Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data.

Emmanuel Bacry[1], Khalil Dayri[1], and Jean-Francois Muzy[1,2] [a] [b].

[1] CMAP UMR 7641 CNRS, Ecole Polytechnique, 91128 Palaiseau, France
[2] SPE UMR 6134 CNRS, Université de Corse, 20250 Corte, France

December 9, 2011

Abstract. We define a numerical method that provides a non-parametric estimation of the kernel shape in symmetric multivariate Hawkes processes. This method relies on second order statistical properties of Hawkes processes that relate the covariance matrix of the process to the kernel matrix. The square root of the correlation function is computed using a minimal phase recovering method. We illustrate our method on some examples and provide an empirical study of the estimation errors. Within this framework, we analyze high frequency financial price data modeled as 1D or 2D Hawkes processes. We find slowly decaying (power-law) kernel shapes suggesting a long memory nature of self-excitation phenomena at the microstructure level of price dynamics.

1 Introduction

Although the concept of self excitement was commonly used for a long time by seismologists (see [30] and the references therein), Alan Hawkes was the first to provide a well defined point process with a self exciting behavior [13, 14,19]. This model was introduced to reproduce the ripple effects generated after the occurrence of an earthquake [31,1]. Since then, it has been successfully used in many areas ranging from seismology (see e.g., [23], for a recent review), biology [28], to even criminology and terrorism [21,10] (cf [19] and references therein for a detailed review of Hawkes processes and their applications). As far as financial applications are concerned, since transactions and price changes are discrete events, Hawkes processes have naturally been generating more and more interest. Applications can be found in the field of order arrival rate modeling [16,7,29], noise microstructure dynamics [2], volatility clustering [11], extreme values and VaR estimation [8] and credit modeling [12].

Hawkes models account for a self exciting behavior of events by which the arrival of one event increases the probability of occurrence of new ones. In its most basic form and in the one dimensional case, the Hawkes process is a counting process defined by λ_t, the rate of arrival of events

by:

$$\lambda_t = \mu + \int_{-\infty}^{t} \phi_{t-s} dN_s \qquad (1)$$

where $\mu > 0$ is a constant *background* rate, N_t is the cumulative counting process and ϕ a positive real function called *decay kernel*. We can clearly see in equation (1) that when some event occurs at time t, we have $dN_t = 1$ and hence $d\lambda_t = \phi_0$. The influence of the event is transmitted to future times through ϕ such that at time $u > t$ the increase in λ_t due to the time t event is ϕ_{u-t}. Thus a self exciting behavior is observed.

A basic issue in applications of Hawkes processes concerns their estimation. In early applications, parametric estimation used spectral analysis by means of the Bartlett spectrum (the Fourier transform of the autocovariance of the process) [4,5] and it was indeed through that light that Hawkes presented his model. A maximum likelihood method for estimating the parameters of exponential, power law and Laguerre class of decay kernels was developed in [26,24] and it became the standard method for estimating Hawkes processes. Furthermore, the Laguerre decay kernels were seen to be very pertinent decay kernels because they allowed to account for long term dependencies as well as offering short term flexibility. More recently, other types of estimation procedures were developed. When the form of the decay kernel is unknown, non-parametric methods are desirable because they give an idea of their general shape. For example, by using the branching property of the Hawkes process [32], the authors in [20] and [18] were able to provide Expectation-Maximization algorithms to

estimate both background rate and the decay kernel. In [28], the authors present also an algorithmic method for decay estimation by using a penalized projection method.

In this paper, we propose an alternative simple non parametric estimation method for multivariate symmetric Hawkes processes based on the Bartlett spectrum. We present the method and its numerical feasibility without going too much in details about convergence speeds or error optimization. By studying one dimensional and 2-dimensional examples, we show that our approach provides reliable estimates on both fast and slowly decaying kernels. We then apply our method to high frequency financial data and find power-law kernels. This implies that the arrival of events display long range correlations, a phenomenon well known in finance, similarly to what was suggested in [6].

This paper is organized as follows. In section 2, we introduce a basic version of a multivariate Hawkes processes. We place ourselves in the context of an $n-$dimensional linear Hawkes process with a constant background rate and a nonnegative decay kernel. We set some notations and give out a martingale representation of the rate function λ_t in Eq. (1). In section 3, we study the autocovariance of the process. Along the same line as [14], we establish its relationship with the decay kernel in both direct and Fourier spaces. The estimation method in the case of symmetric Hawkes processes is then provided in section 4. This method, based on a Hilbert transform phase recuperation method, is explicitly detailed in both univariate and special symmetric bivariate cases. In section 5, the method is illustrated by numerical examples for both exponential and power law kernels. We also address some statistical issues concerning the estimation errors. Finally, in section 6, we apply our method to high frequency financial data for which we deduce a long range nature of the decay kernels.

2 Multivariate Hawkes Processes

2.1 Notations and Definitions

As introduced by Hawkes in [13] and [14], let us consider an n-dimensional point process $(N_t)_{t\geq 0}$, where N_t^i $1 \leq i \leq n$ represents the cumulative number of events in the ith component of the process N_t up to time t.

The conditional intensity vector at any time t is assumed to be a random process that depends on the past as

$$\lambda_t = \mu + \int_{-\infty}^{t} \phi_{t-s} dN_s \qquad (2)$$

where μ is a vector of size n with strictly positive components ($\mu^i > 0$) and ϕ_t is an $n \times n$ matrix referred to as the *decay kernel*.

