. Hall. *Generalized Method of Moments*. Advanced Texts in Econometrics. OUP Oxford, 2004. ISBN 9780198775201.

L. P. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–54, July 1982.

N. Hautsch. *Econometrics of Financial High-frequency Data*. Springer Berlin Heidelberg, 2012.

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

P. Hewlett. Clustering of order arrivals, price impact and trade path optimisation. 2006.

A. Kirilenko, R. B. Sowers, and X. Meng. A multiscale model of high-frequency trading. *Algorithmic Finance*, 2(1): 59–98, Feb. 2013.

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

E. Lewis and G. Mohler. A nonparametric em algorithm for multiscale hawkes processes, 2011.

M. Lourakis. levmar: Levenberg-marquardt nonlinear least squares algorithms in C/C++. [web page] `http://www.ics.forth.gr/~lourakis/levmar/`, Jul. 2004. [Accessed on 31 Jan. 2005.].

S. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, 2nd edition, 2009.

S. P. Meyn and R. L. Tweedie. Stability of Markovian Processes III: Foster-Lyapunov Criteria for Continuous-Time Processes. *Advances in Applied Probability*, 25(3), 1993.

G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita. Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493):100–108, 2011.

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

T. Ozaki. Maximum likelihood estimation of hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics*, 31(1):145–155, 1979.

F. Pomponio and F. Abergel. Multiple-limit trades: empirical facts and application to lead-lag measures. *Quantitative Finance*, 13(5):783–793, 2013.

D. Revuz and M. Yor. *Continuous martingales and Brownian motion*. 3rd ed, Springer-Verlag, 1999.

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

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

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

Appendix

Some Useful Expressions

Skewness

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

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

The stationary regime third moment then writes:

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

Kurtosis

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

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

The stationary regime fourth moment of the process writes:

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

Tables

Symbol	Measure	Poisson λ	λ_∞	α	β
Eurostoxx	Mean	1.4480	0.0625	0.0869	0.0911
	Std. dev.	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
	Std. dev.	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
	Std. dev.	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
	Std. dev.	0.1873	0.0213	0.0212	0.0240
	Median	0.6087	0.0414	0.0700	0.0758

Table 1: Calibration results for two years of data. We calibrate daily a Hawkes process to the trade arrival time of each symbol. For comparison, we put the Poisson equivalent λ, defined as the mean number of trades per second for every day, and characterizing the trading activity on each symbol. For every measure, we put mean, standard deviation and median value of the calibration parameters.

Symbol	Measure	Poisson λ	λ_∞	α	β	$N_{response}$
Bund	Mean	1.2742	0.0671	0.0956	0.1013	19.683
	Std. dev.	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
	Std. dev.	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
	Std. dev.	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
	Std. dev.	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
	Std. dev.	1.7419	0.0271	0.0192	0.0194	19
	Median	4.1956	0.0770	0.1258	0.1282	52.045
GOLD	Mean	2.3191	0.0852	0.1104	0.1149	27.716
	Std. dev.	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
	Std. dev.	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
	Std. dev.	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
	Std. dev.	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
	Std. dev.	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
	Std. dev.	0.4334	0.0242	0.0215	0.0227	8.9511
	Median	1.0926	0.0594	0.1182	0.1244	17.372

Table 2: Calibration results for two 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 every asset class considered we take daily data of the front maturing future to make our calibrations. The last column $N_{response}$ will be defined in section 4.4

