Estimation of slowly decreasing Hawkes kernels: Application to high frequency order book modelling

Emmanuel Bacry[1], Thibault Jaisson[1], and Jean-François Muzy[1,2]

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

Abstract

We present a modified version of the non parametric Hawkes kernel estimation procedure studied in [5] that is adapted to slowly decreasing kernels. We show on numerical simulations involving a reasonable number of events that this method allows us to estimate faithfully a power-law decreasing kernel over at least 6 decades. We then propose a 8-dimensional Hawkes model for all events associated with the first level of some asset order book. Applying our estimation procedure to this model, allows us to uncover the main properties of the coupled dynamics of trade, limit and cancel orders in relationship with the mid-price variations.

Keywords: Hawkes processes, kernel estimations, power law kernels, high frequency, order book dynamics, trades, limit orders, cancel orders, market impact.

1 Introduction

The understanding of the price formation mechanism remains one of the most challenging problem in quantitative finance. In most modern financial exchanges, assets are traded via a continuous double auction: agents can choose to buy or sell a stock at certain prices posting *limit orders* in the order book or to execute the available limit orders using *market orders*. As long as their their limit orders are not executed, agents can also use *cancel orders* to take them off[1]. The intrinsic complexity of the formation of order book (it is the result of market, limit and cancel orders of a large number of anonymous traders) implies that in spite of its great practical and theoretical importance, there are relatively few full order book models (see

[1]See e.g., [11, 37] for more details on the functioning of continuous double auctions.

[35, 37] for "zero-intelligence", i.e. purely random model of book dynamics and [10, 16, 24] for Markovian models of the order book).

A class of simpler models does not take into account the amount of liquidity and focuses exclusively on asset prices which are described as the result of the *impact* of various type of orders: Former models considered exclusively the impact of market orders [6, 7, 25] while some recent models account for the influence of all types of order book events [11, 13]. In a recent work, some of us [4] introduced a price impact model using the framework of Hawkes processes. This model shares features with former impact models but allows one to describe the joint dynamics of both trades and price moves. In particular, unlike previous impact models, this Hawkes model takes into account the influence of price changes on themselves and on trades.

Hawkes processes were defined in [20] as point processes whose intensity function is a linear regression on the past of the process

$$\lambda_t^N = \mu + \int_{(-\infty,t)} \phi(t-s) dN_s.$$

The self and mutually exciting nature of Hawkes processes makes them naturally adapted to the modelling of multivariate counting processes. They have for example been successfully applied in many domains such as the study of earthquakes [32], neurobiology [34] or sociology [30]. In Finance, they have been recently used to model the arrival of trades or limit orders [23, 8, 28], price variations at microstructure level [3, 4, 19], financial contagion [1] or credit risk [14].

In this work, we will propose a generalization of the impact model of [4], in order to account for the interaction between price variations and any event (market, limit or cancel order) at the level I of the order-book. This model turns out to be an 8-dimensional Hawkes process that will allow us to have a direct measure of the endogeneity and the causality between these events, see [15] or [19] for related discussions about the use of Hawkes processes to measure the market endogeneity.

As far as estimation problems are concerned, the few former non parametric estimates of Hawkes kernels in the context of high frequency finance lead to power law decreasing kernels whose exponent is slightly higher than one [2, 4, 19]. As we will see, this implies that one needs to estimate the kernels on a wide range of time scales (typically from 10 microseconds to 100 seconds) to capture the dynamics of the order book. For that purpose we propose an improvement of the numerical scheme introduced in [4, 5] in order to handle slowly decreasing kernels. It will be illustrated on some numerical examples, that this method allows us to estimate the kernel on a large

range of time scales which cannot be done within the Gaussian scheme of [5].

The paper is organized as follows. After recalling the main definitions and properties of multidimensional Hawkes processes, we present the main non-parametric estimation principles in Section 2. Section 3 explains how these principles can be adapted to the non-parametric estimation of slowly decaying kernels. The so-obtained algorithm is presented and tested on numerical simulations at the end of this Section. The level I book model is then introduced in Section 4. It is calibrated on high frequency data corresponding to DAX and BUND front future contracts during one year (from June 2013 to June 2014). We then comment our results as far as endogeneity and mutual influence of all types of events are concerned. Discussion in relationship with former works and prospects for future research are provided in the concluding Section 5.

2 Multidimensional Hawkes processes and estimation principles

In this section we recall the definition of a D-dimensional Hawkes processes and briefly review its main properties. For more details we refer the reader to [5].

2.1 Basic definitions

Definition 2.1. *A D-dimensional point process $N_t = (N_t^i)_{1\leq i\leq D}$ is a Hawkes process if for every i, the ith component of its intensity function λ_t is a linear regression of the past jumps of N_t, i.e.,*

$$\lambda_t^i = \mu^i + \sum_{j=1}^{D} \int_{(-\infty,t)} \phi^{ij}(t-s)dN_s^j, \quad \forall i \in [1, D], \tag{1}$$

where $\mu = \{\mu^i\}_{1\leq i\leq D}$ is the so-called exogenous intensity *(with $\mu^i \geq 0$), and $\phi = \{\phi^{ij}\}_{1\leq i,j\leq D}$ the so-called* Hawkes kernel *matrix, where each $\phi^{ij}(t)$ is a real positive and causal function*[2]*. Using matrix convolution notation, the system of Equation (1) simply rewrites*

$$\lambda_t = \mu + \phi * dN_s. \tag{2}$$

Let us recall the following well known stability condition of the Hawkes processes, see for example [21].

Proposition 2.1. *If the matrix of the norms L^1 of the kernels defined as*

$$||\phi||_{=}(||\phi^{ij}||_1)_{1\leq i,j\leq D}$$

[2]By definition, a function is causal if its support is included in $\mathbb{R}^+$.

has a spectral radius strictly smaller than one then N admits a version with stationary increments.

Let us remark that Brémaud and Massoulié [9] have shown that this stability criterion remains valid when, in Equation (1), λ_t is a nonlinear positive Lipschitz function of $\{dN_s\}_{s<t}$ and where the elements of the kernel ϕ are not restricted to be positive. A particularly interesting generalization of Hawkes processes considered in [18] is

$$\lambda_t = (\mu + \phi * dN_s)^+ \tag{3}$$

where $(x)^+ = x$ if $x > 0$ and $(x)_+ = 0$ otherwise. This extension allows one to account for inhibitory effects when $\phi^{ij}(t) < 0$.

We will now assume that the condition of Proposition 2.1 is satisfied and consider a stationary Hawkes process. The next simple result describes the behavior of the first order statistics of N, see for example [5].

Proposition 2.2. *The average of the intensity vector of N:*

$$\Lambda = (\mathbb{E}[\lambda_t^1], ..., \mathbb{E}[\lambda_t^D])^T$$

satisfies

$$\Lambda = (\mathbb{I} - ||\phi||)^{-1}\mu. \tag{4}$$

2.2 Second-order properties and Wiener-Hopf system

The second order statistics of Hawkes processes are naturally characterized by their infinitesimal covariance functions.

Definition 2.2. *The infinitesimal covariance ν of N is the $D \times D$ matrix whose elements are*

$$\nu^{ij}(t - t')dtdt' = \mathbb{E}[dN_t^i dN_{t'}^j] - \Lambda^i \Lambda^j dtdt',$$

where Σ is the $D \times D$ matrix whose diagonal is the average intensities Λ^i.

The following result is shown in [2].

Proposition 2.3. *ν can be expressed as a functional of ϕ:*

$$\nu = (\delta\mathbb{I} + \psi) * \Sigma(\delta\mathbb{I} + \psi^T) \tag{5}$$

where $\delta(t)$ stands for the Dirac distribution, Σ is the $D \times D$ matrix whose diagonal is the average intensities Λ^i, and where the matrix $\psi(t)$ is is defined as:

$$\psi(t) = \sum_{k \geq 1} \phi^{(*k)}(t), \tag{6}$$

in which we used the notation $\phi^{(*k)}(t) = \phi * \phi * \ldots * \phi$ (where ϕ is repeated k times).

Let us mention that there is an alternative "population representation" of Hawkes processes (see e.g., [22]) according to which they are built as the times of arrival and birth of the following population process:

- There are individuals of type 1, ..., D.
- Migrants of type i arrive at a Poisson rate of μ^i.
- Every individual can have children of (a priori) all types and the children of type i of an individual of type j who was born or migrated in t are distributed as an inhomogeneous Poisson process of intensity $\phi^{ij}(\cdot - t)$.

In this model, $||\phi^{ij}||$ appears as the average number of children of type i of an individual of type j. We thus use $||\phi^{ij}||$ as a measure of "causality" between j and i. Similarly, $\Lambda^j/\Lambda^i||\phi^{ij}||$ appears as the proportion of individuals of type i whose parent is an individual of type j.

In this construction, the elements ψ^{ij} of the matrix ψ also have a natural interpretation. The descendants of type i of each exogenous event of type j occurring at time t are a point process of average intensity $\psi^{ij}(\cdot - t)$. Therefore, $||\phi^{ij}||$ appears as the average number of individuals of type i which descend from an individual of type j. Similarly, $\Lambda^j/\Lambda^i||\psi^{ij}||$ appears as the proportion of individuals of type i whose "ancestor" is an individual of type j.

As explained in [5], since the jumps are always of size one, the second order statistics of the Hawkes process can also be summed up by its conditional laws.

Definition 2.3. *We define the conditional laws denoted $(g^{ij})_{1\leq i,j\leq D}$ as the non singular part of the measure $\mathbb{E}[dN_t^i|dN_0^j = 1]$:*

$$g^{ij}(t)dt = \mathbb{E}[dN_t^i|dN_0^j = 1] - \mathbb{1}_{i=j}\delta(t) - \Lambda^i dt.$$

They are linked with the infinitesimal covariance by the equation

$$g(t) = \nu(t)\Sigma^{-1} - \delta(t)\mathbb{I}. \tag{7}$$

The following proposition, see [5], links the conditional laws and the kernels of Hawkes processes through a Wiener-Hopf system of equations.

Proposition 2.4. *Given the conditional laws of a Hawkes process g, its kernel ϕ is the only*[3] *solution of the Wiener-Hopf system:*

$$g(t) = \phi(t) + g * \phi(t), \quad \forall t > 0. \tag{8}$$

[3]The uniqueness of the solution of Equation (8), considered as an equation in ϕ, is shown in [5].