Notations 1 *In the following*

– $\mathcal{M}_{n,p}(\mathbb{R})$ *(resp. $\mathcal{M}_{n,p}(\mathbb{C})$) denotes the set of $n \times p$ matrices with values in $\mathbb{R}$ (resp. $\mathbb{C}$). For any matrix M (resp. vector v), M^{ij} (resp. v^i) denotes its elements.*

– *For any function f_t, $\widehat{f}_z = \int_{\mathbb{R}} e^{-zt} f_t dt$ corresponds to its Laplace transform.*

– *By extension if $M_t \in \mathcal{M}_{n,p}(\mathbb{R})$ then $\widehat{M}_z \in \mathcal{M}_{n,p}(\mathbb{C})$ corresponds to the matrix whose elements are the Laplace transforms of the elements of M_t.*

Using these notations, one of the main results of Hawkes is that if

H1 the kernel $\phi_t \in \mathcal{M}_{n,n}$ is positive ($\phi_t^{ij} \geq 0$, $\forall t$) and causal ($\phi_t = 0$, $\forall t < 0$),

and

H2 the spectral radius of $\widehat{\phi}_0$ (i.e., its largest eigen value) is strictly smaller than 1,

then $(N_t)_{t\geq 0}$ is a n-dimensional point process with stationary increments. The conditional intensity λ_t is itself a stationary process with mean

$$\Lambda = E(\lambda_t) = E(dN_t)/dt. \qquad (3)$$

Combining this last equation with Eq. (2), one easily gets $\Lambda = \mu + \int_{-\infty}^{t} \phi_{t-s} ds \Lambda = \mu + \int_{0}^{\infty} \phi_u du \Lambda$ and consequently

$$\Lambda = (\mathbb{I} - \widehat{\phi}_0)^{-1} \mu, \qquad (4)$$

where $\mathbb{I}$ refers to the $n \times n$ identity matrix.

Before moving on, we need to introduce some more notations that will be used all along the paper.

Notations 2 *If $A_t \in \mathcal{M}_{m,n}(\mathbb{R})$ and $B_t \in \mathcal{M}_{n,p}(\mathbb{R})$ then the convolution product of A_t and B_t is naturally defined as $A \star B_t = \int_{\mathbb{R}} A_s B_{t-s} ds = \int_{\mathbb{R}} A_{t-s} B_s ds$. Of course it is associative and distributive however it is generally not commutative (unless A_t and B_t are commutative). The neutral element is $\delta \mathbb{I}_t$, i.e., the diagonal matrix with Dirac distribution on the diagonal. In the following we will use the notation :*

$$A \star dN_t = \int_{\mathbb{R}} A_{t-s} dN_s.$$

Combining both Notations 1 and 2, it is easy to show that the convolution theorem on matrices translates in

$$\widehat{A \star B}_z = \widehat{A}_z \widehat{B}_z.$$

2.2 Martingale representation of λ_t

We roughly follow a similar path to Hawkes in [14]. Let $(M_t)_{t\geq 0}$ be the martingale compensated process of $(N_t)_{t\geq 0}$ defined by:

$$dM_t = dN_t - \lambda_t dt . \qquad (5)$$

Then λ_t can be represented as a stochastic integral with respect to the martingale $(M_t)_{t\geq 0}$:

Proposition 1 *One has:*

$$\lambda_t = \Lambda + \Psi \star dM_t, \qquad (6)$$

where Ψ_t is defined as

$$\Psi_t = \sum_{n=1}^{\infty} \phi_t^{(\star n)}, \tag{7}$$

where $\phi_t^{(\star n)}$ refers to the nth auto-convolution of ϕ_t (i.e., $\widehat{\phi^{(\star n)}}_z = (\widehat{\phi}_z)^n$)

Proof Using equations (2) and (5) one has:

$$\lambda_t = \mu + \phi \star dN_t = \mu + \phi \star dM_t + \phi \star \lambda_t,$$

and consequently

$$(\delta \mathbb{I} - \phi) \star \lambda_t = \mu + \phi \star dM_t.$$

Let us note that the inverse of $\delta \mathbb{I}_t - \phi_t$ for the convolution product is nothing but $h_t = \delta \mathbb{I}_t + \Psi_t$. Thus convoluting on each side of the last equation by h_t, one gets (since $h \star \phi_t = \Psi_t$)

$$\lambda_t = h \star \mu + \Psi \star dM_t.$$

Since μ is a constant one has $h \star \mu = \mu \widehat{h}_0$. Using the convolution theorem one gets $\mu \widehat{h}_0 = (\mathbb{I} - \widehat{\phi}_0)^{-1} \mu = \Lambda$ which proves the proposition.

3 The covariance matrix of Hawkes processes

The kernel estimator we are going to build is based on the empirical auto-covariance of $(N_t)_{t \geq 0}$. This section is devoted to the covariance matrix of the n-dimensional Hawkes processes. However, we first discuss some useful results about their *infinitesimal* auto-covariance function.

3.1 The infinitesimal covariance

Let us define the infinitesimal covariance matrix:

$$\nu_{t-t'} = E(dN_t dN_{t'}^{\dagger}),$$

where $M^{\dagger}$ denotes the hermitian conjugate of a matrix M. Along the same way as Hawkes in [14], $\nu_{t-t'}$ can be related to the decay kernel ϕ. We present in proposition 2 an equation linking the infinitesimal covariance matrix to Λ and Ψ but, unlike Hawkes, we express ν explicitly as a function of Ψ and Λ instead of an integral equation in ν. This result will be at the heart of the estimation method we propose in this paper.

Proposition 2 *Let $(N_t)_{t \geq 0}$ be an n-dimensional Hawkes process with intensity λ_t as defined in Section 2.1 (assuming both **H1** and **H2**). Let $\widetilde{\Psi}_t = \Psi_{-t}$. We have the following result:*

$$E(dN_t dN_{t'}^{\dagger}) = \left(\Lambda \Lambda^{\dagger} + \Sigma \delta_{t-t'} + \Psi_{t-t'} \Sigma + \Sigma \Psi_{t'-t}^{\dagger} + \widetilde{\Psi} \star \Sigma \Psi_{t'-t}^{\dagger} \right) dt dt' \tag{8}$$

where Σ is the diagonal matrix defined by $\Sigma^{ii} = \Lambda^i$ for all $1 \leq i \leq n$ and δ_t is the Dirac distribution.