		Symbol							
		Eurostoxx		Dax		BNPP		Sanofi	
Horizon	Measure	Hawkes	Poisson	Hawkes	Poisson	Hawkes	Poisson	Hawkes	Poisson
1	Mean	-0.0083	0.0523	-0.0207	0.0499	0.0087	0.0330	0.0084	0.0234
	Std. dev.	1.1836	1.4243	1.3865	1.7028	0.6252	0.6833	0.4031	0.4278
	Skew	-0.3341	-0.9205	-0.3237	-0.7918	-0.4882	-1.4757	-0.9462	-2.2286
5	Mean	0.0322	0.0498	0.0312	0.0489	0.0295	0.0330	0.0195	0.0214
	Std. dev.	1.3200	1.4262	1.5742	1.7028	0.6402	0.6812	0.4111	0.4301
	Skew	-0.9645	-0.9184	-0.8519	-0.7904	-1.4394	-1.4712	-2.1188	-2.2276
10	Mean	0.0442	0.0472	0.0467	0.0492	0.0358	0.0321	0.0205	0.0186
	Std. dev.	1.3742	1.4280	1.6382	1.7009	0.6632	0.6813	0.4275	0.4352
	Skew	-0.9915	-0.9152	-0.8625	-0.7884	-1.5231	-1.4626	-2.2051	-2.2064
30	Mean	0.0348	0.0300	0.0499	0.0448	0.0299	0.0225	0.0085	0.0046
	Std. dev.	1.4297	1.4331	1.6861	1.6896	0.6839	0.6853	0.4482	0.4493
	Skew	-0.8972	-0.8929	-0.7794	-0.7784	-1.4160	-1.4238	-2.1542	-2.1466

Table 3: For every symbol, mean, standard deviation and skew of the forecast residues are given for Hawkes process and Poisson process prediction method. The superiority of a Hawkes process forecast over a Poisson process forecast disappears when the forecast horizon increases.

Symbol	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=6$
Eurostoxx	0.0020	0.0122	0.0446	0.3681	0.5691	0.0040
Dax	0	0.0020	0.0100	0.1896	0.7944	0.0040
BNPP	0	0.0123	0.1284	0.6403	0.2190	0
Sanofi	0.0021	0.0576	0.1725	0.6350	0.1328	0

Table 4: For every symbol, we present the frequency of the number of lags that were necessary to make the forecast in the AR(p) model of the number of trades occurring over a time interval. The system (36) is solved for a sufficiently large dimension. We consider p to be the rank from which the autoregressive coefficients become non significant. We considered a cut off absolute value of 0.001.

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

Table 5: $N_{response}$ as a characteristic of market liquidity. Symbols are ranked from the most liquid to the less liquid.

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%

Table 6: Median values of asymptotic volatilities as calculated by our toy model, as well as that calculated thanks to BDHM model (Bacry et al., 2013) compared to realized volatilities of the day. We also put median values of Hawkes model parameters. Both of the models give plausible values. BDHM volatility underestimates systematically the realized volatility, whereas our toy model systematically overestimates it.

Figures

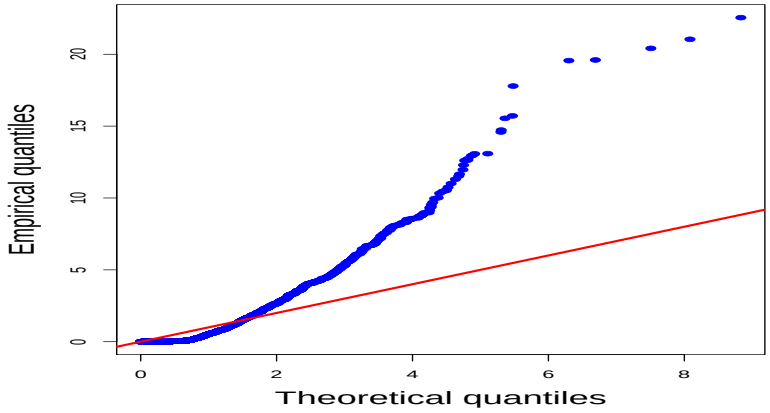

Figure 1: QQplot of inter-trade durations against exponential distribution. Interarrival trade times are clearly not exponential. Graph for Eurostoxx futures trade time durations on 2011/03/03 for the first trading hour.