Remark 2.1. *The linear structure of Hawkes processes implies that they have properties similar to that of autoregressive processes. In particular, the previous proposition can be seen as the counter part of the Yule-Walker Equation.*

2.3 Existence and uniqueness of solutions of the Wiener-Hopf system

The Wiener-Hopf system (8) is at the core of our estimation procedure. In [5], it is shown that if g is the conditional law of a Hawkes process as defined in Definition 2.3 then the system (8) admits the unique causal solution ϕ.

Let us now assume that Σ, ν and g are respectively the average, the infinitesimal covariance function and the conditional law of any stationary D-dimensional point process (not necessarily a Hawkes process). Using the Wiener-Khintchine Theorem, we get that the matrix $\widehat{\nu}(z)$ is positive definite for any $z \in \mathbb{R}$. We can thus apply Theorem 8.2 of [17] to decompose ν as

$$\nu(t) = Y(t) * Y^T(-t)$$

where Y is causal. We then set

$$\psi = (Y - \delta\mathbb{I})\sqrt{\Sigma^{-1}}$$

and ϕ such that

$$\hat{\phi} = \hat{\psi}(\mathbb{I} + \hat{\psi})^{-1}$$

to get that ϕ satisfies Equation (5) and thus Equation (8) (these two equations are equivalent for $g(t) = \nu(t)\Sigma^{-1} - \delta(t)\mathbb{I}$).

Therefore, even if the point process that we are studying is not a Hawkes process Equation (8) has one and only one solution. It thus makes sense to try to find the solution of the Wiener-Hopf system for empirical data[4]. Of course, the so-obtained solution is no longer necessarily positive. We will comment the interpretation of a negative ϕ later.

Remark 2.2. *It is noteworthy that, from a fundamental point of view, solving the Wiener-Hopf system amounts to finding the kernel matrix ϕ (and the intensity vector μ) that provide the best linear predictor (in the sense of mean-square error) of the conditional intensity. This remark leads the authors of [18, 34] to consider an alternative estimation algorithm relying on the minimization of a contrast function, a proxy for the mean square error (see [5] for more details).*

[4]Even if the empirical conditional laws are not equal to the real conditional laws.

2.4 Estimation principles

As explained in [5], the procedure for solving numerically the Wiener-Hopf system (8) follows Nyström method [31]. More precisely, given a realization if the process N on an interval $[0, T]$, the procedure for non parametric estimation is the following

1. Compute an estimation $\tilde{g}$ of the matrix function g using a "fine enough" time grid (values of g outside of this time grid will be computed using an interpolation scheme).

2. Use a quadrature method to discretize the Wiener-Hopf system (8) on an interval $[T_{min}, T_{max}]$. If we use the quadrature points $\{t_k\}_{1\leq k\leq K}$ along with the quadrature weights $\{w_k\}_{1\leq k\leq K}$, one gets, the system

$$\tilde{g}^{ij}(t_n) = \tilde{\phi}^{ij}(t_n) + \sum_{l=1}^{D}\sum_{k=1}^{K} w_k \tilde{g}^{il}(t_n - t_k)\tilde{\phi}^{lj}(t_k), \quad \forall n \in [0, K], \quad (9)$$

   for all $i, j \in [1, D]$.

3. Inverse this so-obtained KD^2 linear system. This leads to the estimation of the matrix kernel at the quadrature points $\{\tilde{\phi}^{ij}(t_n)\}_{i,k\in[1,D],n\in[1,K]}$ as well as an estimation of $||\phi||_1$ (using quadrature).

4. Estimate empirically the average intensity Λ (just counting the number of points on each component of the process) and estimate μ using (4).

A thorough study of this algorithm has been done in [5] (in the framework on a Gaussian quadrature) and showed that it performs a fast and efficient non parametric estimation of a Hawkes process. It proved particularly competitive in the case the realization process has a large number of jumps and the matrix kernel is not well localized in time. Let us point out that [5] showed that this algorithm leads to precise result even if some elements of the kernel matrix have moderate negative values.

However, as we will show in the next section, when some elements of the kernel matrix are power-law, the quality of the estimation can be improved significantly. In this case, the use of Gaussian quadrature is not adapted and leads to inaccurate estimations because the "mass" of the power-law kernel exponents are spread over a large range of scales. We shall study this phenomenon in the framework of high frequency financial data.

3 Non parametric estimation of slowly decaying kernels

One of the main issue when dealing with high frequency financial data is that the range of scales involved in the dynamics is rather large. This translates

into kernels with large support, typically from $T_{min} = 100$ microseconds to $T_{max} = 100$ seconds, slowly decaying (power-law like), varying very quickly at small lags (around T_{min}) and much slower at larger lags (around T_{max}). Let us follow step by step the algorithm of Section 2.4 and explain how to adapt it to this framework.

3.1 Using an adapted sample-grid for g

Step 1. of the algorithm (estimation of g) is clearly the first sensitive step. Since the kernels are varying very quickly at small lags, more care should be taken for estimating the behavior of g in the neighbourhood of 0 than anywhere else. In Appendix A, we describe precisely the algorithm we shall use in this paper for the estimation of g. It basically consists in two step procedure: (i) empirical computation of g on a time-grid which is more dense around 0 than anywhere else, (ii) linear interpolation for computing the values of g at any time (in all our numerical experiments, we have checked than higher-order approximation does not bring relevant precision to the overall algorithm).

3.2 Towards an adapted quadrature scheme

The choice of the quadrature (step 2. in the algorithm) is the next (and last) sensitive step. The Gaussian quadrature used in [5] for solving the Wiener-Hopf system is clearly not adapted to this situation. Indeed, the Gaussian quadrature points are roughly uniform over $[0, T]$ whereas the kernels are varying very quickly around 0 and slowly anywhere else. In order to get precise estimation of kernels on the whole range $[T_{min}, T_{max}]$, the Gaussian quadrature is constrained by their behavior at small lags ($t \simeq T_{min}$) and would imply using an number of quadrature points of the order of $K = T_{max}/T_{min} = 10^6$ which, even for $D = 1$, is far to many to be able to solve numerically the system (9).

One way to circumvent this difficulty could consist in adding points close to zero by using the change of variable proposed in [36]. Let $\{t_k, w_k\}_{1 \leq k \leq K}$ correspond to the Gaussian quadrature scheme on $[\log(T_{min}), \log(T_{max})]$, we then set

$$(t'_k, w'_k) = (e^{t_k}, e^{t_k} w_k). \tag{10}$$

We then replace the system (9) by the new system:

$$\tilde{g}^{ij}(t'_n) = \tilde{\phi}^{ij}(t'_n) + \sum_{l=1}^{D} \sum_{k=1}^{K} w'_k \tilde{g}^{il}(t'_n - t'_k) \tilde{\phi}^{lj}(t'_k). \tag{11}$$

In order to test this new estimation scheme, we simulate a one-dimensional

Hawkes process with parameters

$$\phi(t) = \frac{0.06}{(0.005+t)^{1.3}} \quad \text{and} \quad \mu = 0.05. \tag{12}$$

The simulation is performed on a time-period long enough (10^7 seconds) so that the Monte Carlo error is small enough compared to the "quadrature error".

Remark 3.1. *This kernel checks the same kind of "multiscale behavior" as kernels involved in high frequency financial dynamics. Indeed, 90% of its L^1 norm is spread over 5 decades: from 10^{-3} to 10^2 seconds, in the sense that $\int_0^{+\infty} \phi(t)dt = 0.98$, $\int_0^{10^{-3}} \phi(t)dt = 0.05$ and $\int_{10^2}^{+\infty} \phi(t)dt = 0.05$.*

We applied the algorithm described in Section 2.4, in which we use the procedure described in Appendix A to estimate g using $h_{min} = 1ms$, $h_{max} = 1000s$ and $h_\delta = 0.05$, and in which we replaced (9) by (11). We chose $K = 200$, $T_{min} = 1ms$ and $T_{max} = 2000s$. Figure 1, displays the so-obtained results. On the left, the log-log plot of both the estimated kernel and the theoretical kernel (as well as the conditional law g) are displayed. On the right, the integral on $[0, t]$ of both theoretical and estimated kernels are displayed.

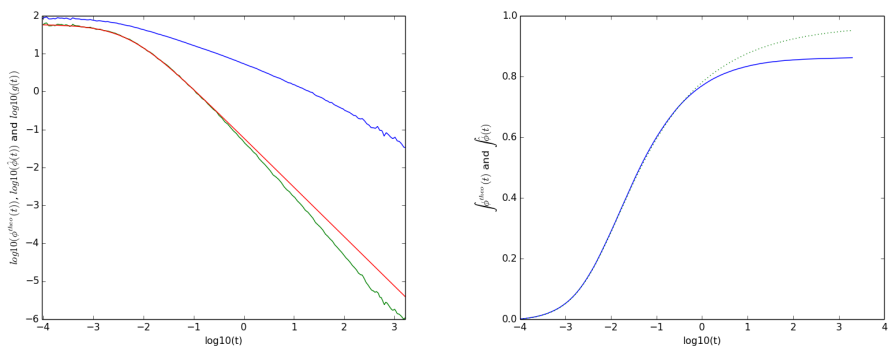


Figure 1: Left: Estimation of the conditional law g and kernel ϕ in the case of a 1-dimensional Hawkes of size $T = 10^7$ seconds, with parameters (12). The estimation of g is performed using the procedure described in Appendix A using $h_{min} = 1ms$, $h_{max} = 1000s$ and $h_\delta = 0.05$, and kernel estimation is performed using algorithm of Section 9 with Gaussian quadrature followed by the change of variable (10) ($K = 200$, $T_{min} = 1ms$, $T_{max} = 2000s$ and thus $\delta = 0.077$). The three curves correspond to $\log_{10} - \log_{10}$ empirical conditional law (blue), theoretical kernel (red) and estimated kernel (green) for 200 quadrature points.
Right: Cumulated theoretical kernel $\int_0^t \phi(s)ds$ (green dots) and cumulated estimated kernel $\int_0^t \tilde{\phi}(s)ds$ (blue) as a function of $\log_{10}(t)$.

We observe that, for large times ($t \geq 1s$), the estimated kernel does not approximate well the theoretical kernel. Moreover, this procedure significantly underestimates the cumulated kernel. Indeed, we get $||\tilde{\phi}||_1 = 0.85$ instead of $||\phi||_1 = 0.98$. This error on the estimation of the norms will be very important in our study of the causality between market events. Indeed the norms of the kernels $||\phi^{ij}||_1$ will have a direct interpretation as the average number of events of type i caused by an event of type j.