Proof By using equation (5), we can write:

$$E(dN_t dN_{t'}^{\dagger}) = E((dM_t + \lambda_t dt)(dM_{t'} + \lambda_{t'} dt')^{\dagger})$$

$$= E(dM_t dM_{t'}^{\dagger}) \tag{9}$$

$$+ E(\lambda_t dM_{t'}^{\dagger}) dt \tag{10}$$

$$+ E(dM_t \lambda_{t'}^{\dagger}) dt' \tag{11}$$

$$+ E(\lambda_t \lambda_{t'}^{\dagger}) dt dt' \tag{12}$$

We begin by noticing that (thanks to martingale property), $E(dM_t dM_{t'}^{\dagger}) = 0$, $\forall t \neq t'$. As for when $t = t'$, we have for all $1 \leq i < j \leq n$

$$E(dM_t^i dM_t^j) = 0$$

since the N^i's (and hence the M^i's) for $1 \leq i \leq n$ have no jump in common. Moreover, for $i = j$ we have

$$E(dM_t^i dM_t^i) = \Lambda^i dt$$

because $E(dM_t^i dM_t^i) = E(dN_t^i dN_t^i)$ and $E(dN_t^i dN_t^i) = E(dN_t^i) = \Lambda^i dt$ since the jumps of N_t are of size 1. To sum up, the term (9) becomes:

$$E(dM_t dM_{t'}^{\dagger}) = \Sigma \delta_{t-t'} dt dt' \tag{13}$$

The remaining terms can be then calculated along the same line. Replacing λ_t's expression from equation (6) in the term (10), gives:

$$E(\lambda_t dM_{t'}^{\dagger}) dt = E \left(\Psi \star dM_t dM_{t'}^{\dagger} \right) dt,$$

or in other words

$$E(\lambda_t dM_{t'}^{\dagger}) dt = \int_{\mathbb{R}} \Psi_{t-s} E(dM_s dM_{t'}^{\dagger}) dt ds$$

and thanks to equation (13),

$$E(\lambda_t dM_{t'}^{\dagger}) dt = \int_{\mathbb{R}} \Psi_{t-s} \Sigma \delta_{s-t'} ds dt' dt$$

that is simplified to:

$$E(\lambda_t dM_{t'}^{\dagger}) dt = \Psi_{t-t'} \Sigma dt dt' \tag{14}$$

Similarly, the term (11) becomes:

$$E(dM_t \lambda_{t'}^{\dagger}) dt' = \Sigma \Psi_{t'-t}^{\dagger} dt dt' \tag{15}$$

and finally, using Eq. (13), the term (12) can be written as:

$$E(\lambda_t \lambda_{t'}^{\dagger}) dt dt' = \left(\Lambda E(\lambda_{t'}^{\dagger}) + E((\Psi \star dM_t) \lambda_{t'}^{\dagger}) \right) dt dt'$$

$$= \left(\Lambda \Lambda^{\dagger} + \int_{\mathbb{R}} \Psi_{t-s} E(dM_s \lambda_{t'}^{\dagger} ds) \right) dt dt'$$

$$= \left(\Lambda \Lambda^{\dagger} + \int_{\mathbb{R}} \Psi_{t-s} \Sigma \Psi_{t'-s}^{\dagger} ds \right) dt dt'$$

By setting $u = t' - s$, we get:

$$E(\lambda_t \lambda_{t'}^{\dagger}) dt dt' = \left(\Lambda \Lambda^{\dagger} + \widetilde{\Psi} \star \Sigma \Psi_{t'-t}^{\dagger} \right) dt dt', \tag{16}$$

which ends the proof of the proposition.

3.2 The covariance matrix

The (normalized) covariance matrix of the Hawkes process can be defined, at scale h and lag τ, by

$$v_\tau^{(h)} = h^{-1}\text{Cov}\left(N_{t+h} - N_t, N_{t+h+\tau} - N_{t+\tau}\right) , \quad (17)$$

where we normalized by h in order to avoid a trivial scale dependence. Let us note that, since the increments of N_t are stationary, the previous definition does not depend on t. Thus, it can be rewritten as

$$v_\tau^{(h)} = \frac{1}{h}E\left((\int_0^h dN_s - \Lambda h)(\int_\tau^{\tau+h} dN_s^\dagger - \Lambda^\dagger h)\right) \quad (18)$$

It is clear that this quantity can be easily estimated on real data using empirical means. As we will see, the non parametric estimation of the decay kernel we propose in this paper is based on these empirical estimations. More precisely, it is based on the following Theorem.

Theorem 1 *Let $g_t^{(h)} = (1 - \frac{|t|}{h})^+$. $v_\tau^{(h)}$ can be expressed as a function of $g_\tau^{(h)}$ and Ψ_τ:*

$$v_\tau^{(h)} = g_\tau^{(h)}\Sigma + g^{(h)}\star\Psi_{-\tau}\Sigma + g^{(h)}\star\Sigma\Psi_\tau^\dagger + g^{(h)}\star\widetilde{\Psi}\star\Sigma\Psi_\tau^\dagger \quad (19)$$

Proof Let us begin the proof by noticing that for any function f with values in $\mathbb{R}^+$ we have:

$$h^{-1}\int_0^h\int_\tau^{\tau+h} f_{t-t'}dt'dt = f\star g_{-\tau}^{(h)} \quad (20)$$

It follows that

$$v_\tau^{(h)} = \frac{1}{h}E\left(\int_0^h dN_s\int_\tau^{\tau+h} dN_s^\dagger\right.$$
$$\left. - \int_0^h dN_s\Lambda^\dagger h - \Lambda h\int_\tau^{\tau+h} dN_s^\dagger + \Lambda\Lambda^\dagger h^2\right)$$
$$= \frac{1}{h}E\left(\int_0^h\int_\tau^{\tau+h} dN_t dN_{t'}^\dagger - \Lambda\Lambda^\dagger h^2\right) .$$

Using equation (8), we can split $dN_t dN_{t'}^\dagger$ into four parts and write:

$$v_\tau^{(h)} = \frac{1}{h}\int_0^h\int_\tau^{\tau+h}\left(\Sigma\delta_{t-t'}\right.$$
$$\left. + \Psi_{t-t'}\Sigma + \Sigma\Psi_{t'-t}^\dagger + \widetilde{\Psi}\star\Sigma\Psi_{t'-t}^\dagger\right)dtdt'$$

By applying equation (20) to each of the terms under the double integral we get equation (19) and achieve the proof.

