Economics
The Open-Access, Open-Assessment E-Journal

Vol. 6, 2012-22 | June 14, 2012 | http://dx.doi.org/10.5018/economics-ejournal.ja.2012-22

Modelling Trades-Through in a Limit Order Book Using Hawkes Processes

Ioane Muni Toke
Ecole Centrale Paris and University of New Caledonia

Fabrizio Pomponio
Ecole Centrale Paris and BNP Paribas, Paris

Abstract The authors model trades-through, i.e. transactions that reach at least the second level of limit orders in an order book. Using tick-by-tick data on Euronext-traded stocks, they show that a simple bivariate Hawkes process fits nicely their empirical observations of trades-through. The authors show that the cross-influence of bid and ask trades-through is weak.

Special Issue
New Approaches in Quantitative Modeling of Financial Markets

JEL C32, C51, G14
Keywords Hawkes processes; limit order book; trades-through; high-frequency trading; microstructure

Correspondence Ioane Muni Toke, Chair of Quantitative Finance, MAS Laboratory, Ecole Centrale Paris, Grande Voie des Vignes, 92295 Châtenay-Malabry Cedex, France.
E-mail: ioane.muni-toke@ecp.fr

Citation Ioane Muni Toke and Fabrizio Pomponio (2012). Modelling Trades-Through in a Limit Order Book Using Hawkes Processes. *Economics: The Open-Access, Open-Assessment E-Journal*, Vol. 6, 2012-22. http://dx.doi.org/10.5018/economics-ejournal.ja.2012-22

1 Introduction

Recent contributions have emphasized that Hawkes processes exhibit interesting features for financial modelling. For example, these self- and mutually exciting point processes can model arrival times of orders in an order book model (Large, 2007; Muni Toke, 2011), or explain the Epps effect in a microstructure toy model (Bacry et al., 2012). An econometric framework has been derived by Bowsher (2007).

In this paper, we are interested in modelling trades-through, i.e. transactions that reach at least the second level of limit orders in an order book. Trades-through are very important in price formation and microstructure. Since traders usually minimize their market impact by splitting their orders according to the liquidity available in the order book, trades-through may contain information. They may also reach gaps in orders books, which is crucial in price dynamics.

In a first part, we give basic statistical facts on trades-through, focusing on their arrival times and clustering properties. Our second part is a general introduction to Hawkes processes. In a third part, using tick-by-tick data on Euronext-traded stocks, we show that a simple bi-dimensional Hawkes process fits nicely our empirical data of trades-through. We show that the cross-influence of bid and ask trades-through is weak. Following Bowsher (2007), we improve the statistical performance of our maximum likelihood calibrations by enhancing the stationary model using deterministic time-dependent base intensity.

2 Trades-through

2.1 Orders splitting and trades-through

It has been shown several times that the time series built from trading flows are long-memory processes (see e.g. Bouchaud et al., 2009). Lillo and Farmer (2004) argue that this is mainly explained by the splitting of large orders. Indeed, let us assume that a trader wants to trade a large order. He does not want to reveal its intentions to the markets, so that the price will not “move against him”. If he were to submit one large market order, he would eat the whole liquidity in the order book, trading at the first limit, then the second, then the third, and so on. When "climbing the ladder" this way, the last shares would be bought (resp. sold) at a price much higher (resp. lower) than the first ones. This trader will thus split its large order in several smaller orders that he will submit one at a time, waiting between each submitted order for some limit orders to bring back liquidity in the order book. We say that the trader tries to minimize its market impact.

In practice, this mechanism is widely used: traders constantly scan the limit order book and very often, if not always, restrict the size of their orders to the quantity available at the best limit. But sometimes speed of execution is more important than minimizing market impact. In this case, orders larger than the size of the first limit may be submitted: thus, trades-through are precisely the trades that stand outside the usual trading pattern, and as such are worth being thoroughly studied.

Trades-through have already been empirically studied by Pomponio and Abergel (2010): their occurrences, links with big trades, clustering, intraday time-stamps distribution, market impact, spread relaxation and use in lead-lag relation. In this paper, we model trades-through with Hawkes processes.

2.2 Definition of trades-through

In general, we call a n-th limit trade-through any trade that consumes at least one share at the n-th limit available in the order book. For example, a second limit trade-through completely consumes the first limit available and begins to consume the second limit of the order book. Our definition is inclusive in the sense that, if p is greater than q, any p-th limit trade-through is also a q-th limit trade-through. In this study, we will focus on second limit trades-through, and simply call them trades-through in what follows. Figure 1 shows an example of trade-through.

2.3 Empirical data

We now describe the empirical data that will be used in the remaining of the paper. We use Thomson-Reuters tick-by-tick data of the Euronext-Paris limit order book for the stock BNP Paribas (BNPP.PA) from June 1st, 2010 to October 29th, 2010, i.e. 109 trading days. This data gives us trades (timestamp to the millisecond, volume and price) and quotes (volume, price, side of the order book) for the stock,