Let us try to give an intuitive explanation of the bad performance of this method: Recall that the quadrature approximates

$$\int_0^{+\infty} g(t'_n - s)\phi(s)ds \simeq \sum_{k=1}^{K} w'_k g(t'_n - t'_k)\phi(t'_k), \quad \forall n \in [1, K]$$

Thanks to the change of variable (10) there are more quadrature points around 0 than around any other time. That ensures that the quadrature approximation captures the fast variation of $\phi(s)$ around $s = 0$. However, g is also varying quickly around 0, the estimation error illustrated in Figure 1 comes from the bad approximation of $g(t'_n - s)$ around $s = t'_n$ (i.e., $k = n$ in the previous quadrature formula).

In the next section, we propose a new quadrature scheme that will capture all behaviour of the conditional law g next to t_n for any n.

3.3 The adapted quadrature scheme for slow-decaying kernels

Let us consider the time grid which is uniform from 0 to T_{min} and log-uniform between T_{min} and T_{max} using the same grid-size $\delta < 1$

$$\{t_k\}_{1 \leq k \leq K} = [0, \delta T_{min}, 2\delta T_{min}, ..., T_{min}, T_{min}e^{\delta}, T_{min}e^{2\delta}, ..., T_{max}]. \quad (13)$$

The adapted quadrature scheme consists in considering that the kernels are piecewise affine on $[t_k, t_{k+1}]$:

$$\phi(t) = \phi(t_k) + \frac{t - t_k}{t_{k+1} - t_k}(\phi(t_{k+1}) - \phi(t_k)).$$

Under this linear hypothesis, Equation (8), in the $D = 1$ dimensional case, becomes (the generalization to dimension $D > 1$ is straightforward):

$$\begin{aligned} g(t_n) &= \phi(t_n) + \sum_{k=0}^{N-1} \phi(t_k) \int_{t_k}^{t_{k+1}} g(t_n - s) ds \\ &+ \sum_{k=0}^{N-1} (\phi(t_{k+1}) - \phi(t_k)) \int_{t_k}^{t_{k+1}} \frac{s - t_k}{t_{k+1} - t_k} g(t_n - s) ds \\ &= \phi(t_n) + \sum_{k=0}^{N-1} \phi(t_k) \int_{t_n - t_{k+1}}^{t_n - t_k} g(u) du \\ &+ \sum_{k=0}^{N-1} \frac{\phi(t_{k+1}) - \phi(t_k)}{t_{k+1} - t_k} \int_{t_n - t_{k+1}}^{t_n - t_k} (t_n - t_k - u) g(u) du \end{aligned}$$

Thus, using the adapted quadrature, the system (9) becomes (in the $D = 1$ case)

$$\begin{aligned} \tilde{g}(t_n) &= \tilde{\phi}(t_n) + \sum_{k=0}^{N-1} \tilde{\phi}(t_k) \int_{t_n - t_{k+1}}^{t_n - t_k} \tilde{g}(u) du \\ &+ \sum_{k=0}^{N-1} \frac{(\tilde{\phi}(t_{k+1}) - \tilde{\phi}(t_k))(t_n - t_k)}{t_{k+1} - t_k} \int_{t_n - t_{k+1}}^{t_n - t_k} \tilde{g}(u) du \\ &- \sum_{k=0}^{N-1} \frac{\tilde{\phi}(t_{k+1}) - \tilde{\phi}(t_k)}{t_{k+1} - t_k} \int_{t_n - t_{k+1}}^{t_n - t_k} u \tilde{g}(u) du. \end{aligned} \tag{14}$$

The empirical values of the integrals $\int_0^x \tilde{g}(u) du$ and $\int_0^x u\tilde{g}(u) du$ can easily be computed beforehand at all abscissa of the form $x = t_n - t_k$ using the linear interpolation of the conditional laws described in Section 3.1 (see also Appendix A).

As before, this is a linear equation in the values of ϕ at the quadrature points. Solving it, one gets an estimation of ϕ at these points. A simple linear interpolation can be performed to get estimation at other points.

Applying the same testing procedure as before, we get, see Figure 2 that this scheme works perfectly with a reasonable number of quadrature points.

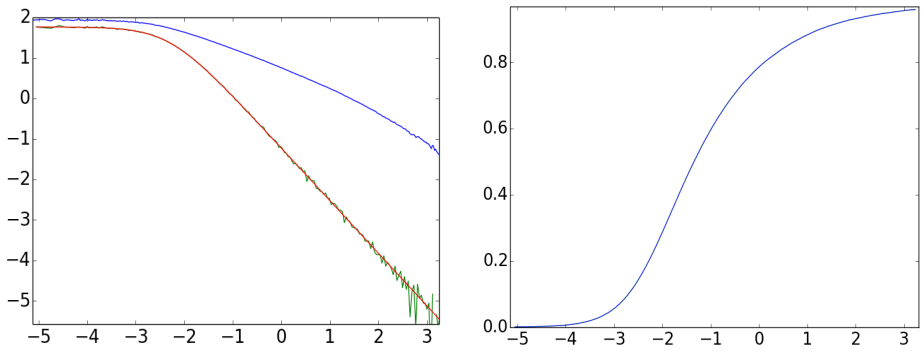


Figure 2: Left: Estimation of the conditional law g and kernel ϕ for the same Hawkes process realization as in Figure 1. The kernel estimation is performed using the algorithm of Section 9 where the quadrature of step 3. has been replaced by the adapted quadrature given by (14) ($K = 200$, $T_{min} = 1ms$, $T_{max} = 2000s$ and thus $\delta = 0.077$). The three curves correspond to $\log_{10} - \log_{10}$ empirical conditional law (blue), theoretical kernel (red) and estimated kernel (green) for 200 quadrature points. The estimation and the theoretical kernels perfectly match.
Right: Cumulated theoretical kernel $\int_0^t \phi(s)ds$ (green dots) and cumulated estimated kernel $\int_0^t \tilde{\phi}(s)ds$ (blue) as a function of $\log_{10}(t)$.

3.4 The adapted estimation procedure

Here is the step by step estimation procedure that will be used all along this paper:

1. Compute an estimation $\tilde{g}$ of the matrix function g using the procedure described in Section 3.1.

2. Use the adapted quadrature method to discretize the Wiener-Hopf system (8) on an interval $[T_{min}, T_{max}]$. The quadrature points are given by (13) and the scheme by (14) (this last equation corresponds to the one dimensional scheme, but generalization to any dimension is obvious).

3. Inverse this so-obtained KD^2 linear system. This leads to the estimation of the matrix kernel at the quadrature points $\{\tilde{\phi}^{ij}(t_n)\}_{i,k\in[1,D],n\in[1,K]}$ as well as an estimation of $||\phi||_1$ (using quadrature).

4. Estimate empirically the average intensity Λ and estimate μ using (4).

4 Hawkes model for Level I Order book events

4.1 Definition of the model

Multivariate Hawkes models provide a natural framework to account for the impact of past events of various type on the rate of arrival of future events. In [4], this framework has been considered to extent classical impact price models (as in e.g. [7]) by accounting in real time for the "impact" of market orders on price changes but also for the auto-regressive dynamics of price changes and their retro-action on the rate of market order arrivals. Along the same line, one can extend the model of [4] by accounting, within a multidimensional Hawkes model, for the cross and self influencing dynamics of all event types in the order book. For the sake of parsimony and simplicity, we consider only events occurring at the Level I of the order book, i.e., events that change the state of the order book at the best bid or best ask levels. More precisely we consider the following 8 dimensional counting process:

$$N_t = (P_t^{(a)}, P_t^{(b)}, T_t^{(a)}, T_t^{(b)}, L_t^{(a)}, L_t^{(b)}, C_t^{(a)}, C_t^{(b)})$$

where:

- $P^{(a)}$ (resp. $P^{(b)}$) counts the number of upward (resp. downward) mid-price moves.
- $T^{(a)}$ (resp. $T^{(b)}$) counts the number of market orders at the ask (resp. bid) that do not move the price.
- $L^{(a)}$ (resp. $L^{(b)}$) counts the number of limit orders at the ask (resp. bid) that do not move the price.
- $C^{(a)}$ (resp. $C^{(b)}$) counts the number of cancel orders at the ask (resp. bid) that do not move the price.

Remark 4.1. *Let us stress that mid-price moves can correspond to the occurrence of a market order or a cancel order that eats all the available liquidity at best bid or best ask or to a limit order placed in the spread between best bid and best ask. We choose to not distinguish these events in order to handle a process with relatively low dimension.*

The previous counting process can be considered as an 8-dimensional Hawkes process characterized by 8 exogenous intensities and 64 kernels. We denote by $\phi^{N^j \to N^i}$ the kernel ϕ^{ij} coding the influence of the jth process on the ith intensity. For example $\phi^{T^{(b)} \to P^{(a)}}$ corresponds to the influence of the trades at the bid that left the mid-price unchanged on the upward price moves. These intensities and kernels can be estimated from empirical book data using the procedure described in Section 3.4.

Notice that, in the following, we never impose "by hand" any (bid,downward)-(ask,upward) symmetry. However, it can be seen in Appendix B that, as expected, this symmetry is rather well satisfied (for example, $\phi^{T^{(b)}\to P^{(a)}} \simeq \phi^{T^{(a)}\to P^{(b)}}$) in our estimations. This indicates that, as suggested on simulations, while we perform our kernel estimations on processes with many dimensions and on many time scales, our method and the amount of data that we use, allows us to retrieve significant results about the kernels. Indeed, an unstable method would hardly allows one to retrieve this symmetry.

4.2 Description of the database

The financial data used in this paper have been provided by the company QuantHouse EUROPE/ASIA (http://www.quanthouse.com). It consists in all level-I order book data[5] of BUND and DAX future contracts. For every day, we only keep the most liquid maturity and we use data over one year from June 2013 to June 2014. Each file lists all the changes in the first limit (best ask or best bid) of the order book at a micro second precision. We can thus easily precisely compute from this data the times of the different market events that will be of interest here (ask or bid market, limit or cancel orders at the first limit and upward and downward mid price moves).

At the time scales that we shall study (from 10^{-5} to 10^{2} seconds), the order book dynamics strongly depend on the tick size of the asset (or more precisely by the quotient between the average spread and the tick size, see [6] or [12] for characterization and differences between small and large tick assets). In that respect, the DAX corresponds to a "small-tick" asset while the Bund is a "large tick" asset.

The number of events in our sample is summarized in Table 1.