It is more convenient to rewrite the result of theorem 3.2 in Laplace domain. Since $\widehat{\phi^{(\star n)}}_z = \widehat{\phi}_z^n$, Eq. (7) translates into :

$$\widehat{\Psi}_z = \sum_{n=1}^{+\infty}\widehat{\phi}_z^n = \widehat{\phi}_z(\mathbb{I} - \widehat{\phi}_z)^{-1} \quad (21)$$

and conversely:

$$\widehat{\phi}_z = (\mathbb{I} + \widehat{\Psi}_z)^{-1}\widehat{\Psi}_z \quad (22)$$

Theorem 3.2 gives way to the following corollary.

Corollary 1 *In Laplace domain equation (19) becomes:*

$$\widehat{v}_z^{(h)} = \widehat{g}_z^{(h)}(\mathbb{I} + \widehat{\Psi}_z^*)\Sigma(\mathbb{I} + \widehat{\Psi}_z^*)^\dagger \quad (23)$$

4 Non-parametric estimation of the kernel ϕ_t

4.1 The estimation principle

In this paper, we aim at building an estimator of ϕ_t based on empirical measurements of $v_\tau^{(h)}$. Let us note that empirical measurements of $v_\tau^{(h)}$ are naturally obtained replacing probabilistic mean by empirical mean in Eq. (18) (see [3] for proof of convergence of the empirical mean towards the probabilistic mean). Thus, in order to build an estimator, we need to express ϕ_t as a function of $v_\tau^{(h)}$. In the Laplace domain, since a given $\widehat{\psi}_z$ corresponds to a unique $\widehat{\phi}_z$ (through Eq. (22)), it translates in trying to express $\widehat{\psi}_z$ as a function of $\widehat{v}_z^{(h)}$ which exactly corresponds to inverting Eq. (23) (i.e., computing the square root of $(\mathbb{I} + \widehat{\Psi}_z^*)\Sigma(\mathbb{I} + \widehat{\Psi}_z^*)^\dagger$). Indeed, knowing $\widehat{\psi}_z$, one easily gets $\widehat{\phi}_z$ (from Eq. (22)) and finally ϕ_t. Actually, let us note that, from a practical point of view, we don't need to work in the full complex domain $z \in \mathbb{C}$ of the Laplace transform. Working with the Fourier transform restriction (i.e., $z = i\omega$ with $\omega \in \mathbb{R}$) is enough to recover ϕ_t.

Dealing with cancelations of $\widehat{g}_z^{(h)}$. The first problem that seems to appear for inverting this formula (23) (i.e., expressing $\widehat{\psi}_z$ as a function of $\widehat{v}_z^{(h)}$ only for $z = i\omega$) is the scalar term $\widehat{g}_z^{(h)}$ that may vanish. Indeed, since $g_t^{(h)} = (1 - \frac{|t|}{h})^+$, its Fourier transform, $\widehat{g}_{i\omega}^{(h)} = (4/\omega^2 h)\sin^2(\omega h/2)$, cancels for all ω of the form $\frac{2n\pi}{h}$, $n \in \mathbb{Z}$, $n \neq 0$. Actually, this is not a real problem as long as

- τ in the empirical estimation of $v_\tau^{(h)}$ is sampled using a sampling period Δ greater than $h/2$
- Δ is small enough so that $\widehat{v}_{i\omega}^{(h)}$ can be considered to have a compact support

Indeed, if this is the case, then the estimation of $\widehat{v}_{i\omega}^{(h)}$ (practically obtained by taking the Discrete Fourier Transform (DFT) of the sampled signal $(v_{k\Delta}^{(h)})_k$) will be equal on $[-\pi/\Delta, \pi/\Delta]$ to the product of the DFT of $g_\tau^{(h)}$ (which does not vanish since $[-\pi/\Delta, \pi/\Delta] \subset [-\pi/h, \pi/h]$) and the DFT of $(\delta_\tau + \widetilde{\Psi})\Sigma\star(\delta_\tau + \widetilde{\Psi})_\tau^\dagger$. Consequently, as long as h is small enough (so that Δ can be chosen greater than h and small enough), dividing on both hand sides Eq. (23) by $\widehat{g}_z^{(h)}$ is not, from a practical estimation point of view, a real problem. So, in the following we will write

$$(\mathbb{I} + \widehat{\Psi}_z^*)\Sigma(\mathbb{I} + \widehat{\Psi}_z^*)^\dagger = \widehat{v}_z^{(h)}/\widehat{g}_z^{(h)}, \quad (24)$$

without bothering with eventual cancelations of $\widehat{g}_z^{(h)}$.

Computing the square root of $(\mathbb{I} + \widehat{\Psi}_z^*)\Sigma(\mathbb{I} + \widehat{\Psi}_z^*)^\dagger$. In the estimation process, once we have estimated $v_\tau^{(h)}$ and consequently (through DFT) $\widehat{v}_z^{(h)}$ (for $z = i\omega$), using Eq (24), we can estimate $(\mathbb{I} + \widehat{\Psi}_z^*)\Sigma(\mathbb{I} + \widehat{\Psi}_z^*)^\dagger$. We need to go from there to the estimation of $\widehat{\Psi}_z$ (then using Eq. (22), we get $\widehat{\phi}_z$ and then by inverse Fourier transform ϕ_t). This problem requires therefore to take the square root of the left hand side of Eq. (24). In dimension $n = 1$, it means being able to go from $|1 + \widehat{\Psi}_z|^2$ to $\widehat{\Psi}_z$. There is clearly a phase determination problem. We will see that, in dimension $n = 1$, this phase is uniquely determined by the hypothesis **H1** and **H2**. However, in dimension $n > 1$, they are many possible solutions and determining the correct one is not necessarily possible in general. We need to make a strong additional hypothesis.