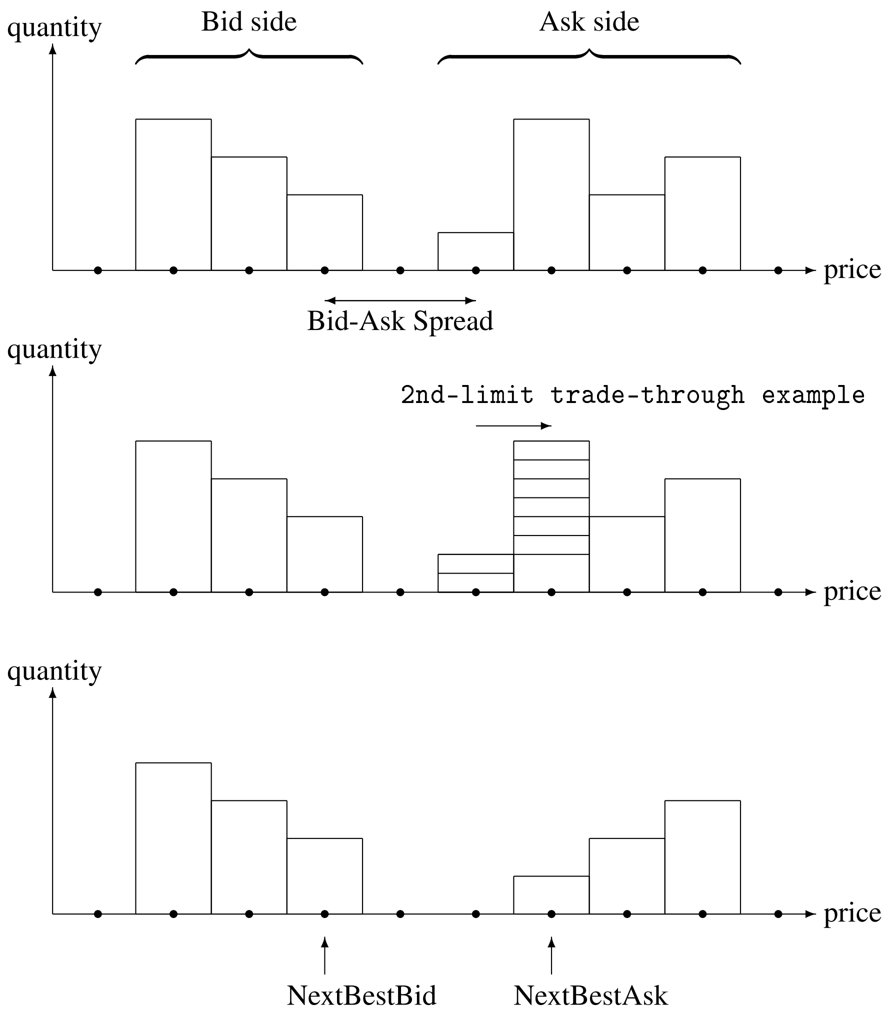

Figure 1: Example of a trade-through: *(up)* Limit order book configuration before the trade-through; *(middle)* Trade-through; *(down)* Limit order book configuration after the trade-through.

Limit considered	Number of trades-through per day (all)	Number of trades-through per day (bid side)	Number of trades-through per day (ask side)
2	829.0	401.8	427.2
3	124.1	59.0	65.1
4	30.5	14.6	15.9

Table 1: Occurrences of trades-through at bid and ask sides for BNP Paribas.

from the opening to the close of the market. For each trading day, we extract the series of timestamps $\left(t_i^A\right)_{i\geq 1}$ and $\left(t_i^B\right)_{i\geq 1}$ of the trades-through.

2.4 Occurrences of trades-through

We look at the occurrences of trades-through on the different sides of the order book. Basic statistics are given in Table 1. We can see that for second limit trades-through, there are around 400 events per day on each side of the book.

2.5 Clustering

Trades-through are clustered both in physical time and in trade time (see Pomponio and Abergel, 2010). Here we study in detail several aspects of this problem that will be helpful for further modelling: is the global clustering of trades-through still true when looking only at one side of the book? If so, is there an asymmetry in trades-through clustering on the bid and on the ask sides? Is there a cross-side effect for trades-through, in other words will a trade-through on one side of the book be followed more rapidly than usual by a trade-through on the other side of the book? Which is the stronger from those different effects?

In order to grasp the clustering of trades-though, we compute the mean of the distribution of waiting times between two consecutive trades-through, and we compare it with the mean waiting time between one trade (of any kind) and the next trade-through.

Table 2 summarizes our result on BNP Paribas stock in the considered period of study. We use the notation λ when looking at trades-through and Λ when looking at all the trades. When a specific side of the book is under scrutiny we mention it with a $+$ for ask side and a $-$ for bid side. For example, $(\Lambda^+) \to (\lambda^+ + \lambda^-)$

Impact studied	Mean waiting time until next trade-through (in seconds)
$(\lambda^+ + \lambda^-) \to (\lambda^+ + \lambda^-)$	36.9
$(\Lambda^+ + \Lambda^-) \to (\lambda^+ + \lambda^-)$	51.8
$(\lambda^+) \to (\lambda^+ + \lambda^-)$	36.3
$(\Lambda^+) \to (\lambda^+ + \lambda^-)$	51.7
$(\lambda^-) \to (\lambda^+ + \lambda^-)$	37.5
$(\Lambda^-) \to (\lambda^+ + \lambda^-)$	51.7
$(\lambda^+) \to (\lambda^+)$	76.1
$(\Lambda^+) \to (\lambda^+)$	107.9
$(\lambda^-) \to (\lambda^-)$	71.6
$(\Lambda^-) \to (\lambda^-)$	98.1
$(\lambda^+) \to (\lambda^-)$	80.4
$(\Lambda^+) \to (\lambda^-)$	101.8
$(\lambda^-) \to (\lambda^+)$	91.1
$(\Lambda^-) \to (\lambda^+)$	111.6

Table 2: Clustering of trades-through on bid and ask sides (on BNP Paribas data).

means that we look at the time interval between a trade at the ask side and the next trade-through, whatever its sign.

Analysing the first group of statistics ($(\lambda^+ + \lambda^-) \to (\lambda^+ + \lambda^-)$ and $(\Lambda^+ + \Lambda^-) \to (\lambda^+ + \lambda^-)$), we see that previous result on global clustering of trades-through is confirmed: you wait less the next trade-through when you already are on a trade-through, compared to when you are on a trade. Moreover, when looking at the second group of statistics, we see there is no asymmetry in this effect: both trades-through at the ask and at the bid are more closely followed in time by trades-through (whatever their sign), than trades at the bid and trades at the ask are.

The third group of statistics indicates that if you restrict the study to only one side of the book, the clustering is still valid. Finally, the fourth group of statistics shows that there seems to be a cross-side effect of clustering of trades-through:

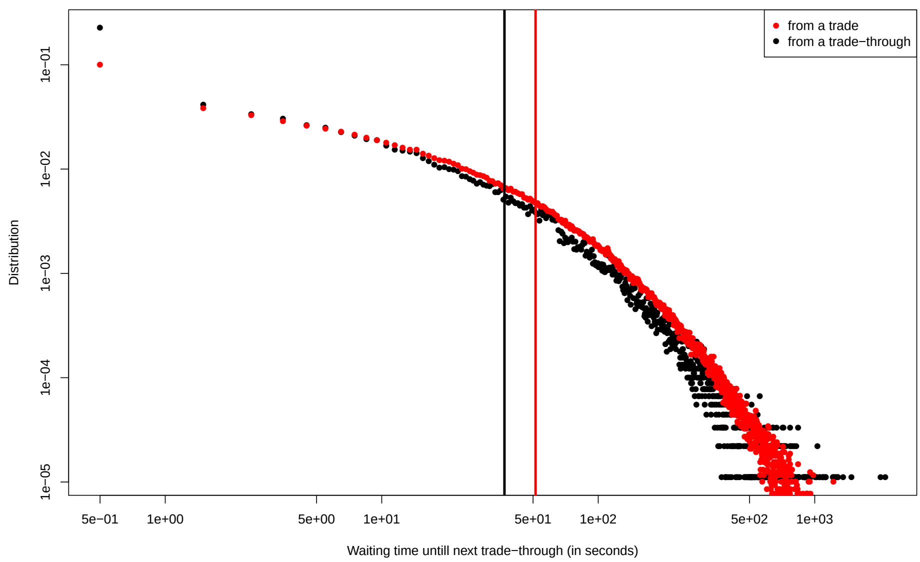

Figure 2: Global trades-through clustering for BNP Paribas: Empirical distribution of the time intervals between a trade or a trade-through and the next trade-through.

a trade-through at one side of the book will be more closely followed in time by a trade-through on the other side of the book. But comparing the relative difference between mean waiting times of $(\lambda^+) \to (\lambda^+)$ and $(\Lambda^+) \to (\lambda^+)$, we have approximately a 30% decrease on the same side of the book. Whereas there is only a 20% decrease of mean waiting time between $(\lambda^+) \to (\lambda^-)$ and $(\Lambda^+) \to (\lambda^-)$, which reflects that cross-side clustering effect is weaker than same side clustering for trades-through.