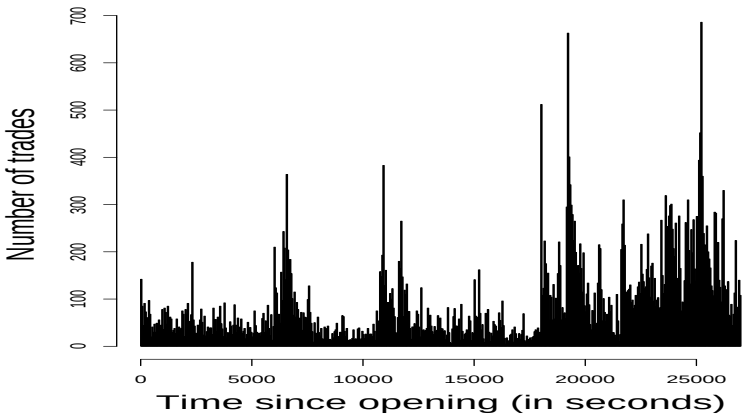

Figure 2: Number of trades by time intervals of 1 minute: presence of clusters is graphically apparent. Graph for Eurostoxx futures trade time durations on 2011/03/03.

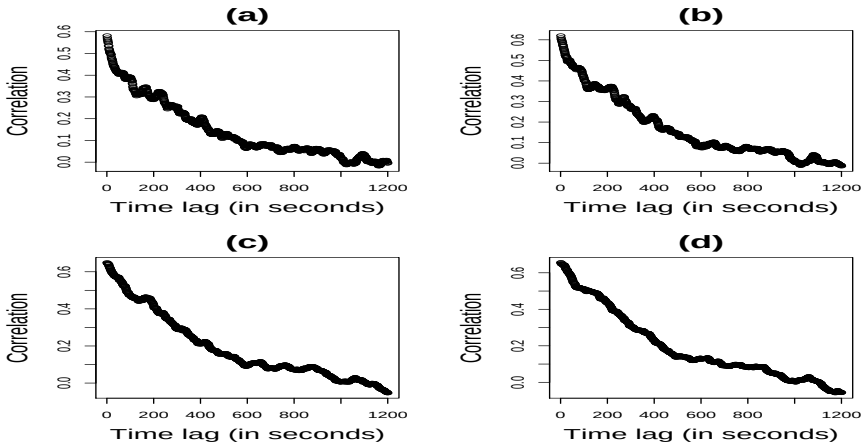

Figure 3: Empirical autocorrelation of the trade numbers occurring on consecutive intervals separated by a time lag, as a function of this latter. In the four plots, we varied the length of the time interval considered. (a) correponds to 20 seconds, (b) 30 seconds, (c) 60 seconds and (d) 90 seconds. The shape of the function remains identical, even if the plot is noisier when the time interval length decreases. The data are the Eurostoxx futures trades on 2010/01/07.

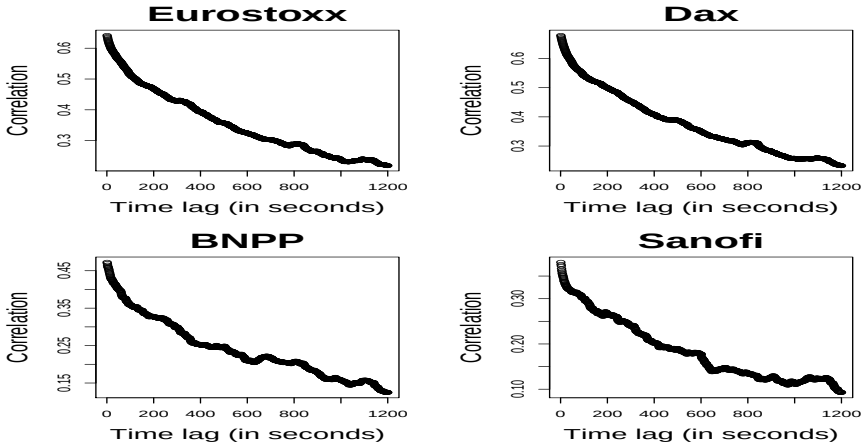

Figure 4: Empirical autocorrelation of the trade numbers of 1-minute intervals for Eurostoxx, Dax, BNPP and Sanofi, averaged every day for the month January 2010.