	$P^{(a)}$	$P^{(b)}$	$T^{(a)}$	$T^{(b)}$	$L^{(a)}$	$L^{(b)}$	$C^{(a)}$	$C^{(b)}$
xFDAX	9.36	9.35	2.66	2.67	19.7	19.6	23.3	23.1
xFGBL	1.72	1.72	3.40	3.48	29.8	29.8	26.6	26.5

Table 1: Total number of events (in millions).

We have checked using numerical simulations of multidimensional Hawkes processes with power-law kernels, that the size of the sample is sufficient to provide reliable estimations of the shapes of the kernels. We have also checked that our main results do not depend on intraday seasonal effects: if

[5] That is the times of market orders and limit and cancel orders at the first limit and the state of the first bid-ask queues at these times.

one selects a 1 hour intraday time slice, we obtain the same results (up to some statistical noise).

4.3 Conditional law estimations

To estimate the 64 conditional laws we proceed as explained in Section 3.1 (see also Appendix A) taking the parameters $h_\delta = 0.05$, $h_{min} = 0.1$ milliseconds and $h_{max} \simeq 1059$ seconds (see (16)) on the period from 1 June 2013 to 1 June 2014 (252 open days).

Surprisingly most of the conditional laws between financial events have very similar properties. As an illustration, Figure 3 displays, for the Bund, $g^{P^{(a)}\to P^{(b)}}$, $g^{T^{(a)}\to T^{(a)}}$, $g^{L^{(a)}\to L^{(a)}}$, $g^{C^{(a)}\to C^{(a)}}$, $g^{L^{(a)}\to L^{(b)}}$ and $g^{L^{(a)}\to C^{(a)}}$ in $\log_{10} - \log_{10}$.

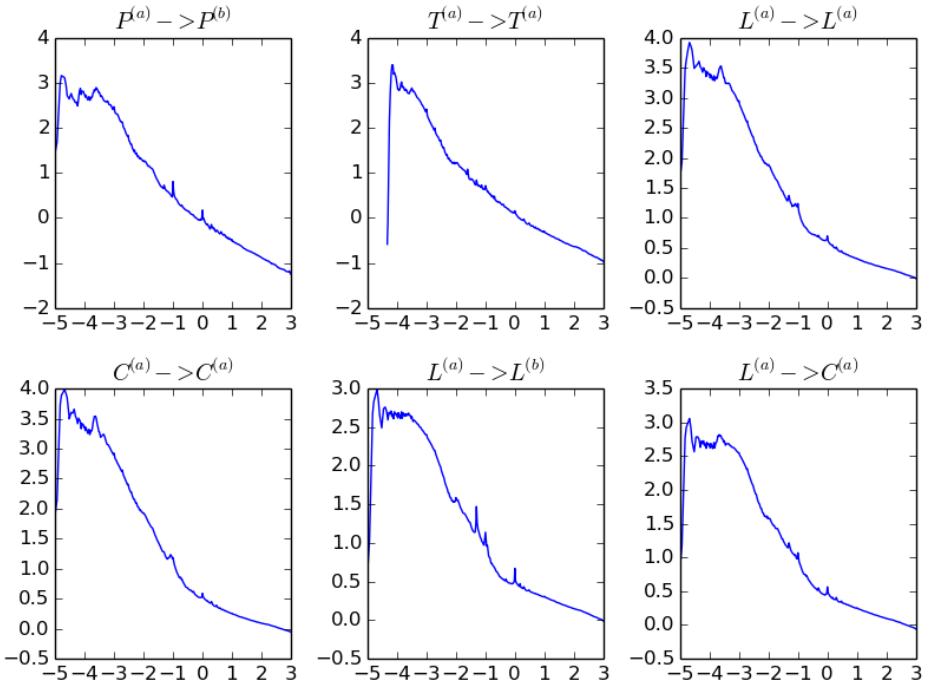


Figure 3: $\log_{10} - \log_{10}$ empirical conditional laws $g^{P^{(a)}\to P^{(b)}}$, $g^{T^{(a)}\to T^{(a)}}$, $g^{L^{(a)}\to L^{(a)}}$ and $g^{C^{(a)}\to C^{(a)}}$ for the Bund.

The first thing that we notice is that above a time scale $\tau_1 \sim 0.1$ seconds, these conditional laws behave as $t^{-\gamma}$ with $\gamma < 1$. For $g^{T^{(a)}\to T^{(a)}}$, this is the well known long-range memory of the order flows. Between τ_1 and $\tau_2 \sim 0.3ms$, the conditional laws roughly behaves as a power law of exponent of order 1. Below, τ_2, the conditional laws "saturate".

We also observe a few "bumps" on the conditional law which do not correspond to noise in the estimation. We believe that there are two kinds of bumps. The first kind of bumps appear at "round" times (0.01, 0.1, 0.5, 1 and 2 seconds), we believe that they are due to the automatization of trading. For example, if an algorithm posts a limit order every second, this implies a bump in the conditional law at 1 second. The second kind of bumps appears around $\tau_2 \sim 0.3$ milliseconds. We believe that it corresponds to the average reaction time (i.e., the average "latency") of the agents to an event. The decreasing of the conditional law below 0.03 milliseconds is an artefact of our data.

Once the conditional laws have been estimated, we solve the Wiener-Hopf system, following steps 2. and 3. in Section 3.4 with the parameters $K = 100$ (so that $\delta \simeq 0.15$), $T_{min} = 0.1$ milliseconds and $T_{max} \simeq 140$ seconds and compute the exogenous intensity estimates following step 4. We shall first comment on the values of the exogenous intensities.

4.4 Exogenous intensities

Let us recall that, within the Hawkes model, the exogenous intensity μ^i can be interpreted as the rate of (Poisson) events of type i that are "coming" from an exogenous source of information, i.e., that are not "caused" by any other past event in the model. In that respect, the ratio:

$$R^i = \frac{\mu^i}{\Lambda^i} \tag{15}$$

represents an exogeneity ratio, namely the ratio between the number of exogenous events and the total number of events of type i. Notice that, in the one dimensional case, according to Proposition 2.2, one has simply: $R = 1 - ||\phi||_1$.

	$P^{(a)}$	$P^{(b)}$	$T^{(a)}$	$T^{(b)}$	$L^{(a)}$	$L^{(b)}$	$C^{(a)}$	$C^{(b)}$
μ	2.37e-2	2.39e-2	1.06e-2	1.14e-2	2.07e-2	2.27e-2	1.53e-2	7.51e-3
R	2.7%	2.7%	4.3%	4.5%	1.1%	1.2%	0.7%	0.4%

Table 2: Estimated exogenous intensities (in s^{-1}) and the corresponding exogenous ratio (15) of the DAX futures.

	$P^{(a)}$	$P^{(b)}$	$T^{(a)}$	$T^{(b)}$	$L^{(a)}$	$L^{(b)}$	$C^{(a)}$	$C^{(b)}$
μ	7.13e-3	7.10e-3	1.41e-2	1.45e-2	3.83e-2	3.83e-2	4.00e-2	4.39e-2
R	4.4%	4.4%	4.5%	4.5%	1.4%	1.4%	1.6%	1.8%

Table 3: Estimated exogenous intensities (in s^{-1}) and the corresponding exogenous ratio (15) of the BUND futures.

The estimated intensities and the corresponding exogenous ratio for the DAX and the BUND futures are reported respectively in Tables 2 and 3. These tables reveal that, for both assets, the level of exogeneity is very low: R is only a few percent, meaning that most of the events can be considered to be directly triggered by past events within this model. See [26, 27] for theoretical studies of the diffusive long term behavior of such almost purely endogenous Hawkes processes. For comparison, a simple one dimensional Hawkes model accounting for mid-price jump events as in refs [15] provides an exogenous ratio of respectively $R = 0.059$ for the DAX and $R = 0.10$ for the BUND future. This means that the 8-dimensional Hawkes model provides a better description of mid-price changes than a simple one dimensional model that has to involve a larger amount of "external" sources of information.

Let us also remark that, for both assets, market orders are more exogenous than limit and cancel orders. This is not surprising since several studies tend to show that market orders are "leading" limit and cancel. Thus, degree of endogeneity of the limit and cancel orders should be greater than the one of the market orders.

Finally, one can see that though limit and market orders have strikingly similar exogenous ratio for small and large tick assets, the mid-price changes and cancel occurrences are more endogenous in the case of small tick asset (DAX).

4.5 Matrix of kernel norms $||\phi||_1$

Before discussing the precise shape of the estimated kernels, we use the values of the norms of the kernels ϕ^{ij} to comment the main mutual and self excitations that occur in the order book[6]. The matrices of the kernel norms $||\phi||_1$ we obtained for the DAX and the BUND are reported in Figure 4. For the sake of simplicity, we have represented the norm values using a colormap from blue to red. Notice that blue values correspond to negative

[6]Let us recall that the norm $||\phi^{ij}||_1$ represents the mean number of events of type i triggered by an event of type j.

kernel norms[7]. This feature has a natural interpretation as an inhibitory effect within the non-linear version of the Hawkes model described by Equation (3). We have checked, on numerical simulations, that as long as the realized quantity (2) remains most of the time positive, the non parametric estimation procedure introduced in Section 3.4, when applied to the non linear Hawkes process (3), leads to reliable estimations even for negative kernels.

As expected, the overall symmetry (ask/bid) is fairly well recovered empirically on the matrix shape. Any 2×2 sub-matrix with homogeneous inputs (i.e., same type of inputs to be chosen among Price changes, Trades, Cancels or Limits) and homogeneous outputs is symmetric. Thus for instance the kernel $\phi^{T^{(a)} \to C^{(b)}}$ seems to have the same norm as the kernel $\phi^{T^{(b)} \to C^{(a)}}$. Let us point out that, since the Dax is a stock index we could have expected some discrepancy between the ask side and the bid side (statistics of downward jumps are slightly different than statistics of upward jumps on equity markets) however, this discrepancy is negligible on intraday data.

One striking feature that clearly appears on both matrices is the anti-diagonal shape of the $(P^{(a)}, P^{(b)})$ sub-matrix and the diagonal shape of the $(T^{(a)}, T^{(b)})$, $(L^{(a)}, L^{(b)})$, $(C^{(a)}, C^{(b)})$ sub-matrices, see also Paragraphs B.1, B.2, B.3 and B.4 in appendix where the corresponding normalized cumulated kernels are plotted. This means that, on the one hand, market, limit and cancel orders mainly cause events of the same type and sign. This property can be mainly attributed to the splitting of metaorders, see [29], and to a less extent to some herding behaviour of agents. We will see in the next section that these diagonal kernels are well described by roughly power-law decreasing functions. As far as mid-price jumps are concerned, the cross-exciting structure between $P^{(a)}$ and $P^{(b)}$ implies a strongly mean-reverting behaviour, which is of the main characteristic of the price microstructure and which somehow guarantees the absence of long range correlation in prices, see [4], in agreement with market efficiency. Note that this cross price kernel can be linked to the kernel of the propagator model, see [7, 25].