Figure 2 plots the distributions of waiting times $(\lambda^+ + \lambda^-) \to (\lambda^+ + \lambda^-)$ and $(\Lambda^+ + \Lambda^-) \to (\Lambda^+ + \Lambda^-)$ studied in this paragraph.

In brief, looking at these distributions of durations gives us global tendencies on clustering and relative comparisons of the influences of trades-through with respect

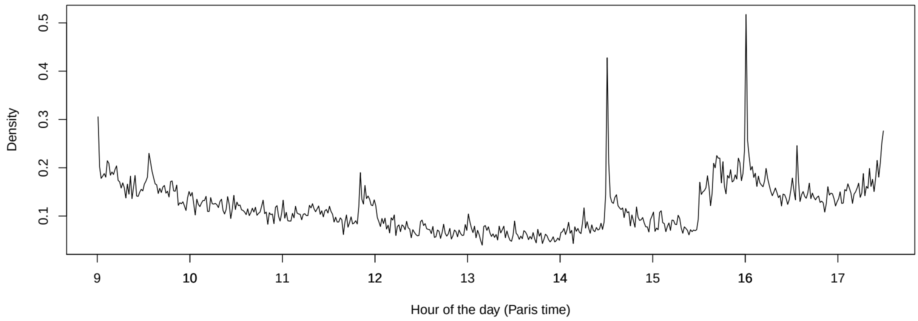

Figure 3: Intraday distribution of timestamps of trades-through for the stock BNP Paribas on June–October 2010, using one-minute bins.

to limit order book sides. A more quantitative measurement of those effects will be done in the following part using the analysis of calibrated parameters of a fitted stochastic model, namely Hawkes processes.

2.6 Intraday timestamp distribution

We now look at the intraday distribution of timestamps for second-limit trades-through on the BNP Paribas stock, plotted on Figure 3. We can see that the distribution is globally the sum of two parts: a U-shape curve (linked to the global U-shape trading activity curve) and two peaks at very precise hours (2:30 pm and 4:00 pm – Paris time) reflecting the impact of major macro-economic news released at that moment of the day.

What is important for further modelling is to notice that it seems very difficult to find a pure stochastic model able to capture both the local behaviour and fluctuations of trades-through arrival times and the two big peaks at very precise hours of the day. A first attempt may be to simply remove those peaks in the distribution. In the remaining of the paper, we will restrict ourselves to a two-hour interval, thus removing major seasonality effects.

3 Hawkes processes

Let us first recall standard definitions and properties of Hawkes processes. Hawkes processes are self-exciting processes which were introduced by Hawkes (1971). Roughly speaking, in a self-exciting process, the occurrence of an event increases the probability of occurence of another event. Hawkes (1971) specifically studies the case in which this increase is instantaneous and exponentially decaying, case which we use in this paper. This section only gives results necessary to the remaining of the paper. Many references may provide more details : Bremaud (1981) and Daley and Vere-Jones (2003) are well-known general textbook treatments of point processes ; more precisely in our case, Bowsher (2007) provides a framework for "Generalized Hawkes models" which fully encompasses the processes used in this study.

In the following, we will note $\overline{\mathbb{R}}_+^* = \mathbb{R}_+^* \cup \{+\infty\}$ and $\overline{\mathbb{N}} = \mathbb{N} \cup \{+\infty\}$.

3.1 Definition

Let $M \in \mathbb{N}^*$. We start by defining a M-variate simple point process. Let $\{T_i, Z_i\}_{i\in\mathbb{N}^*}$ be a double sequence of random variables on some probability space $(\Omega, \mathscr{F}, \mathbf{P})$. $\{T_i\}_{i\in\mathbb{N}^*}$ is $\overline{\mathbb{R}}_+^*$-valued, (i) almost surely increasing, (ii) such that $T_i < T_{i+1}$ almost surely on $\{T_i < +\infty\}$, and (iii) such that $\lim_i T_i = +\infty$ almost surely. For all i, T_i represents the time of occurence of the i-th event. $\{Z_i\}_{i\in\mathbb{N}^*}$ is $\{1,\dots,M\}$-valued. For all i, Z_i represents the type of the i-th event. For all $m = 1,\dots,M$, let us define $N^m(t) = \sum_{i\in\mathbb{N}^*} \mathbf{1}_{\{T_i\le t\}}\mathbf{1}_{\{Z_i=m\}}$ the counting process associated to events of type m. $N = (N^1,\dots,N^M)$ is the M-dimensional vector of counting processes. N (or equivalently $\{T_i, Z_i\}_{i\in\mathbb{N}^*}$) is called a M-variate simple point process.

Let us now precise the notion of (conditional) intensity. Let $\{\mathscr{F}_t\}_{t\in\mathbb{R}_+}$ be a filtration such that for all $t \in \mathbb{R}_+$, $\mathscr{F}_t \supset \mathscr{F}_t^N$, where $\mathscr{F}_t^N$ is the natural filtration of N. For all $m = 1,\dots,M$, a $\mathscr{F}_t$-intensity of the monovariate simple point process N^m is a nonnegative $\mathscr{F}_t$-progressively measurable process $\{\lambda^m(t)\}_{t\in\mathbb{R}_+}$ such that

$$\forall 0 \le s \le t,\ \mathbf{E}\left[N^m(t) - N^m(s)) | \mathscr{F}_s\right] = \mathbf{E}\left[\int_s^t \lambda^m(u)du \middle| \mathscr{F}_s\right] \ a.s. \tag{1}$$

With sufficient regularity conditions, one can derive that :

$$\lim_{t \to s, t > s} \frac{1}{t-s} \mathbf{E}\left[N^m(t) - N^m(s)) | \mathscr{F}_s\right] = \lambda^m(s) \ \ a.s., \tag{2}$$

which is often taken as a definition for λ^m (see e.g. Hautsch (2004, Definition 2.1)). The intensity λ^m can thus be interpreted as the conditional probability to observe an event of type m per unit of time.