4.2 Further hypothesis on the kernel ϕ_t

In the following, we suppose that

H3 $\Sigma = \bar{\lambda}\mathbb{I}$ with $\forall\, i,\, E(\lambda_t^i) = \bar{\lambda} \in \mathbb{R}$, and $\widehat{\phi}_z$ can be diagonalized into a matrix D with some constant unitary matrix U (that does not depend on z):

$$\widehat{\phi}_z = U^\dagger D_z U. \tag{25}$$

Let us point out that, though always true in dimension $n = 1$, **H3** is a strong hypothesis for dimension $n \geq 2$. Clearly it is satisfied in the case all the components of the process are identically distributed, i.e. the process is invariant under arbitrary permutations. Other cases are very specific and consequently are not discussed in this paper.

4.3 The estimator

Using Eq. (21) along with **H3**, one gets that $\mathbb{I} + \widehat{\Psi}_z^*$ is diagonalizable in the same basis U and that

$$\mathbb{I} + \widehat{\Psi}_z^* = U^\dagger(\mathbb{I} - D_z^*)^{-1}U, \tag{26}$$

and thus from (24), one gets

$$E_z = U\widehat{v}_z^{(h)}U^\dagger / (\bar{\lambda}\widehat{g}_z^{(h)}), \tag{27}$$

where E_z is the diagonal matrix with the real positive coefficients :

$$E_z^{kk} = |1 - D_z^{kk}|^{-2}. \tag{28}$$

So the estimation problem reduces to being able to recover the coefficients D_z^{kk} from the coefficients $E_z^{kk} = |1 - D_z^{kk}|^{-2}$. This problem is solved by the following Lemma.

Lemma 1 *Let k be fixed. Let $z = i\omega$ (with $\omega \in \mathbb{R}$). Then*

$$(1 - D_{i\omega}^{(kk)})^{-1} = e^{\frac{1}{2}\log|E_{i\omega}^{kk}| - iH(\frac{1}{2}\log|E_{i\omega}^{kk}|)}, \tag{29}$$

where the operator $H(.)$ refers to the Hilbert transform [25].

The proof is based on the following theorem:

Theorem 2 (Paley-Wiener [27]) *Let's suppose that we observe the amplitude $|\widehat{f}_{i\omega}|$ of the Fourier transform of a real filter f_t. If*

$$\int_{\mathbb{R}} \frac{\log(|\widehat{f}_{i\omega}|)}{1 + \omega^2} d\omega \quad < \quad \infty, \tag{30}$$

then the filter g_t defined by its Fourier transform

$$\widehat{g}_{i\omega} = e^{\log(|\widehat{f}_{i\omega}|) - iH(\log(|\widehat{f}_{i\omega}|))}, \tag{31}$$

is the only causal filter (i.e., supported by $\mathbb{R}^+$), which is a phase minimal filter[1] and which satisfies $|\widehat{g}_{i\omega}| = |\widehat{f}_{i\omega}|$.

Proof of the Lemma. It is a simple application of this theorem with $|\widehat{f}_{i\omega}| = \sqrt{E_z^{kk}} = |1 - D_{i\omega}^{(kk)}|^{-1}$. Indeed, let us first show that $\widehat{g}_z = (1 - D_z^{(kk)})^{-1}$ is a minimal phase filter, i.e., that both the poles and zeros are such that $\Re(z) < 0$.

- Let z be a zero of $(1 - D_z^{(kk)})^{-1}$, then it is a pole of D_z^{kk} and consequently of $\widehat{\phi}_z$. However, from **H1** and **H2** one concludes that $|\widehat{\phi}_z| \leq \int_0^{+\infty} |e^{-zt}|\phi_t dt$ cannot be infinite, unless $\Re(z) < 0$.
- Let z be a pole of $(1 - D_z^{kk})^{-1}$. It thus satisfies $D_z^{kk} = 1$. Thus 1 is an eigenvalue of $\widehat{\phi}_0$ which is in contradiction with **H2**. Thus $(1 - D_z^{kk})^{-1}$ has no pole.

Consequently $\widehat{g}_z = (1 - D_z^{(kk)})^{-1}$ is a minimal phase filter. Moreover, since every coefficients of ϕ_t are positive and in L^1, the Fourier transforms $D_{i\omega}^{(kk)}$, for any k, are continuous functions of ω and goes to 0 at infinity. Along with the fact that $(1 - D_z^{kk})^{-1}$ has no pole and has its zeros on the half-plane $\Re(z) < 0$ we easily conclude that $|\widehat{f}_{i\omega}| = |1 - D_z^{kk}|^{-1}$ satisfies (30). The theorem above can be applied and the Lemma follows.

Main steps for kernel estimation. The different steps for the final kernel estimator, in the fully symmetric case, can be summarized as follows

- Set Δ small enough (see Section 4.1) and fix $h = \Delta$,
- Estimate the unconditional intensity $\bar{\lambda}$,
- Estimate the auto-covariance operator $v_t^{(h)}$ and compute its Fourier transform $\widehat{v}_{i\omega}^{(h)}$,
- Compute $(\mathbb{I} + \widehat{\Psi}_{i\omega}^*)(\mathbb{I} + \widehat{\Psi}_{i\omega}^*)^\dagger$ using Eq. (24). Diagonalize it and compute the matrix $E_{i\omega}$ defined by Eq. (27),
- Compute the diagonal matrix $D_{i\omega}$ using (29),
- Go back to the initial basis and Inverse Fourier transform to get the estimation of ϕ_t.

4.4 Some particular cases

4.4.1 The one dimensional case $n = 1$

As we already pointed out, the hypothesis **H3** is always true in this case since all the functions are scalar functions.