Finally, let us point out that that price changes appear to be the components that influences the most the other components, i.e., a price change "drives" the dynamics more than anything else. We will come back on this feature and on other features of this matrix (impact of orders on the price, impact of the price on events, impact of trades on the liquidity,...) in the following sections.

[7] Let us remark that we allow an abuse of language in the sense that $||\phi||_1$ stands for $\int_0^\infty \phi(s)ds$ which is not really a norm unless $\phi(t) > 0$.

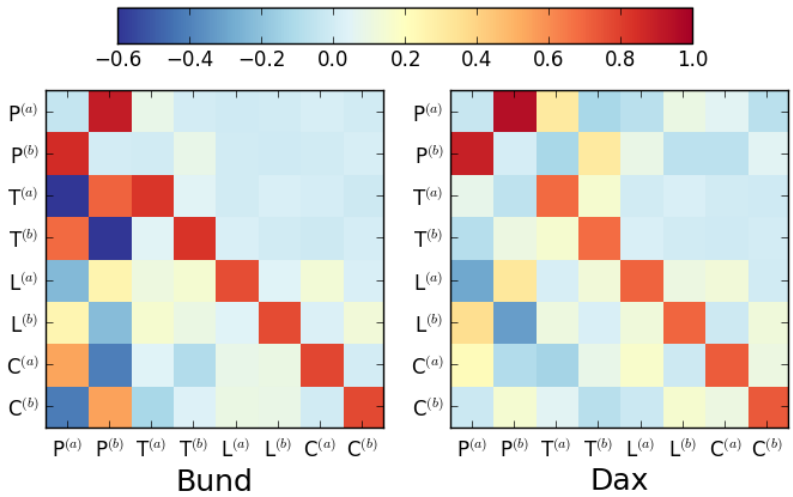


Figure 4: The matrix of estimated $||\phi^{ij}||$.

4.6 Matrix of norms $||\psi||_1$

Let us recall that the kernel ϕ^{ij} can be considered as the "bare" impact of an event of type j on an event of type i. If one wants to account for the "dressed" impact, i.e. the impact associated with all the cascade triggered by some event, one has to estimate the function ψ^{ij} as defined by (6) whose matrix of norms can be computed as:

$$||\psi||_1 = ||||\phi||_1(\mathbb{I} - ||\phi||_1)^{-1}.$$

More precisely, if one introduces:

$$||\bar{\psi}^{ij}||_1 = \frac{\mu^j}{\Lambda^i}||\psi^{ij}||_1$$

then, according to the population interpretation mentionned in section 2.2 and Proposition 2.2, $||\bar{\psi}^{ij}||$ corresponds to the fraction of events of type i triggered by *exogenous* events of type j (while $||\psi||_1$ corresponds to the average number of events of type i indirectly generated by an event of type j). The matrix norm $||\bar{\psi}||$ is displayed in Figure 5.

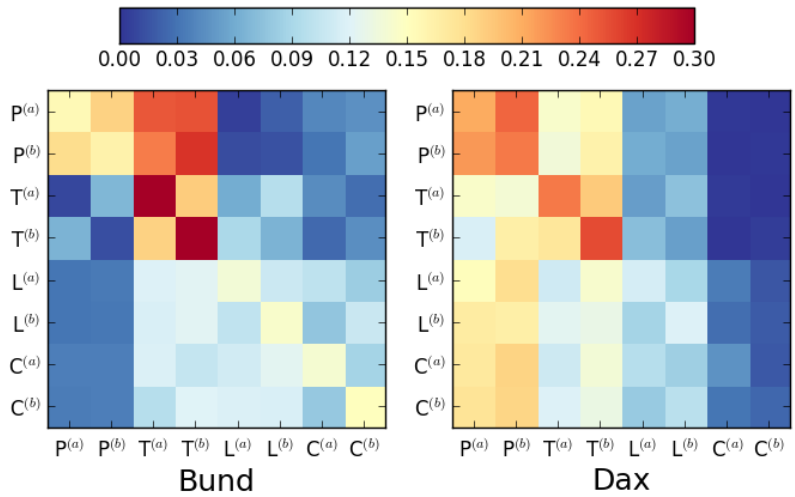


Figure 5: The matrix of estimated $||\bar{\psi}^{ij}||$.

One can see that exogenous limit and cancel orders have a poor influence on mid-price variations that are mainly caused by exogenous trades and price jumps. This indicates that exogenous information is mainly incorporated into prices through trades or orders that directly shift the mid-price. One can also remark that in the case of large tick asset (BUND), an exogenous price move mainly impacts the price itself but not the order flow that does not move the price while for the small tick asset (DAX), an exogenous price variation impacts all type of events.

4.7 Shape of the kernels

In all former studies where non parametric estimations of Hawkes kernels involved in the dynamics of order flows were performed [2, 4, 19], power-law kernels with exponents close to 1 have been observed. This property can be directly linked to the strong persistence of the order flow dynamics, see [25], mainly caused by the splitting of large orders and, to a lesser extent, to the herding behavior of agents. This feature remains true in the 8-dimensional description adopted in this paper when one focuses on the kernels involved in the self-excitation of market, limit and cancel orders and in the cross excitation of mid-price moves (the kernels with highest norms, see Figure 4).

Figure 6 displays $log_{10} - log_{10}$ plots of the kernels $\phi^{T^{(a)}\to T^{(a)}}$, $\phi^{L^{(a)}\to L^{(a)}}$, $\phi^{C^{(a)}\to C^{(a)}}$ and $\phi^{P^{(a)}\to P^{(b)}}$. It clearly appears that these kernels loosely behave as a power law of exponent slightly higher than one.

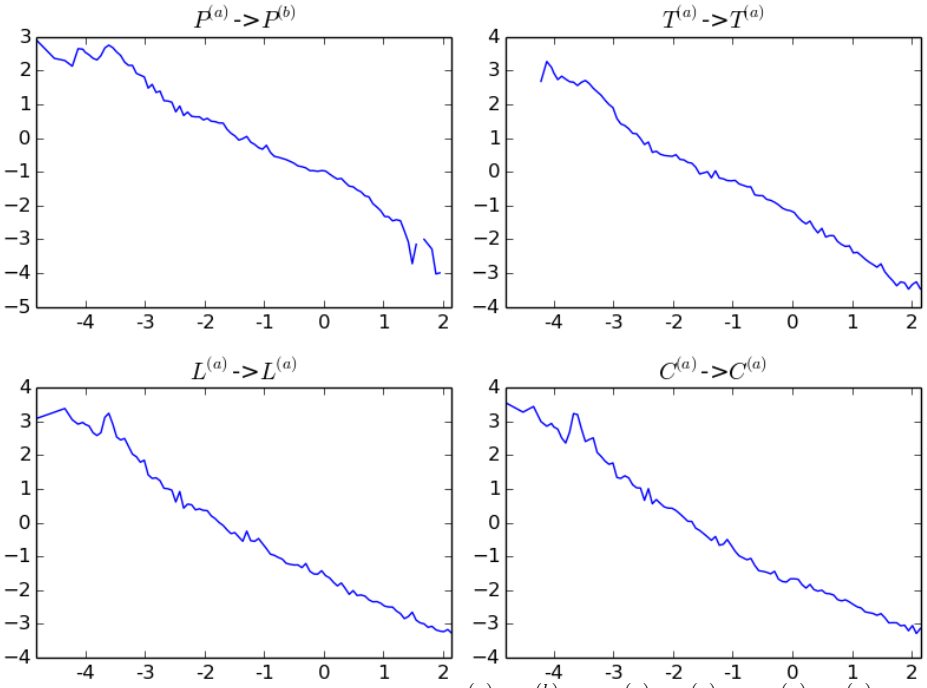


Figure 6: $\log_{10} - \log_{10}$ plots of $\phi^{P^{(a)} \to P^{(b)}}$, $\phi^{T^{(a)} \to T^{(a)}}$, $\phi^{L^{(a)} \to L^{(a)}}$ and $\phi^{C^{(a)} \to C^{(a)}}$. These kernels display loose power-law behavior with an exponent close to 1. For illustration purposes, fits were performed on the entire time-scale range, though they clearly display different regime.

Let us stress that this "roughly power law behavior" is not true for all kernels. For example, the kernels $\phi^{T^{(a)} \to P^{(a)}}$ and $\phi^{P^{(a)} \to L^{(a)}}$ are "localised" below the millisecond, see Figure 7 and the cumulated kernels in Appendix B. Empirical kernels have a much "richer" behavior than conditional laws which as stated earlier all have the same stylized facts.

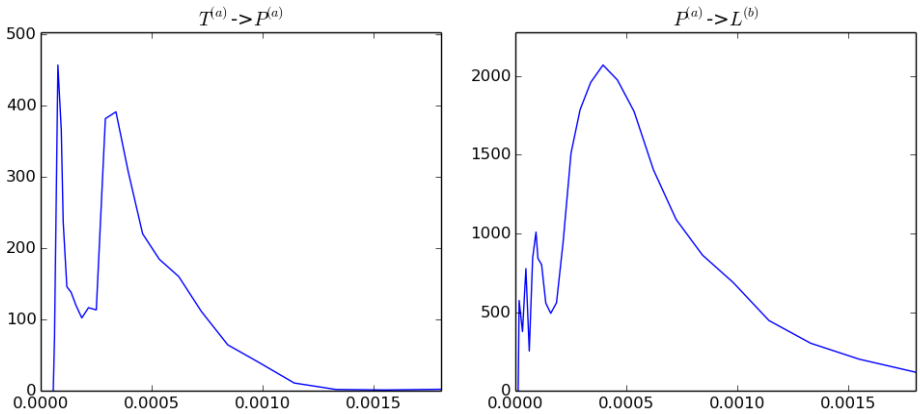


Figure 7: $\phi^{T^{(a)}\to P^{(a)}}$ and $\phi^{P^{(a)}\to L^{(a)}}$ as a function of time below the millisecond.