In this paper, we're interested in the special case of Hawkes processes with an exponentially-decaying kernel, i.e. processes for which the m-th coordinate N^m, $m \in \{1, \dots, M\}$, admits an $\mathscr{F}_t^N$-intensity of the form

$$\lambda^m(t) = \lambda_0^m(t) + \sum_{n=1}^{M} \int_0^t \sum_{j=1}^{P} \alpha_j^{mn} e^{-\beta_j^{mn}(t-s)} dN_s^n, \tag{3}$$

where $\lambda_0^m : \mathbb{R}_+ \to \mathbb{R}_+$ is a deterministic function, the number P of exponential kernels is a fixed integer, and for all $m, n = 1, \dots, M$, and $j = 1, \dots, P$, α_j^{mn} and β_j^{mn} are positive constants.

In its simplest version with $P = 1$ and $\lambda_0^m(t) = \lambda_0^m$ a positive constant, the definition becomes for all $t \in \mathbb{R}_+$:

$$\lambda^m(t) = \lambda_0^m + \sum_{n=1}^{M} \int_0^t \alpha^{mn} e^{-\beta^{mn}(t-s)} dN_s^n, \tag{4}$$

Parameters α^{mn} and β^{mn} express the influence (scale and decay) of the past events of type n on the m-th coordinate of the process. It follows from this definition that two phenomena are present: self-excitation ($m = n$) and mutual excitation ($m \neq n$).

3.2 Stationarity condition

Taking here $P = 1$ and rewriting equation (4) using vector notations $\lambda = (\lambda^1, \dots, \lambda^M)$, we have:

$$\lambda(t) = \lambda_0 + \int_0^t G(t-s) dN_s, \tag{5}$$

where G is a matrix-valued function of time defined by

$$G(t) = \left(\alpha^{mn} e^{-\beta^{mn}(t)}\right)_{m,n=1,\ldots,M}. \tag{6}$$

A multivariate point process is stationary if the joint distribution of any number of types of events on any number of given intervals is invariant under translation, i.e. if, forall integer r, the distribution of $\{N^{i_1}(t_1+t) - N^{i_1}(s_1+t), \ldots, N^{i_r}(t_r+t) - N^{i_r}(s_r+t)\}$ for $i_j \in \{1,\ldots,M\}, s_j \leq t_j, j = 1,\ldots,r$ does not depend on $t \in \mathbb{R}$. This of course cannot be the case if λ_0 is not constant. In the financial models of the following sections, such an assumption translates the fact that the level of market activity for all the events considered does not dramatically changes in the period considered, which may be far from obvious, as discussed in section 4.2.

In the monovariate case $M = 1$, Hawkes and Oakes (1974, Theorem 1) show that a sufficient condition for the existence of stationary point process with intensity as in equation (5) is that $\int_0^{+\infty} G(u)du < 1$. In our special case of exponential kernels, this gives the sufficient stationarity condition $\frac{\alpha^{11}}{\beta^{11}} < 1$.

Bremaud and Massoulie (1996, Theorem 7) generalizes the result to the multivariate case (and even much further to the non-linear case, which we don't need here). The condition becomes that the spectral radius of the matrix

$$\Gamma = \int_0^{\infty} G(u)du = \left(\frac{\alpha^{mn}}{\beta^{mn}}\right)_{m,n=1,\ldots,M} \tag{7}$$

is strictly smaller than 1. We recall that the spectral radius of the matrix G is defined as $\rho(G) = \max_{a\in\mathscr{S}(G)} |a|$, where $\mathscr{S}(G)$ denotes the set of all eigenvalues of G. In the two-dimensional case that will be used in the following sections, this can be written:

$$\frac{1}{2}\left(\frac{\alpha^{11}}{\beta^{11}} + \frac{\alpha^{22}}{\beta^{22}} + \sqrt{\left(\frac{\alpha^{11}}{\beta^{11}} - \frac{\alpha^{22}}{\beta^{22}}\right)^2 + 4\frac{\alpha^{12}}{\beta^{12}}\frac{\alpha^{21}}{\beta^{21}}}\right) < 1. \tag{8}$$

3.3 Maximum-likelihood estimation

In this section, we introduce the notation $\{T_i^m\}_{i\in\mathbb{N}^*}$ which is the re-enumerated sequence of events of type m. $\{T_i, Z_i\}_{i=1,\ldots,N}$ is thus the ordered pool of all

$\{\{T_i^m, m\}_{m=1,\dots,M}\}$. The log-likelihood of a multidimensional Hawkes process can be computed as the sum of the likelihood of each coordinate, and is thus written:

$$\ln \mathscr{L}(\{N(t)\}_{t\leq T}) = \sum_{m=1}^{M} \ln \mathscr{L}^m(\{N^m(t)\}_{t\leq T}), \tag{9}$$

where each term is defined by:

$$\ln \mathscr{L}^m(\{N^m(t)\}_{t\leq T}) = \int_0^T (1-\lambda^m(s))\,ds + \int_0^T \ln \lambda^m(s) dN^m(s). \tag{10}$$

This partial log-likelihood can be computed as:

$$\begin{aligned}\ln \mathscr{L}^m(\{N^m(t)\}_{t\leq T}) &= T - \Lambda^m(0,T) \\ &+ \sum_{i:T_i\leq T} \mathbf{1}_{\{Z_i=m\}} \ln \left[\lambda_0^m(T_i) + \sum_{n=1}^{M}\sum_{j=1}^{P}\sum_{T_k^n<T_i} \alpha_j^{mn} e^{-\beta_j^{mn}(T_i-T_k^n)}\right]\end{aligned} \tag{11}$$

where $\Lambda^m(0,T) = \int_0^T \lambda^m(s)ds$ is the integrated intensity. Following Ozaki (1979), we compute this in a recursive way by observing that, thanks to the exponential form of the kernel:

$$\begin{aligned}R_j^{mn}(l) &= \sum_{T_k^n<T_l^m} e^{-\beta_j^{mn}(T_l^m-T_k^n)} \\ &= \begin{cases} e^{-\beta_j^{mn}(T_l^m-T_{l-1}^m)} R_j^{mn}(l-1) + \sum_{T_{l-1}^m\leq T_k^n<T_l^m} e^{-\beta_j^{mn}(T_l^m-T_k^n)} & \text{if } m\neq n, \\ e^{-\beta_j^{mn}(T_l^m-T_{l-1}^m)}\left(1+R_j^{mn}(l-1)\right) & \text{if } m=n.\end{cases}\end{aligned} \tag{12}$$

The final expression of the partial log-likelihood may thus be written:

$$\begin{aligned}\ln \mathscr{L}^m(\{N^m(t)\}_{t\leq T}) = T &- \int_0^T \lambda_0^m(s)ds \\ &- \sum_{i:T_i\leq T}\sum_{n=1}^{M}\sum_{j=1}^{P} \frac{\alpha_j^{mn}}{\beta_j^{mn}}\left(1-e^{-\beta_j^{mn}(T-T_i)}\right) \\ &+ \sum_{l:T_l^m\leq T} \ln\left[\lambda_0^m(T_l^m) + \sum_{n=1}^{M}\sum_{j=1}^{P} \alpha_j^{mn} R_j^{mn}(l)\right],\end{aligned} \tag{13}$$

where $R_j^{mn}(l)$ is defined with equation (12) and $R_j^{mn}(0) = 0$.