---

[1] a minimal phase filter [25] is a filter whose all the zeros and the poles of its Laplace transform satisfy $\Re(z) < 0$

Thus, the phase determination problem is solved without adding any assumption apart from **H1** and **H2**. The kernel estimator simply consists in first computing

$$|1 + \widehat{\Psi}_{i\omega}|^2 = \frac{\widehat{v}_{i\omega}^{(h)}}{\bar{\lambda}\widehat{g}_{i\omega}^{(h)}}, \tag{32}$$

and then inverse Fourier transform of

$$\widehat{\phi}_{i\omega} = 1 - e^{-\log|1+\widehat{\Psi}_{i\omega}| + iH(\log|1+\widehat{\Psi}_{i\omega}|)}. \tag{33}$$

4.4.2 The bisymmetric 2-dimensional case $n = 2$

In the two-dimensional case, the hypothesis **H3** is satisfied in the particular case

$$\mu^1 = \mu^2, \tag{34}$$

and the kernel ϕ_t is *bisymmetric*, i.e., has the form

$$\phi_t = \begin{pmatrix} \phi_t^{11} & \phi_t^{12} \\ \phi_t^{12} & \phi_t^{11} \end{pmatrix} \tag{35}$$

The matrix of the Laplace transform $\widehat{\phi}_z$ can be indeed decomposed as follows:

$$\widehat{\phi}_z = U^\dagger \begin{pmatrix} -\widehat{\phi}_z^{11} + \widehat{\phi}_z^{12} & 0 \\ 0 & \widehat{\phi}_z^{11} + \widehat{\phi}_z^{12} \end{pmatrix} U$$

where $U = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}$.

Diagonalizing and identifying the diagonal coefficients in both hand sides of (27) leads to

$$\frac{\widehat{v}_{i\omega}^{(h)11} + \widehat{v}_{i\omega}^{(h)12}}{\bar{\lambda}\widehat{g}_{i\omega}^{(h)}} = |1 + \widehat{\Psi}_{i\omega}^{11} + \widehat{\Psi}_{i\omega}^{12}|^2 \tag{36}$$

$$\frac{\widehat{v}_{i\omega}^{(h)11} - \widehat{v}_{i\omega}^{(h)12}}{\bar{\lambda}\widehat{g}_{i\omega}^{(h)}} = |1 + \widehat{\Psi}_{i\omega}^{11} - \widehat{\Psi}_{i\omega}^{12}|^2 \tag{37}$$

Applying the same method used in the case $n = 1$ to these last two equations, we get an estimate of $\widehat{\Psi}$. Finally we apply

$$\widehat{\phi}_z = (\mathbb{I} + \widehat{\Psi}_z)^{-1}\widehat{\Psi}_z \tag{38}$$

giving $\widehat{\phi}$. Applying the inverse transform to $\widehat{\phi}$ we finally get ϕ.

In the following section we illustrate these results and our methods on numerical simulations of 1D and 2D Hawkes processes.

5 Numerical illustrations

Let us discuss some examples illustrating the estimation method as defined previously using simulated Hawkes processes. All the simulations have been performed with the thinning algorithm described in [22].

We estimate $v_\tau^{(h)}$ from the realization of a Hawkes process (X_t), $t \geq 0$ (1D or 2D). We then strictly follow the method described in previous section in order to estimate the decay kernel ϕ. $v_\tau^{(h)}$ is sampled at rate Δ (i.e., $\tau = n\Delta$) up to a maximum lag τ_{max}. As explained in Section 4.1, in order to avoid problems related to the zeros of $\widehat{g}^{(h)}$, we naturally choose $h = \Delta$. From a practical point of view, Δ has to be chosen small enough in order to avoid Fourier aliasing.

5.1 The case $n = 1$

5.1.1 Exponential kernel

We first simulate a one dimensional Hawkes process with an exponential kernel:

$$\phi_t = \alpha e^{-\beta t}\mathbb{1}_{t \geq 0} \tag{39}$$

where we choose ($\mu = 1$, $\alpha = 1$, $\beta = 4$). This gives $\widehat{\phi}_0 = 1/4$, $\bar{\lambda} = 4/3$. The simulated sample contains 130000 jumps (it is approximately $T \simeq 10^5$ seconds long). In figure 1(a) we have reported the estimated kernel function ϕ_t (circles) on top of the true kernel (solid line). We see that the estimated kernel function is, up to some noise, very close to the real kernel

5.1.2 Power Law Decay

We now consider a power-law decaying kernel defined as:

$$\phi_t = \alpha(t + \gamma)^\beta \mathbb{1}_{t \geq 0} \tag{40}$$

with $\beta < -1$. In this case we have $\widehat{\phi}_0 = -\frac{\alpha}{\beta+1}\gamma^{\beta+1}$. We choose $\alpha = 32$ and $\beta = -5$ and $\gamma = 2$ making $\widehat{\phi}_0 = 0.5 < 1$ and $\bar{\lambda} = 2$ again with 130000 jumps ($T \simeq 65000$ secs). The kernel estimated on a single sample is reported in 1(b). Once again, one can see that, up to an additive noise, the estimated kernel fits well the real one. Let us point out that, since the decay is much slower than in the previous exponential case, we chose the maximum lag τ_{max} to be ten times as much.

5.1.3 Error Analysis

Let us briefly discuss some issues related to the errors associated with our kernel estimates. In ref. [3], a central limit theorem has been proved that shows that, asymptotically, the errors of the empirical covariance function estimates are normally distributed with a variance that decreases as T^{-1} (or N^{-1} for Δ fixed). One thus expects the same kind of results in the estimates of the components of ϕ.