4.8 The impact of the order flows on the price

For the sake of clarity, we split the two 8×8 matrices of kernels into 2×2 sub-matrices. This representation of the kernels is provided in Appendix B for both DAX and BUND. Each sub-plot represents the normalized cumulated kernels $\frac{\Lambda^j}{\Lambda^i}\int_0^t \phi^{ij}(u)du$ as a function of $\log_{10}(t)$ (so there are 2×64 of them). This normalization for the kernel is natural in the population dynamics interpretation of Hawkes processes, see [22]. Indeed, $\frac{\Lambda^j}{\Lambda^i}||\phi^{ij}||$ corresponds to the proportion of i whose parent is a j (while $||\phi^{ij}||$ corresponds to the average number of children of type i for an individual of type j).

In this section, we focus on the impact of the flow of orders on the price, i.e., on the events which do not instantaneously move the price and that have an impact on mid-price variations. It corresponds to Paragraphs B.5, B.6 and B.7 in appendix.

One can see that the trades have a very localized delayed *self* impact on price variations (by *self* we mean *in the same direction*, e.g., an ask market order implies an eventual upward price move). Indeed, it seems that it takes 0.3 milliseconds for the market to incorporate the information of the trade. This is the averaged time for other agents to react to this order (this corresponds to the "latency" delay we have observed on the conditional laws in Section 4.3 and the kernels in Section 4.7). After this short time scale, the kernel is negligible, that is the cumulated kernel remains constant, and the trade does not imply any more price changes.

As far as limit and cancel orders are concerned, their impacts appear to be less localized and to be different when one compares the DAX (small tick) and the BUND (large tick). For the DAX, limit orders impact the price

by exiting the price moves in the opposite direction (limit orders at the bid trigger upward price moves and conversely) and by slightly inhibiting the price moves in the same direction. For the BUND, limit orders mainly have an inhibitory effect one price changes (limit orders at the bid inhibit downward price moves). As opposed to the impact of market orders, the impact of limit orders is not very fast. While the information of market orders is almost immediate (0.3 milliseconds), a limit order needs time to be significant.

Cancel orders impact the price in a very similar way as limit orders but in the opposite direction.

4.9 The impact of price jumps on the order flows

The kernels that correspond to the impact of price jumps on the order flows correspond to Paragraphs B.8, B.9 and B.10 where the normalized cumulated kernels[8] are plotted.

The most striking feature is that the effect of a mid-price change on market orders strongly depends on the tick size (as it is also observed on Figure 4). For the DAX (small tick), this effect is rather small and *self-exciting* (upward price moves imply ask trades). On the contrary, for the BUND (large tick), this effect is far more important and is *cross-exciting* (upward price moves trigger bid trades and inhibit ask trades). This can be interpreted in terms of adverse selection: On the one hand, for large tick assets, after an upward price move, agents will not want to execute market orders at the new ask price because this price will be too high. On the other hand , the new bid price will be more interesting and agent will thus execute bid market orders.

On both assets, the influence of price moves on the liquidity is mostly *self* in the sense that upward price moves excite bid limit orders and ask cancel orders (and inhibit ask limit orders and bid cancel orders). The excitations are mostly localized (around 0.3 milliseconds) and correspond to the averaged time reactions of market makers to the new information while the inhibition kernels are long term. One can also notice that for the DAX, the short term exciting effect of say $\phi^{P^{(a)} \to C^{(a)}}$ is balanced by a longer term inhibitory effect (around 10 seconds). We believe that this effect is linked to the reversion of the liquidity to its stationary state a "long" time after the price change has occurred.

[8] See beginning of the previous section or of Appendix B

4.10 The impact of market orders on liquidity and vice-versa

The kernels that correspond to the impact of the market order flow on limit and cancel order flows are in Paragraphs B.11 and B.12 where the normalized cumulated kernels are plotted.

For both assets, trades have very fast positive influence on the opposite limit orders. This is due to the fact that the underlying "efficient" price has moved and the liquidity must adapt itself. At longer time scales, the effect is reversed (the cross kernels become negative, i.e., trades inhibit cross limit orders). This is due to the reversion of the liquidity towards its stationary state. As far as the influence of market orders on the rate of cancel orders is concerned, for the BUND, trades imply cancel orders in the same direction at very short time scale and then, at longer time scales, this influence becomes negative. This effect is very small which means that the proportion of limit and cancel orders directly implied by a trade is small. However, we see on Figure 5 that when we "dress" this impact with intermediate events, exogenous trades indirectly imply a significant proportion of limit and cancel orders.

Concerning the reverse impact of limit and cancel order flows on market order flow, the corresponding kernels are displayed in Paragraphs B.13 and B.14. One can see that ask (resp. bid) cancel and bid (resp. ask) limit orders excite the trading rate at the ask (resp. the bid).

4.11 The impact of the limit order flow on the cancel order flow and vice-versa

The kernels that correspond to the dynamics between limit and cancel order flows correspond to Paragraphs B.3, B.4, B.15 and B.16 which display the corresponding normalized cumulated kernels.

As for market orders, the limit/limit kernels and the cancel/cancel kernels are mostly self-exciting (e.g., $\phi^{L^{(a)}\to L^{(a)}}$ is dominant and $\phi^{L^{(a)}\to L^{(b)}}$ is negligible). Remark also that cancel orders have a short term influence on cross limit orders ($\phi^{C^{(a)}\to L^{(b)}}$ is localized around 0.3 milliseconds) which corresponds to the information of the cancel order and a longer term impact on self limit orders ($\phi^{C^{(a)}\to L^{(a)}}$ is a power law) which corresponds to the return of the liquidity.

In a similar way, limit orders have a short term influence on cross cancel orders and a longer term impact on self cancel orders.

5 Discussion and concluding remarks

We have presented a numerical solution of the Wiener-Hopf Equation (8) when some kernels behave as power laws of exponents slightly higher than one and are thus significant over a very wide range of time scales (from 100 microseconds to 100 seconds!) and in a rather large number of dimensions. This lead us to an efficient algorithm for non-parametric estimation of these kernels.

Using the natural causal interpretation of Hawkes processes, we applied this algorithm to the study of high frequency financial data (timestamped with a precision of a micro-second). It allowed us to disentangle the cross-influences between eight types of first limit events occurring in order books.

Our approach allowed us to retrieve some well known stylized facts about market dynamics:

- The order flows are strongly self excited: The main influence of market, limit and cancel orders is on themselves. This is linked to the well known persistence of order flows and to the splitting of meta-orders into sequences of orders. Moreover, orders also have an influence of orders of the same "direction". For example, ask trades excite bid limit orders and ask cancel orders. The same kind of behavior had already been observed in [13] by directly looking at the correlation functions between these events.

- The prices are efficient: To balance the strong persistence of the order flow imbalance, the influence of price changes on themselves is mostly cross. That is upward price moves cause downward price moves. This is consistent the measure of the "bare propagators" of [13]. Indeed, in [13], the bare impact of events on the price is shown to decrease in time. Not however that this bare impact is not exactly equivalent to our kernels since it does not fully "disentangle" the influences of all the events on all the events but only the influence of orders on price changes.

- The orders impact the price even if they do not move it mechanically. Again, this is consistent with the measures of the propagators, of [13], of the events that do not move the price.

The generality of our Hawkes model allowed us to account for much richer dynamics than previous works and to describe and quantify the influences between all types of events. We have thus found some new and more subtle results that to our knowledge had never been observed.

- Price moves have a retro influence on orders flows: These are the kernels that the tick size influences the most. For large tick assets,

adverse selection prevails and this effect is cross: upward price moves excite bid trades and vice versa. For small tick assets, this effect is self because of the persistence of order flows.

- The Hawkes framework and our estimation of the kernels allows us not only to have a measure of "causality" between the different events but also of the time scales at which this causality appears. We have found that there are loosely two kinds of influences: Fast and localized influences, for which the volume of the kernel is localized around the reaction time of the market (on our data 0.3 milliseconds) and influences which are spread over a wide range of time scales and whose corresponding kernels thus behave somehow as power law functions of exponents slightly higher than one.

- Note also that as opposed to [13], our model is a real time model and not a discrete time model where each period corresponds to an event. This allows us for example, to see that market and limit orders have different impacts. Market orders have localized impacts while limit orders take time to have an impact.

- Finally, we can measure the endogeneity of all market events. We find in particular that market orders are more exogenous than limit and cancel orders. Therefore, when looking at the "dressed kernels" ψ, we find that the trades are leading in the sense that, exogenous trades have more influence than exogenous limit and cancel orders.

One thing that probably lacks in our purely Hawkes order book model is that the arrival of events does not directly depend on the state of the order book but only on the past events. For example, it is not clear that in our model, when the spread is important, liquidity will arrive. On the other side of the spectrum of order book models, we can mention [16, 24] where the order book dynamics are purely Markovian and thus do not depend on the history of the order flow. We believe that "reality" is between these two approaches: the intensity of market events depends both on their history and on the state of the order book.

The stability and tractability of our numerical estimation method over extremely wide ranges of time scales and as the dimension of the model increases imply that we have many leads to improve our order book model. For example, we can add marks to account for the size of the different events, see [5] or exogenous terms to account for external news, see [33].

We conclude by underlying that of course, the relevance of Hawkes models is not restricted to finance. Therefore, our procedure can be applied to the study of any field where the measure of complex relations between point processes is necessary.

Acknowledgements

We thank M. Rosenbaum for useful discussions. This research benefited from the support of the Chair Markets in Transition, under the aegis of Louis Bachelier Finance and Sustainable Growth laboratory, a joint initiative of Ecole Polytechnique, Université d'Evry Val d'Essonne and Fédération Bancaire Française.

A Multiscale estimation of the conditional laws

The aim of this appendix is to present our multi scale estimation procedure of the conditional law g defined by (2.3).

A.1 Estimation procedure of g

Let us assume that, for all (i,j), we have the times of the events i $(T_k^i)_{k\leq n_i}$ and j $(T_k^j)_{k\leq n_j}$ and we want an empirical estimation of the conditional laws g^{ij}.

To do that, we will choose a time grid $(t_l)_{l\leq n}$ small enough so that we can approximate $g^{ij}(\frac{t_l+t_{l+1}}{2})$ by $\frac{1}{t_{l+1}-t_l}\int_{t_l}^{t_{l+1}} g^{ij}(s)ds$ but large enough so that there is enough points (k,k') such that $T_k^i - T_{k'}^j \in [t_l, t_{l+1}]$ to have a good approximation of $\int_{t_l}^{t_{l+1}} g^{ij}(s)ds$ by $\frac{1}{n_j}\sum_{k=1}^{n_i}\sum_{k'=1}^{n_j} \mathbb{1}_{T_k^i - T_{k'}^j \in [t_l,t_{l+1}]}$. We thus approximate $g^{ij}(\frac{t_l+t_{l+1}}{2})$ by

$$\hat{g}^{ij}(\frac{t_l+t_{l+1}}{2}) = \frac{1}{t_{l+1}-t_l}\frac{1}{n_j}\sum_{k=1}^{n_i}\sum_{k'=1}^{n_j} \mathbb{1}_{T_k^i - T_{k'}^j \in [t_l,t_{l+1}]}$$

and we take a piecewise affine interpolation of $\hat{g}^{i,j}$ between the points $\frac{t_l+t_{l+1}}{2}$.