3.4 Testing the calibration

A general result on point processes theory states that a given non-Poisson process can be transformed into a homogeneous Poisson process by a stochastic time change. A standard monovariate version of this result and its proof can be found in Bremaud (1981, Chapter II, Theorem T16). Bowsher (2007) has shown that this can be generalized in a multidimensional setting, which provides specification tests for multidimensional Hawkes models. We reproduce here its result, with slightly modified notations to accommodate the notations chosen here.

Theorem (Bowsher (2007, Theorem 4.1))**.** *Let N be a M-variate point process on $\mathbb{R}_+^*$ with internal history*[1] $\{\mathscr{F}_t^N\}_{t\in\mathbb{R}_+}$*, and $M \geq 1$. Also let $\{\mathscr{F}_t\}_{t\in\mathbb{R}_+}$ be a history of N (that is, $\mathscr{F}_t^N \subseteq \mathscr{F}_t, \forall t \geq 0$), and suppose, for each m, that N^m has the $\mathscr{F}_t$-intensity λ^m where λ^m satisfies $\int_0^\infty \lambda^m(s)ds = \infty$ almost surely. Define for each m and all $t \geq 0$ the $\mathscr{F}_t$-stopping time $\tau^m(t)$ as the (unique) solution to*

$$\int_0^{\tau^m(t)} \lambda^m(u)du = t. \tag{14}$$

Then the M point processes $\{\tilde{N}^m\}_{m=1,\ldots,M}$ defined by $\tilde{N}^m(t) = N^m(\tau^m(t))$, $\forall t \geq 0$, are independent Poisson processes with unit intensity. Furthermore, the durations of each Poisson process $\tilde{N}^m$ are given by

$$\Lambda^m(T_{i-1}^m, T_i^m) = \int_{T_{i-1}^m}^{T_i^m} \lambda^m(s)ds, \ \ \forall i \geq 2. \tag{15}$$

Let us compute this integrated intensity of the m-th coordinate of a multidimensional Hawkes process between two consecutive events T_{i-1}^m and T_i^m of type m:

[1] Which we previously called *natural filtration*.

$$\begin{aligned}\Lambda^m(T_{i-1}^m, T_i^m) = & \int_{T_{i-1}^m}^{T_i^m} \lambda_0^m(s)ds \\ & + \sum_{n=1}^{M}\sum_{j=1}^{P}\sum_{T_k^n < T_{i-1}^m} \frac{\alpha_j^{mn}}{\beta_j^{mn}} \left[e^{-\beta_j^{mn}(T_{i-1}^m - T_k^n)} - e^{-\beta_j^{mn}(T_i^m - T_k^n)} \right] \\ & + \sum_{n=1}^{M}\sum_{j=1}^{P}\sum_{T_{i-1}^m \le T_k^n < T_i^m} \frac{\alpha_j^{mn}}{\beta_j^{mn}} \left[1 - e^{-\beta_j^{mn}(T_i^m - T_k^n)} \right].\end{aligned} \tag{16}$$

As in the log-likelihood computation, following Ozaki (1979), we observe that:

$$\begin{aligned} A_j^{mn}(i-1) &= \sum_{T_k^n < T_{i-1}^m} e^{-\beta_j^{mn}(T_{i-1}^m - T_k^n)} \\ &= e^{-\beta_j^{mn}(T_{i-1}^m - T_{i-2}^m)} A_j^{mn}(i-2) + \sum_{T_{i-2}^m \le T_k^n < T_{i-1}^m} e^{-\beta_j^{mn}(T_{i-1}^m - T_k^n)}, \end{aligned} \tag{17}$$

so that the integrated density can be written for all $i \geq 2$:

$$\begin{aligned}\Lambda^m(T_{i-1}^m, T_i^m) = & \int_{T_{i-1}^m}^{T_i^m} \lambda_0^m(s)ds + \sum_{n=1}^{M}\sum_{j=1}^{P} \frac{\alpha_j^{mn}}{\beta_j^{mn}} \Bigg[A_j^{mn}(i-1) \\ & \times \left(1 - e^{-\beta_j^{mn}(T_i^m - T_{i-1}^m)}\right) + \sum_{T_{i-1}^m \le T_k^n < T_i^m} \left(1 - e^{-\beta_j^{mn}(T_i^m - T_k^n)}\right) \Bigg],\end{aligned} \tag{18}$$

where A_j^{mn} is defined as in equation (17) with for all $j = 1, \ldots, P, A_j^{mn}(0) = 0$.

Hence, simply following the method in Bowsher (2007), we can easily define tests to check the goodness-of-fit of a Hawkes model to our empirical data. Since the integrated intensity $\Lambda^m(T_{i-1}^m, T_i^m)$ is a time interval of a homogeneous Poisson Process, we can test for each $m = 1, \ldots, M$: (i) whether the variables $\left(\Lambda^m(T_{i-1}^m, T_i^m)\right)_{i\geq 2}$ are exponentially distributed ; (ii) whether the variables $\left((\Lambda^m(T_{i-1}^m, T_i^m)\right)_{i\geq 2}$ are independent. In Section 4.3, the independence test will be carried out with a Ljung-Box test up to the twentieth term, and we will use a standard Kolmogorov-Smirnov test the empirical data against the exponential distribution.

In this section, we have recalled results with detailed computations that we will now use to build and estimate models for trades-through. Several reasons make this approach attractive. Studying point processes taking the stochastic intensity as the main tool allows to define straightforwardly multivariate models, while multidimensional generalizations of duration-based models for example may be cumbersome, and at least quite artificial (Daley and Vere-Jones, 2003: 41). Furthermore, as for the special case of linear Hawkes processes with exponentially-decaying kernel, the results recalled in this section show that interpretation of the few parameters (base intensity as background market activity, jumps and exponential decay of the influence of past events), statistical estimation and specification tests are straightforward. Note also that simulation algorithms of Hawkes processes are easily available (see e.g. Ogata, 1981; Moller and Rasmussen, 2005, and Appendix A), allowing further developments for the use of the estimated models. These practical properties make Hawkes processes good candidates for the study of the clustering of events empirically observed in order book models (Hautsch, 2004; Bowsher, 2007; Large, 2007; Bacry et al., 2012).