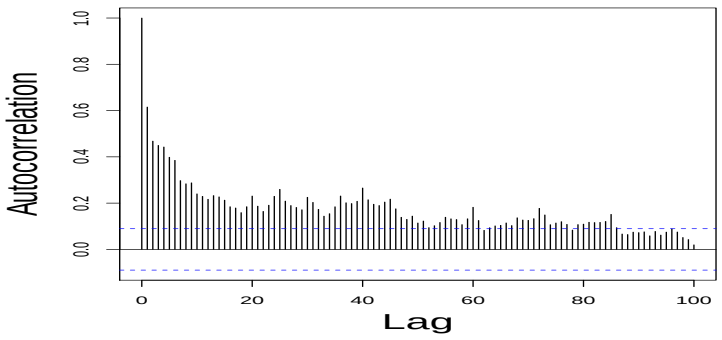

Figure 5: Number of trades empirical autocorrelation. The number of trades are computed for an interval $\tau = 1$ minute and the lag δ is also measured in minutes. The symbol is Eurostoxx on 2010/10/26.

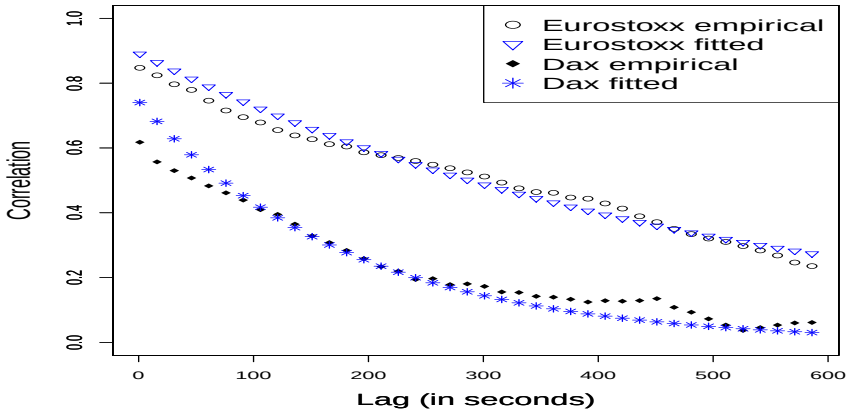

Figure 6: Empirical autocorrelation of the time series of the number of trades occurring during $\tau = 1$ minute versus the theoretically fitted one. For the Dax, fitted parameters are : $\lambda_\infty = 0.0326806$, $\alpha = 0.0431643$ and $\beta = 0.0486235$ and for Eurostoxx, fitted parameters are : $\lambda_\infty = 0.033282$, $\alpha = 0.04259$ and $\beta = 0.0446049$.

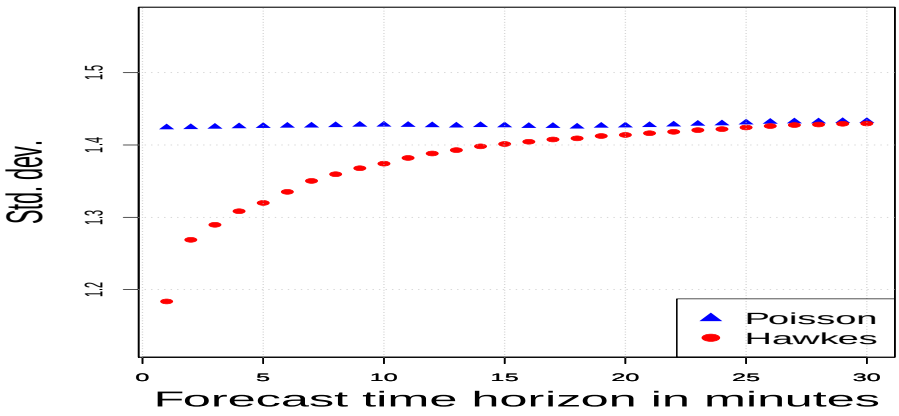

Figure 7: Standard deviation of residues of Hawkes forecasts vs. Poisson forecasts as a function of the forecast time horizon.