Let us define the L^2 estimation error as:

$$e^2 = \sum_{k=1}^{\frac{\tau_{max}}{\Delta}} |\phi_{k\Delta} - \phi_{k\Delta}^{(e)}|^2 \tag{41}$$
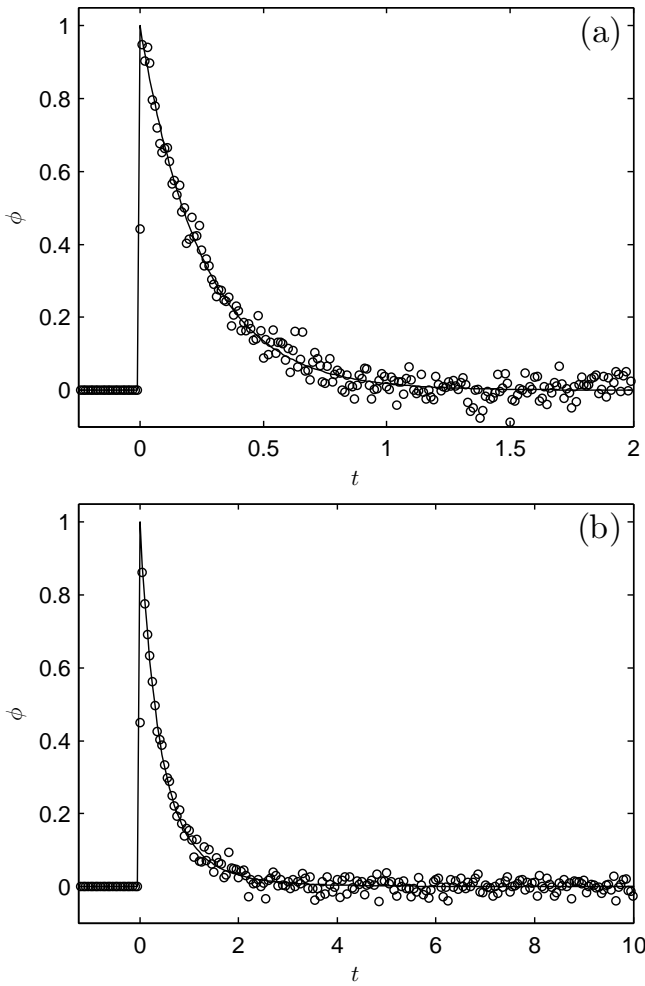
Fig. 1. Non parametric estimation of a one dimensional Hawkes process using method described in Section 4.4.1 from a unique realization with 130000 jumps. Estimated ($\circ$) and analytical kernel (solid line) are shown. (a) Case of the exponential decay kernel ϕ (Eq. (39)) with $\alpha = 1$, $\beta = 4$. We used $\Delta = 0.01$ and $\tau_{\max} = 2$. (b) Case of the power-law decay kernel ϕ (Eq. ((40))), with $\alpha = 32$, $\beta = -5$ and $\gamma = 2$. We used $\Delta = 0.05$ and $\tau_{\max} = 20$.
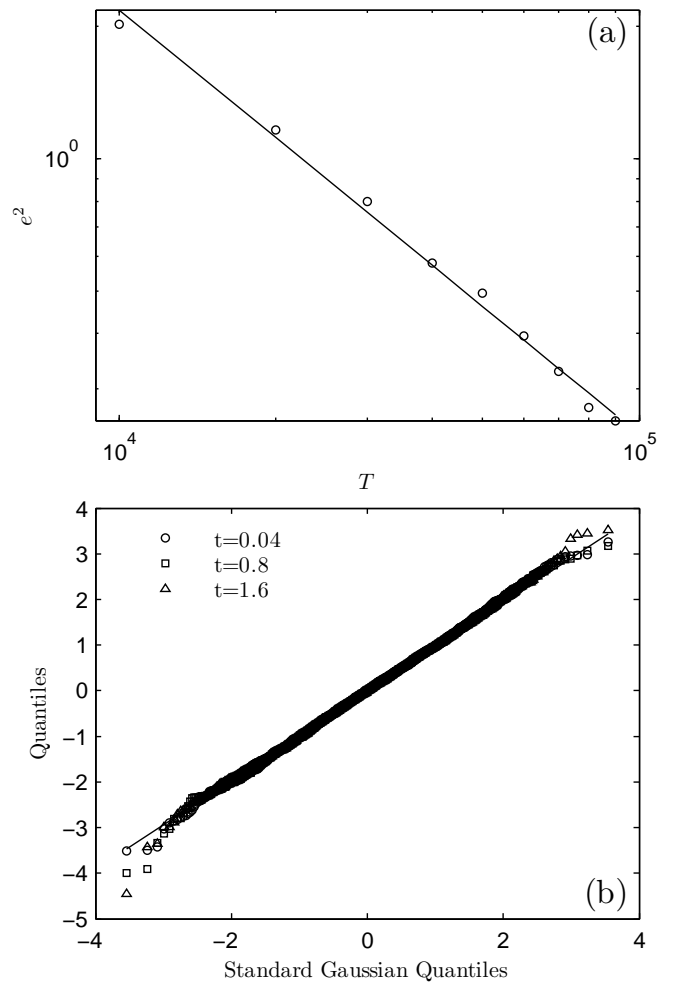
Fig. 2. Error analysis of a one dimensional Hawkes process with the same exponential kernel as in Fig. 1(a) with $\Delta = 0.01$ and $\tau_{max=2}$. (a) Mean square error as a function of the sample length T in log-log coordinates. The solid-line corresponds to the curve the expected T^{-1} behavior. (b) qq-plots of the Normal probability distribution function (pdf) versus empirical error pdf for ϕ_t for three different values of t. Here $T = 10^5$ seconds is fixed. The empirical error pdf's have been normalized to have the same variance (the variance increases as t goes to 0).

where $\phi^{(e)}$ is the estimated kernel. In Fig. 2(a), we have reported e^2 as a function of the sample length T for the same realization as the one used in Fig. 1(a). As expected, one observes a behavior very close to T^{-1}. Let us now look at the error between the analytical ϕ and the estimated one for a fixed $t > 0$. We find that, for each t, the series of errors are centered gaussian and mainly uncorrelated. Their variance increases as t decreases to 0. The normality of the observed errors is illustrated in Fig. 2(b) where we report, for 3 different values of t, the qq-plots of the empirical error pdf (with standardized variance) versus the normal pdf.