We now turn to the modelling of trades-through in an order book model.

4 A simple Hawkes model for trades-through

4.1 Model

Since empirical evidence shows that trades-through obviously occur in a clustered way, it makes sense to try to model them with self-exciting Hawkes processes. We thus define our basic model as follows. Let $\left(T_i^A\right)_{i\geq 1}$ be the point process of trades-through occurring on the ask side of the limit order book, and $\left(T_i^B\right)_{i\geq 1}$ be the point process of trades-through occurring on the bid side. Let N^A and N^B denote the associated counting processes. These two processes are assumed to form a two-dimensional Hawkes process with intensities λ^A and λ^B defined with parameters $(\alpha_{ij}, \beta_{ij})_{(i,j)\in\{A,B\}^2}$ as follows:

$$\begin{aligned} \lambda^A(t) &= \lambda_0^A(t) + \int_0^t \alpha_{AA} e^{-\beta_{AA}(t-s)} dN_s^A + \int_0^t \alpha_{AB} e^{-\beta_{AB}(t-s)} dN_s^B, \\ \lambda^B(t) &= \lambda_0^B(t) + \int_0^t \alpha_{BA} e^{-\beta_{BA}(t-s)} dN_s^A + \int_0^t \alpha_{BB} e^{-\beta_{BB}(t-s)} dN_s^B. \end{aligned} \tag{19}$$

This is a standard bivariate Hawkes model of Section 3 with $P = 1$.

4.2 Calibration

We compute the maximum-likelihood estimates for the parameters of our model, using computations presented in Section 3, and the empirical data previoulsy described in Section 2.3. In a first step, we make the assumptions that base intensities λ_0^A and λ_0^B are constants. Taking into account the huge variations of trading activity during the day, we restrict our empirical observations to a two-hour interval, from 9:30 am to 11:30 am. Considering Figure 3, this may make a bit more realistic the stationarity assumption discussed in Section 3.2. During these five months of trading, we count in average 2737 trades each day during this time interval, 206 of which are trades-through (100 on the ask side and 106 on the bid side). Thus roughly 8% of the recorded transactions are trades-through.

Tables 3 and 4 summarize the statistics on the estimated values on the ask and bid sides. These tables show that the median half-lives associated to the kernels AA, AB, BA and BB are respectively 145, 521, 474 and 35 milliseconds. It appears that we observe very large variations in the results of the numerical maximization of the likelihood. However, whatever the absolute size of the parameters, it is clear that the cross-excitation effect, i.e. the excitation of trades-through of a given side by the occurrence of trades-through on the opposite side, is much weaker than the self-excitation effect, which translates the clustering of trades-through on a given side. The average value of α_{AB} is 9.5 times smaller that the average value of α_{AA}, while at the same time the associated exponential decay β_{AB} is only 1.5 times smaller than the average β_{AA}. The instantaneous effect is thus much smaller while its half-life is not significantly longer. This observation is also valid for the average α_{BA} which is 13 times smaller than the average α_{BB}, while the average exponentials decays differ only by a factor 4.

	Average	Median	Min	Max	Stdev
λ_0^A	1.01E-02	8.42E-03	6.62E-06	3.52E-02	6.27E-03
α_{AA}	4.13E+00	6.45E-01	3.53E-02	3.09E+01	6.03E+00
α_{AB}	4.33E-01	1.05E-01	1.00E-10	4.78E+00	8.41E-01
β_{AA}	3.70E+01	4.78E+00	1.84E-01	2.34E+02	5.21E+01
β_{AB}	2.48E+01	1.33E+00	1.00E-10	1.48E+03	1.44E+02

Table 3: Statistics summary for the maximum-likelihood estimates of the ask side of model (19).

	Average	Median	Min	Max	Stdev
λ_0^B	1.09E-02	9.08E-03	2.46E-03	3.98E-02	6.35E-03
α_{BA}	3.68E-01	7.56E-02	3.83E-13	4.46E+00	6.77E-01
α_{BB}	4.81E+00	3.04E+00	2.08E-02	4.62E+01	7.00E+00
β_{BA}	9.61E+00	1.46E+00	1.00E-10	1.00E+02	1.92E+01
β_{BB}	3.98E+01	2.00E+01	3.71E-02	3.75E+02	5.52E+01

Table 4: Statistics summary for the maximum-likelihood estimates of the bid side of model (19).

In other words, the ratio $\frac{\alpha}{\beta}$, which is equal to the total integrated intensity of an exponential kernel $\int_0^{+\infty} \alpha e^{-\beta u} du$, is much weaker in the cross-excitation cases (taking the average values, $\frac{\alpha_{AB}}{\beta_{AB}} = 0.017$, $\frac{\alpha_{BA}}{\beta_{BA}} = 0.038$) than in the self-excitation cases (still using the average values, $\frac{\alpha_{AA}}{\beta_{AA}} = 0.111$, $\frac{\alpha_{BB}}{\beta_{BB}} = 0.120$). Therefore, we can focus on the calibration and use of a simpler model, where trades-through are modelled by two one-dimensional Hawkes processes, with no cross-excitation:

$$\begin{aligned} \lambda^A(t) &= \lambda_0^A(t) + \int_0^t \alpha_{AA} e^{-\beta_{AA}(t-s)} dN_s^A, \\ \lambda^B(t) &= \lambda_0^B(t) + \int_0^t \alpha_{BB} e^{-\beta_{BB}(t-s)} dN_s^B. \end{aligned} \tag{20}$$

Table 5 summarizes the statistics of the estimated values of this simplified model with the assumption λ_0^A and λ_0^B constant. Values are similar to the previous case, confirming that the cross-effects were negligible. The effect of this simplification will be further discussed with the goodness-of-fit tests.

Finally, in an attempt to grasp small variations of activity independent of the clustering of the trades-through, following ideas presented in Bowsher (2007), we test a third version of the model by getting rid of the assumptions stating that λ_0^A and

	Average	Median	Min	Max	Stdev
λ_0^A	1.18E-02	9.77E-03	3.47E-03	3.66E-02	6.53E-03
α_{AA}	6.02E+00	4.85E+00	1.00E-10	3.09E+01	6.14E+00
β_{AA}	4.76E+01	3.63E+01	3.92E-02	2.34E+02	4.87E+01
λ_0^B	1.26E-02	1.13E-02	3.80E-03	4.25E-02	7.00E-03
α_{BB}	8.05E+00	5.61E+00	1.28E-02	4.79E+01	8.81E+00
β_{BB}	6.64E+01	4.72E+01	2.64E-02	3.91E+02	7.22E+01

Table 5: Statistics summary for the maximum-likelihood estimates of the simplified model (20) with λ_0^A and λ_0^B constant.