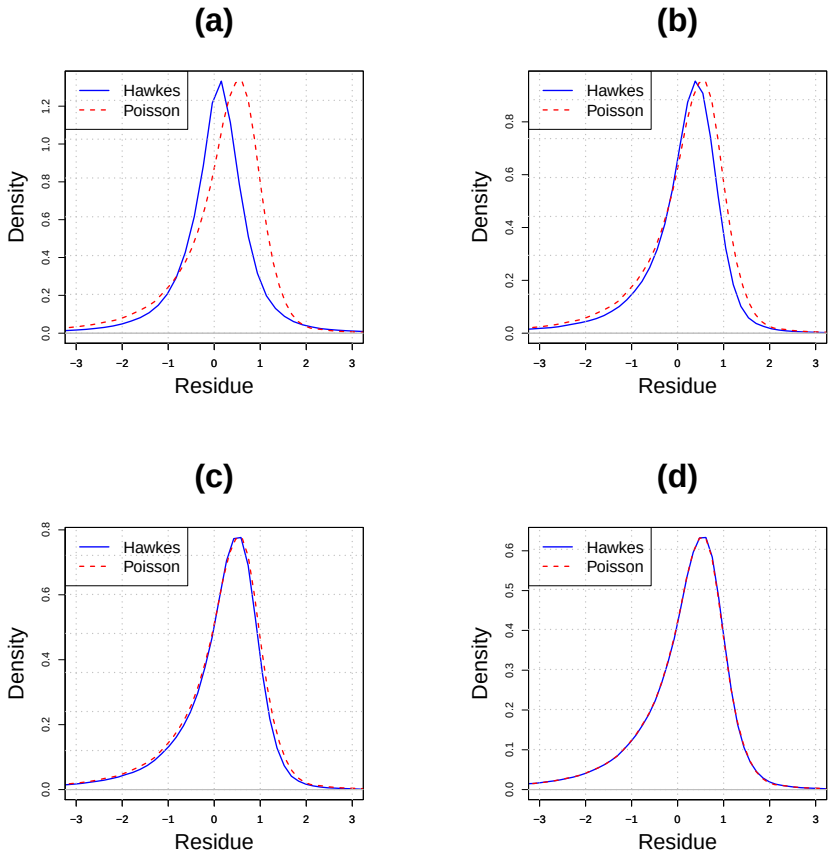

Figure 8: Densities of residues collapsing when forecast horizon increases. (a), (b), (c) and (d) correspond to forecast horizons of 1 minute, 5 minutes, 10 minutes and 30 minutes, respectively.

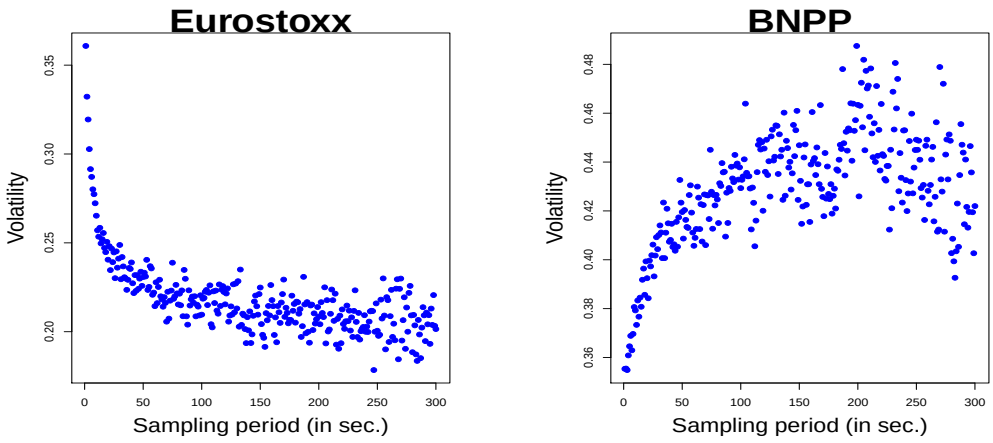

Figure 9: Signature Plot of Eurostoxx futures and BNPP stock on 2011/04/01, computed on mid prices to eliminate bid-ask bounce.