A.2 Choice of the grid

The first natural grid that one can take is a uniform grid: $[0, h, 2h, ..., h_{max}]$. However, if one does that, one cannot get a good estimation of the conditional law on a large range of time scales. For example, if one takes $h = 0.0001s$ to have a good estimation of the conditional law at low time scales, one cannot estimate get a good estimation around one second. Indeed, there will be very few points of i between 1 and 1.0001 seconds after a point of j.

To solve this problem, we will consider a grid that is uniform between 0 and h_{min} and log-uniform between h_{min} and h_{max}:

$$[0, h_{min}h_\delta, h_{min}2h_\delta, ..., h_{min}, h_{min}e^{h_\delta}, h_{min}e^{2h_\delta}, ..., h_{max}]. \tag{16}$$

Doing this, we will have enough points between two points of the grid but the claw will not vary too much between two points of the grid.

B Estimation results for the full order book model: the cumulated kernel matrix $\int_0^t \phi(s)ds$

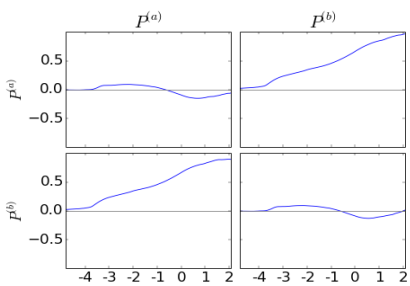
In this Appendix, we represented the resulting kernels estimated on the high frequency database (see Section 4). For each i and j in

$$\{P^{(a)}, P^{(b)}, T^{(a)}, T^{(b)}, L^{(a)}, L^{(b)}, C^{(a)}, C^{(b)}\}$$

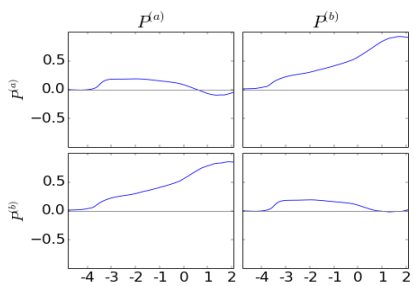
we have represented the estimated normalized cumulated kernels $\frac{\Lambda^j}{\Lambda^i}\int_0^t \phi^{i\to j}(u)du$ as a function of $\log_{10}(t)$. This normalization for the kernel is natural because in the population dynamics interpretation of Hawkes processes $\frac{\Lambda^j}{\Lambda^i}||\phi^{i\to j}||_1$ corresponds to the proportion of i whose parent is a j (while $||\phi^{i\to j}||_1$ corresponds to the average number of children of type i for an individual of type j).

There are 2×64 plots that are organized by assets and by types of events.