λ_0^B are constants. In this version of the simplified model (20), base intensities $\lambda_0^A(t)$ and $\lambda_0^B(t)$ are piecewise-linear continuous functions on the subdivision $(9:30 < 10:00 < 10:30 < 11:00 < 11:30)$ of the time interval $[9:30\text{ am}; 11:30\text{ am}]$. Note that this assumption implies that the process is not stationary anymore. Tables 6 and 7 summarize the statistics on the estimated values on the ask and bid sides. Let us now discuss the goodness-of-fit of these three calibrations.

4.3 Goodness-of-fit

For each trading days, we have extracted the time series $\left(t_i^A\right)_{i\geq 1}$ and $\left(t_i^B\right)_{i\geq 1}$. For each of the three models discussed above and for each trading day, we can compute the integrated intensities $\left(\Lambda^A(t_i^A, t_{i+1}^A)\right)_{i\geq 1}$ and $\left(\Lambda^B(t_i^B, t_{i+1}^B)\right)_{i\geq 1}$ defined as in (18) and perform the four tests of goodness-of-fit described in Section 3. This gives us four tests per model and per trading day. Table 8 shows the results of the tests for a risk of the first kind equal to 1% and 2.5%. These results confirm that the cross-excitation of trades-through of one side of the book on the other side is weak. In the case where λ_0 is constant, the percentage of trading days where the model passes all 4 statistical tests is 76% in the full specification case, and stays at 71% when cross-excitation is not taken into account. And in the case where λ_0 is allowed to vary as a piecewise-linear continuous function, these two percentages are even equal: in this latter case, we don't have any statistical improvement by including the cross-excitation effect.

Moreover, these results show that adding more flexibility in the modelling of λ_0 using piecewise-linear continuous functions helps the model to grasp the dynamics

	Average	Median	Min	Max	Stdev
$\lambda_0^A(9:30)$	1.94E-02	1.65E-02	1.00E-20	5.40E-02	1.29E-02
$\lambda_0^A(10:00)$	1.13E-02	9.81E-03	1.00E-20	3.72E-02	9.73E-03
$\lambda_0^A(10:30)$	1.33E-02	1.25E-02	1.00E-20	5.01E-02	9.30E-03
$\lambda_0^A(11:00)$	7.67E-03	4.97E-03	1.00E-20	5.45E-02	9.93E-03
$\lambda_0^A(11:30)$	1.32E-02	1.03E-02	1.00E-20	1.54E-01	1.68E-02
α_{AA}	6.62E+00	5.10E+00	3.64E-13	3.09E+01	6.25E+00
β_{AA}	5.64E+01	4.61E+01	1.00E-20	2.34E+02	5.14E+01

Table 6: Statistics summary for the maximum-likelihood estimates of the ask side of model (20) with λ_0^A and λ_0^B piecewise-linear continuous functions.

	Average	Median	Min	Max	Stdev
$\lambda_0^B(9:30)$	1.99E-02	1.67E-02	1.35E-03	6.51E-02	1.31E-02
$\lambda_0^B(10:00)$	1.25E-02	1.06E-02	1.00E-20	5.38E-02	1.00E-02
$\lambda_0^B(10:30)$	1.26E-02	1.14E-02	1.00E-20	5.15E-02	9.18E-03
$\lambda_0^B(11:00)$	9.32E-03	7.77E-03	1.00E-20	5.65E-02	9.65E-03
$\lambda_0^B(11:30)$	1.33E-02	9.16E-03	7.06E-14	1.26E-01	1.53E-02
α_{BB}	8.20E+00	5.60E+00	8.70E-04	4.79E+01	8.68E+00
β_{BB}	6.82E+01	5.15E+01	1.25E-03	3.91E+02	6.98E+01

Table 7: Statistics summary for the maximum-likelihood estimates of the bid side of model (20) with λ_0^A and λ_0^B piecewise-linear continuous functions.

Model	Performance	2.5%		1%	
Full Model (19) with λ_0 constant	4 passed	70	(64.2)	83	(76.1)
	3 passed	29	(26.6)	26	(23.9)
	2 or less	10	(9.2)	0	(0.0)
No Cross (20) with λ_0 constant	4 passed	59	(54.1)	77	(70.6)
	3 passed	35	(32.1)	25	(22.9)
	2 or less	15	(13.8)	7	(6.4)
Full Model (19) with λ_0 piecewise-linear	4 passed	84	(77.1)	94	(86.2)
	3 passed	20	(18.3)	13	(11.9)
	2 or less	5	(4.6)	2	(1.8)
No Cross (20) with λ_0 piecewise-linear	4 passed	83	(76.1)	95	(87.2)
	3 passed	20	(18.3)	14	(12.8)
	2 or less	6	(5.5)	0	(0.0)

Table 8: Performance of the calibration of the Hawkes models. For each model, this table gives the number of trading days (out of 109) where 4, 3, or 2 or less tests out of for where successfully passed. The four tests are two independence Ljung-Box tests and two Kolmogorov-Smirnov tests for the exponential distribution. Values in parentheses are percentages.

of trades-through: all tests are passed in more that 87% of the trading days tested in both cases.

5 Conclusion

We have studied in this paper a model for trades-through based on Hawkes processes. We have shown that the clustering properties of trades-through can be well modelled with such self-exciting processes. Although calibration results may vary a lot from trading day to trading day, general patterns remain, such as the weak cross-excitation effects. The clustering of trades-through highlightened here may not be a surprise for a market practictioner, but our results may help describing periods of high liquidity consumption in order book models. Our observations are related to the dynamics of aggressive market orders described e.g. in Large (2007). Note though that the set of aggressive market orders, usually defined as the set of market orders that move the best price, contains, but is not equal to, our set of trades-through defined in Section 2.2. The link between order flows and price changes is of prime interest when modelling order books, and aggressive market orders are essential to a good understanding of the market impact (see e.g. Eisler et al., 2012). Our empirical study of the clustering of trades-through hopefully contribute to that line of work. Both the nature of our observations and the simplicity of the models could lead to future practical work on trades-through-based trading strategies and order book modelling.

Appendix A – Simulation of Hawkes processes

Ogata (1981) proposes an algorithm for the simulation of Hawkes processes, based on a general procedure called thinning. Taking a different point of view, a more recent work by Moller and Rasmussen (2005) claims to provide better quality simulations, without edge effects. In this appendix however, we describe the simple thinning algorithm in a multidimensional setting, because of its intuitive progression and its ability to be straightforwardly generalized to any point process described with its conditional intensity.

Let $\mathscr{U}_{[0,1]}$ denote the uniform distribution on the interval $[0,1]$ and $[0,T]$ the time interval on which the Hawkes process defined by equation (4) is to be simulated. We define $I^K(t) = \sum_{n=1}^{K} \lambda^n(t)$ the sum of the intensities of the first K components of the multivariate process. The algorithm is written as follows.

1. **Initialization**: Set $i \leftarrow 1$, $i^1 \leftarrow 1, \dots, i^M \leftarrow 1$ and $I^* \leftarrow I^M(0) = \sum_{n=1}^{M} \lambda_0^n$.

2. **First event**: Generate $U \rightsquigarrow \mathscr{U}_{[0,1]}$ and set $s \leftarrow -\frac{1}{I^*} \ln U$ (s is exponentially distributed with parameter I^*).

   (a) **If** $s > T$ **Then** go to step 4.

   (b) **Attribution Test**: Generate $D \rightsquigarrow \mathscr{U}_{[0,1]}$ and set $t_1^{n_0} \leftarrow s$ where n_0 is such that $\frac{I^{n_0-1}(0)}{I^*} < D \leq \frac{I^{n_0}(0)}{I^*}$.

   (c) Set $t_1 \leftarrow t_1^{n_0}$.

3. **General routine**: Set $i^{n_0} \leftarrow i^{n_0} + 1$ and $i \leftarrow i + 1$.

   (a) **Update maximum intensity**: Set $I^* \leftarrow I^M(t_{i-1}) + \sum_{n=1}^{M} \sum_{j=1}^{P} \alpha_j^{nn_0}$. I^* exhibits a jump of size $\sum_{n=1}^{M} \sum_{j=1}^{P} \alpha_j^{nn_0}$ as an event of type n_0 has just occurred.

   (b) **New event**: Generate $U \rightsquigarrow \mathscr{U}_{[0,1]}$ and set $s \leftarrow s - \frac{1}{I^*} \ln U$.
   If $s > T$,
   Then go to step 4.

   (c) **Attribution-Rejection test**: Generate $D \rightsquigarrow \mathscr{U}_{[0,1]}$.
   If $D \leq \frac{I^M(s)}{I^*}$,
   Then set $t_{i^{n_0}}^{n_0} \leftarrow s$ where n_0 is such that $\frac{I^{n_0-1}(s)}{I^*} < D \leq \frac{I^{n_0}(s)}{I^*}$, and $t_i \leftarrow t_{i^{n_0}}^{n_0}$ and go through the general routine again,
   Else update $I^* \leftarrow I^M(s)$ and try a new date at step (b) of the general routine.

4. **Output**: Retrieve the simulated process $\{\{t_i^n\}_i\}_{n=1,\ldots,M}$.

Of course, still using using notations introduced in Section 3, one can simulate the double sequence $\{t_i, z_i\}_i$ instead of the M sequences $\{t_i^n\}_i, n = 1,\ldots,M$ with very minor and straightforward modifications.

Acknowledgements: The authors wish to thank their colleagues of the Chair of Quantitative Finance at Ecole Centrale Paris and two anonymous referees for their comments and suggestions, leading to significant improvements of the manuscript.

References

Bacry, E., Delattre, S., Hoffmann, M., and Muzy, J. (2012). Modeling microstructure noise with mutually exciting point processes. *Quantitative Finance*. URL http://www.tandfonline.com/doi/abs/10.1080/14697688.2011.647054. Forthcoming.

Bouchaud, J.-P., Farmer, J. D., and Lillo, F. (2009). How markets slowly digest changes in supply and demand. In T. Hens, and K. R. Schenk-Hoppe (Eds.), *Handbook of financial markets: dynamics and evolution*, pages 57–160. Elsevier, North-Holland.

Bowsher, C. (2007). Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141(2): 876–912. URL http://ideas.repec.org/a/eee/econom/v141y2007i2p876-912.html.

Bremaud, P. (1981). *Point processes and queues, martingale dynamics*. Springer-Verlag New York Inc.

Bremaud, P., and Massoulie, L. (1996). Stability of nonlinear Hawkes processes. *The Annals of Probability*, 24(3): 1563–1588. ISSN 00911798. URL http://www.jstor.org/stable/2244985.

Daley, D., and Vere-Jones, D. (2003). *An introduction to the theory of point processes*, volume I : Elementary Theory and Methods. Springer, 2nd edition.

Eisler, Z., Bouchaud, J.-P., and Kockelkoren, J. (2012). The price impact of order book events: market orders, limit orders and cancellations. *Quantitative Finance*. URL http://arxiv.org/abs/0904.0900. Forthcoming.

Hautsch, N. (2004). *Modelling irregularly spaced financial data: Theory & practice of dynamic duration models*. Springer-Verlag Berlin and Heidelberg.

Hawkes, A. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1): 83 –90. URL http://www.jstor.org/stable/2334319.

Hawkes, A., and Oakes, D. (1974). A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3): 493–503. URL http://www.jstor.org/stable/10.2307/3212693.

Large, J. (2007). Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, 10(1): 1–25. URL http://ideas.repec.org/a/eee/finmar/v10y2007i1p1-25.html.

Lillo, F., and Farmer, J. (2004). The long memory of the efficient market. *Studies in Nonlinear Dynamics & Econometrics*, 8(3): 1. URL http://ideas.repec.org/a/bpj/sndecm/v8y2004i3n1.html.

Moller, J., and Rasmussen, J. (2005). Perfect simulation of Hawkes processes. *Advances in Applied Probability*, 37(3): 629–646. ISSN 00018678. URL http://www.jstor.org/stable/30037347.

Muni Toke, I. (2011). "Market making" in an order book model and its impact on the bid-ask spread. In F. Abergel, B. Chakrabarti, A. Chakraborti, and M. Mitra (Eds.), *Econophysics of Order-Driven Markets*, New Economic Windows, pages 49–64. Springer-Verlag Milan. URL http://www.springerlink.com/content/x811111w104v5hk3/.

Ogata, Y. (1981). On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, 27(1): 23–31. URL http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=01056305.

Ozaki, T. (1979). Maximum likelihood estimation of Hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics*, 31(1): 145–155. URL http://www.springerlink.com/content/hr3q7667x3522235/.

Pomponio, F., and Abergel, F. (2010). Trade-throughs: Empirical facts – Application to lead-lag measures. URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1694103.