B.1 Influence of price moves on price moves

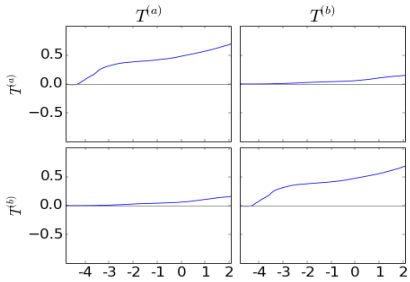


(a) xFDAX

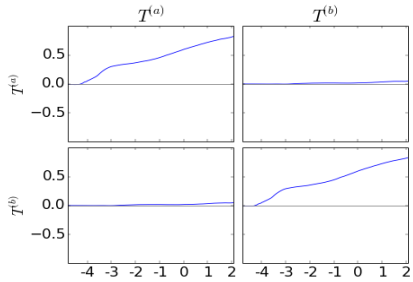


(b) xFGBL

B.2 Influence of trades on trades

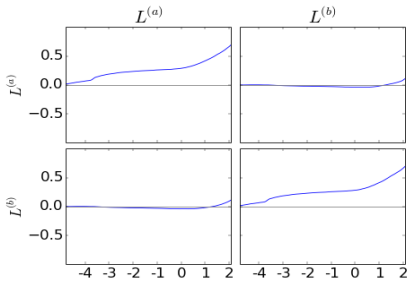


(a) xFDAX

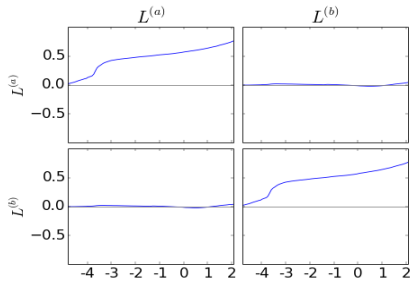


(b) xFGBL

B.3 Influence of limit orders on limit orders

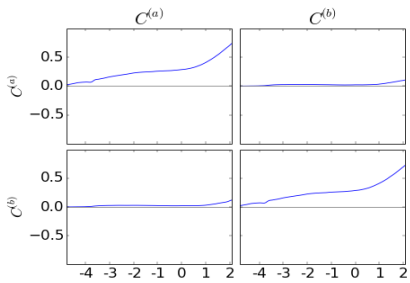


(a) xFDAX

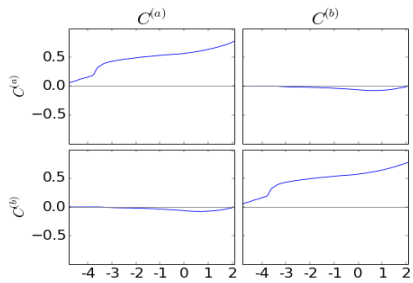


(b) xFGBL

B.4 Influence of cancel orders on cancel orders

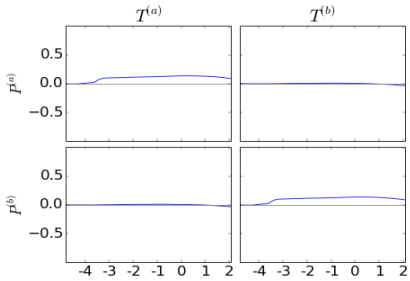


(a) xFDAX

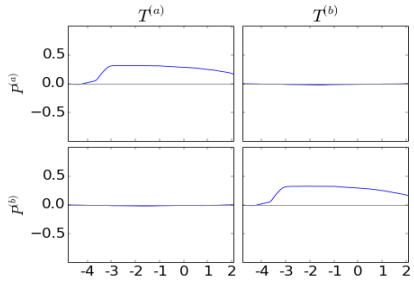


(b) xFGBL

B.5 Influence of trades on price changes

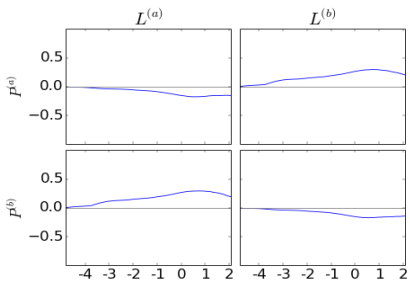


(a) xFDAX

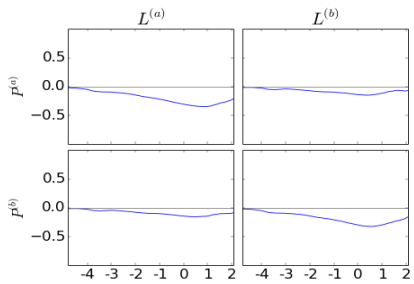


(b) xFGBL

B.6 Influence of limit orders on price changes

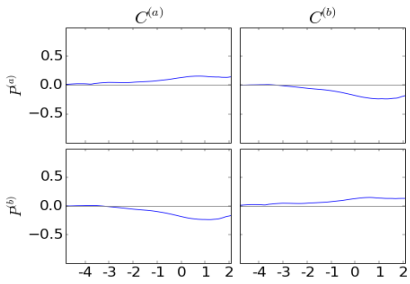


(a) xFDAX

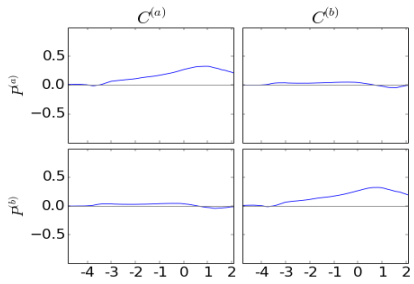


(b) xFGBL

B.7 Influence of cancel orders on price changes

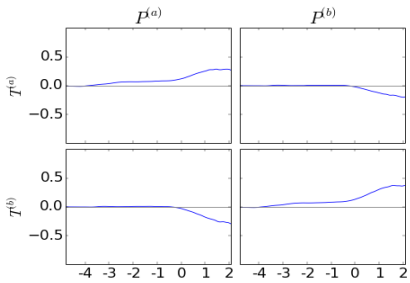


(a) xFDAX

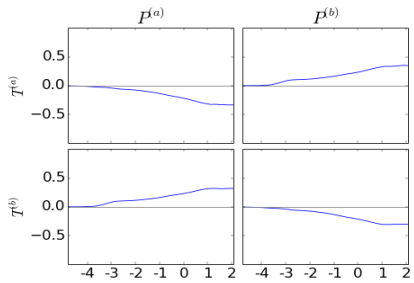


(b) xFGBL

B.8 Influence of price changes on trades

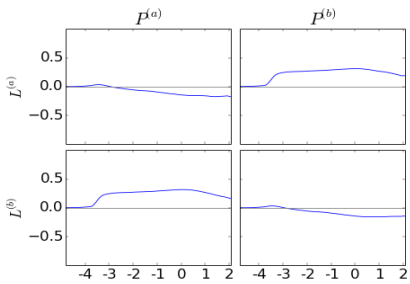


(a) xFDAX

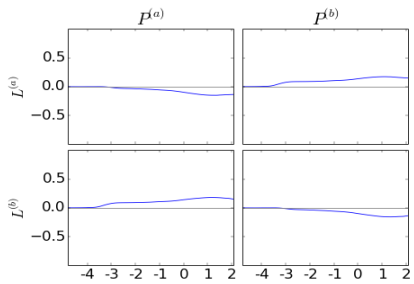


(b) xFGBL

B.9 Influence of price changes on limit orders

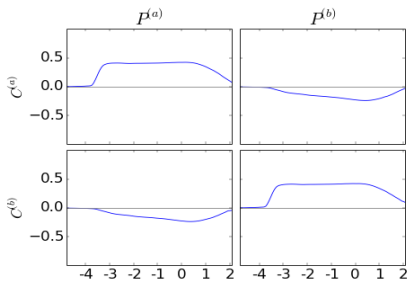


(a) xFDAX

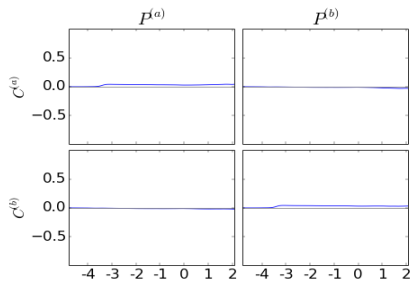


(b) xFGBL

B.10 Influence of price changes on cancel orders

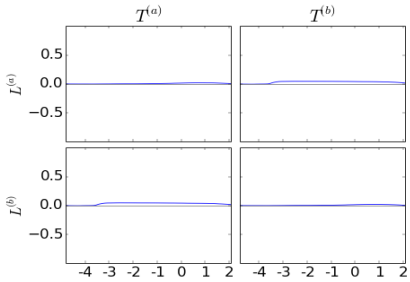


(a) xFDAX

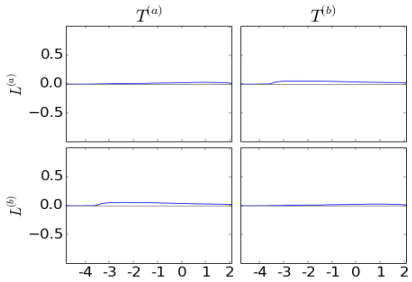


(b) xFGBL

B.11 Influence of trades on limit orders

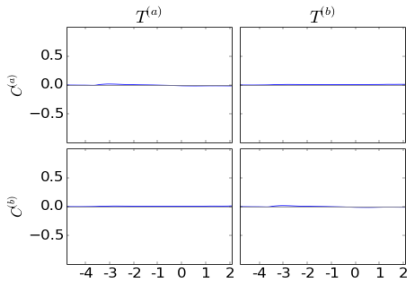


(a) xFDAX

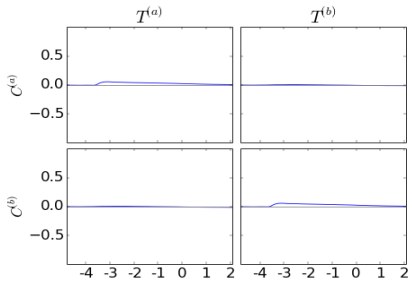


(b) xFGBL

B.12 Influence of trades on cancel orders

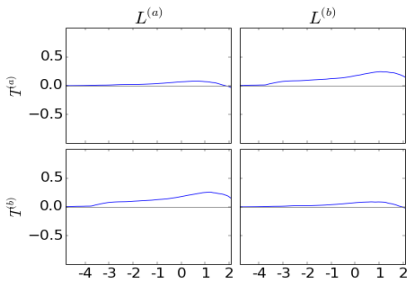


(a) xFDAX

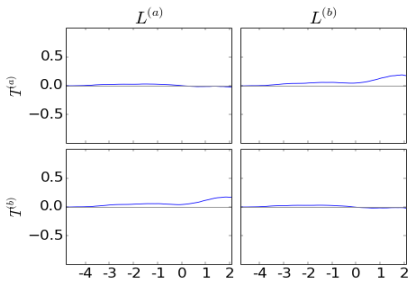


(b) xFGBL

B.13 Influence of limit orders on trades

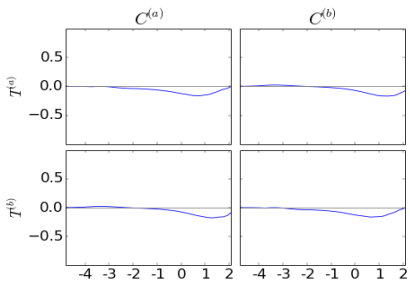


(a) xFDAX

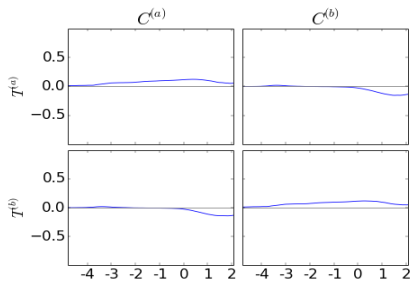


(b) xFGBL

B.14 Influence of cancel orders on trades

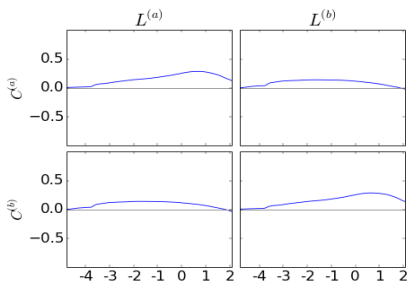


(a) xFDAX

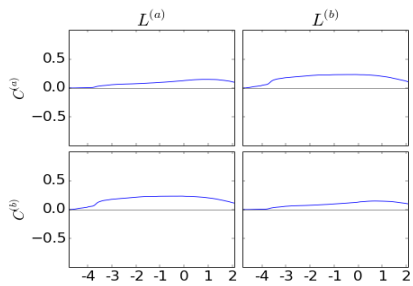


(b) xFGBL

B.15 Influence of limit orders on cancel orders

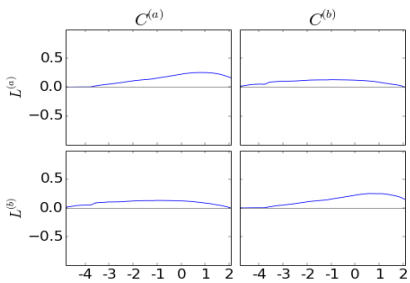


(a) xFDAX

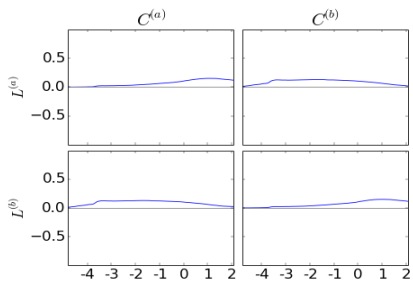


(b) xFGBL

B.16 Influence of cancel orders on limit orders



(a) xFDAX



(b) xFGBL

References

[1] Y. Aït-Sahalia, J. Cacho-Diaz, and R. J. Laeven. Modeling financial contagion using mutually exciting jump processes. Technical report, National Bureau of Economic Research, 2010.

[2] E. Bacry, K. Dayri, and J.-F. Muzy. Non-parametric kernel estimation for symmetric Hawkes processes. application to high frequency financial data. *The European Physical Journal B*, 85(5):1–12, 2012.

[3] E. Bacry, S. Delattre, M. Hoffmann, and J.-F. Muzy. Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, 13(1):65–77, 2013.

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

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

[6] J.-P. Bouchaud, J. D. Farmer, and F. Lillo. How markets slowly digest changes in supply and demand. *Handbook of financial markets: dynamics and evolution*, 1:57, 2009.

[7] J.-P. Bouchaud, Y. Gefen, M. Potters, and M. Wyart. Fluctuations and response in financial markets: the subtle nature of "random" price changes. *Quantitative Finance*, 4(2):176–190, 2004.

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

[9] P. Brémaud and L. Massoulié. Stability of nonlinear Hawkes processes. *Annals of Probability*, 24(3):1563–1588, 1996.

[10] R. Cont and A. de Larrard. Price dynamics in a Markovian limit order book market. *SIAM Journal of Financial Mathematics*, 4:1–25, 2013.

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

[12] K. Dayri and M. Rosenbaum. Large tick assets: implicit spread and optimal tick size. *arXiv preprint arXiv:1207.6325*, 2012.

[13] Z. Eisler, J.-P. Bouchaud, and J. Kockelkoren. The price impact of order book events: market orders, limit orders and cancellations. *Quantitative Finance*, 12(9):1395–1419, 2012.

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

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

[16] A. Gareche, G. Disdier, J. Kockelkoren, and J.-P. Bouchaud. Fokker-Planck description for the queue dynamics of large tick stocks. *Physical Review E*, 88(3):032809, 2013.

[17] I. Gokhberg and M. G. Krein. Systems of integral equations on the half-line with kernels depending on the difference of the arguments. *Uspekhi matematicheskikh nauk*, 13(2):3–72, 1958.

[18] N. R. Hansen, P. Reynaud-Bouret, and V. Rivoirard. Lasso and probabilistic inequalities for multivariate point processes. *arXiv preprint arXiv:1208.0570*, 2012.

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

[20] A. G. Hawkes. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 438–443, 1971.

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

[22] A. G. Hawkes and D. Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, pages 493–503, 1974.

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

[24] W. Huang, C.-A. Lehalle, and M. Rosenbaum. Simulating and analyzing order book data: The queue-reactive model. *arXiv preprint arXiv:1312.0563*, 2013.

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

[26] T. Jaisson and M. Rosenbaum. Limit theorems for nearly unstable Hawkes processes. *arXiv preprint arXiv:1310.2033*, 2013.

[27] T. Jaisson and M. Rosenbaum. Fractional diffusions as scaling limits of nearly unstable heavy-tailed Hawkes processes. *Working paper*, 2014.

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

[29] F. Lillo, S. Mike, and J. D. Farmer. Theory for long memory in supply and demand. *Physical Review E*, 71(6):066122, 2005.

[30] 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), 2011.

[31] E. J. Nyström. Über die praktische auflösung von integralgleichungen mit anwendungen auf randwertaufgaben. *Acta Mathematica*, 54(1):185–204, 1930.

[32] Y. Ogata. Seismicity analysis through point-process modeling: A review. *Pure and applied geophysics*, 155(2-4):471–507, 1999.

[33] M. Rambaldi, P. Pennesi, and F. Lillo. Modeling FX market activity around macroeconomic news: A Hawkes process approach. *arXiv preprint arXiv:1405.6047*, 2014.

[34] P. Reynaud-Bouret, V. Rivoirard, F. Grammont, and C. Tuleau-Malot. Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *The Journal of Mathematical Neuroscience (JMN)*, 4(1):1–41, 2014.

[35] I. Roşu. A dynamic model of the limit order book. *Review of Financial Studies*, page hhp011, 2009.

[36] I. H. Sloan. Quadrature methods for integral equations of the second kind over infinite intervals. *Mathematics of Computation*, 36(154):511–523, 1981.

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