Let us point out that the estimation error depends, a priori, on the sampling rate Δ. Numerical simulations show that there is an optimal choice for the sampling rate parameter Δ. Indeed, we found that the estimation error

is minimum for a finite value $\Delta = \Delta^*$ which is neither large nor close to 0. For instance, for the process used in Fig. 1(a), we found $\Delta^* \simeq 0.15$. The fact that there exists such a minimum is not that surprising. On the one hand, if Δ is too large the error is dominated by Fourier aliasing. On the other hand, if Δ is too small, estimating $v_\tau^{(\Delta)}$ for a fixed $\tau > 0$ corresponds to estimating the correlation between the increments at scale Δ of two point processes. Such an estimation is well known to converge to 0 when the time-scale Δ goes to 0. In the Finance literature (see, e.g., [2]), this is known as the *Epps effect*. Of course, the optimal value Δ^* is not known a priori. However, it is natural to think that it is inversely proportional to $\bar{\lambda}$. On

practical situations we advocate to start choosing Δ of the order of $0.1\bar{\lambda}$ and then play around this value.

5.2 The bisymmetric 2-dimensional case $n = 2$ with an exponential kernel

Let us now consider a 2-dimensional bisymmetric Hawkes process introduced in Section 4.4.2. Both the diagonal term $\phi^{(d)}$ and the anti-diagonal term $\phi^{(a)}$ (cf Eq. (35)) have an exponential form

$$\phi_t^{(d)} = \phi_t^{11} = \phi_t^{22} = \alpha_d e^{-\beta_d t} \mathbb{1}_{t \geq 0} \qquad (42)$$
$$\phi_t^{(a)} = \phi_t^{12} = \phi_t^{21} = \alpha_a e^{-\beta_a t} \mathbb{1}_{t \geq 0}.$$

We use the following parameters for the simulations, $\alpha_d = 0.5$, $\beta_d = 8$, $\alpha_a = 1$, $\beta_a = 4$, and $\mu^1 = \mu^2 = 1$ (consequently $\bar{\lambda} = 1.45$). We simulate about 60000 jumps for each of the two components of the Hawkes process. The Figure 3 below shows the theoretical (solid lines) and estimated ($\circ$)versions of ϕ_d (in Fig. 3(a) and ϕ_a in Fig. 3(b)). We see that, as in the 1D case, we get a reliable estimate of both kernels.
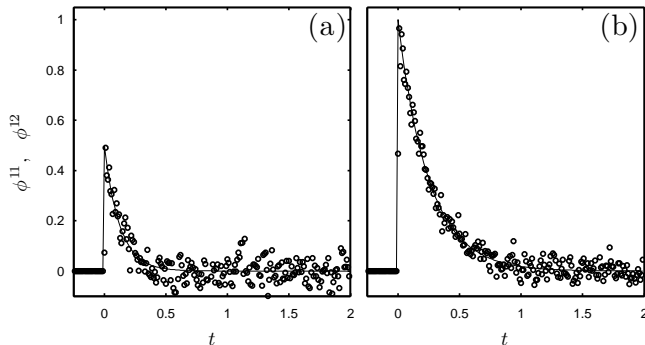

Fig. 3. Non parametric estimation of the two dimensional Hawkes bisymmetric exponential kernel ϕ (Eq. (42)) from a unique realization with 60000 jumps. We chose $\alpha_d = 0.5$, $\beta_d = 8$, $\alpha_a = 1$, $\beta_a = 4$, and $\mu^1 = \mu^2 = 1$ ($\bar{\lambda} = 1.4545$) We used $\Delta = 0.05$ and $\tau_{\max} = 2$. Estimated ($\circ$) and analytical kernel (solid line) are shown. (a) Estimation of the diagonal term $\phi_t^{(d)}$. (b) Estimation of the anti-diagonal term $\phi_t^{(a)}$

6 Application to high frequency market data

As mentioned in the introduction, Hawkes point processes have found many applications and notably as models for high frequency financial data. Indeed, because of the discrete and correlated nature of trade and limit order arrival times, point processes are natural models of market dynamics at the microstructure level. One can mention for instance Ref. [16] where buy and sell trades arrivals are represented by a bivariate Hawkes process with exponential kernels (see also [17]). Another approach, developed in

[2], consists in describing high frequency price dynamics as the difference between two coupled Hawkes processes representing respectively up and down discrete price variations. The authors emphasized that such model allows one to account for the main stylized facts characterizing the observed noise microstructure, namely the signature plot and the Epps effect (see end of Section 6.2.1 for a short remark on these effects).

We use **level 1** data (i.e, trades and best limit data) provided by QuantHouse Trading Solutions[2] of the 10-years Euro-Bund (Bund) and on the futures contracts on the Dax index. In this paper, we use 75 days covering the period between 2009-06-01 and 2009-09-15. The most liquid maturity is always chosen. In order to minimize seasonal effect, every day, only the data between 9 AM and 11 AM (GMT) are kept during which the rate of incoming orders can be considered as stationary, moreover, we ignored the days with too little trades. The data has millisecond accuracy and it has been treated in such a way that each market order is equivalent to exactly one trade [3].

6.1 Estimation in the case of the one-dimensional model for Bund data

In the following we test our method on the process of incoming trade times (market orders) for the Bund Future. This is a 1-dimensional point process, consequently we use the estimator described in Section 4.4.1. Every single day, we compute an estimation of the function $v_{n\Delta}^{(h)}$ with $h = \Delta = 0.1$ and $n\Delta < 100$. These quantities are then averaged on all the 75 days and we perform the ϕ estimation using this averaged quantity.

The results are shown in the figure 4. The so-obtained estimation of ϕ clearly displays a power law decay (Fig. 4(b)). A fit with the function αt^β gives $\alpha \simeq 0.1$, and $\beta \simeq -1$ [4].

Let us point out that we studied the stability of this estimation as the value for Δ is changed. Table 1, clearly shows that, as Δ decreases, there is a pretty large discrepancy on the estimated values of both both α and β. Though these values seem to stabilize when Δ reaches 0.1 which could indicate that this choice is not far from the optimal value Δ^* (see Section 5.1.3)

---

[2] http://www.quanthouse.com/

[3] When one market order hits several limit orders it results in several trades being reported.

[4] Strictly speaking a decay kernel of the form αt^{-1} is not admissible for our model. Indeed, its integral diverges both at $t = 0$ and at infinity. So it is admissible as long as we consider that this behavior has two cut-off, for small and large t. This hypothesis will be implicitly made in the following.
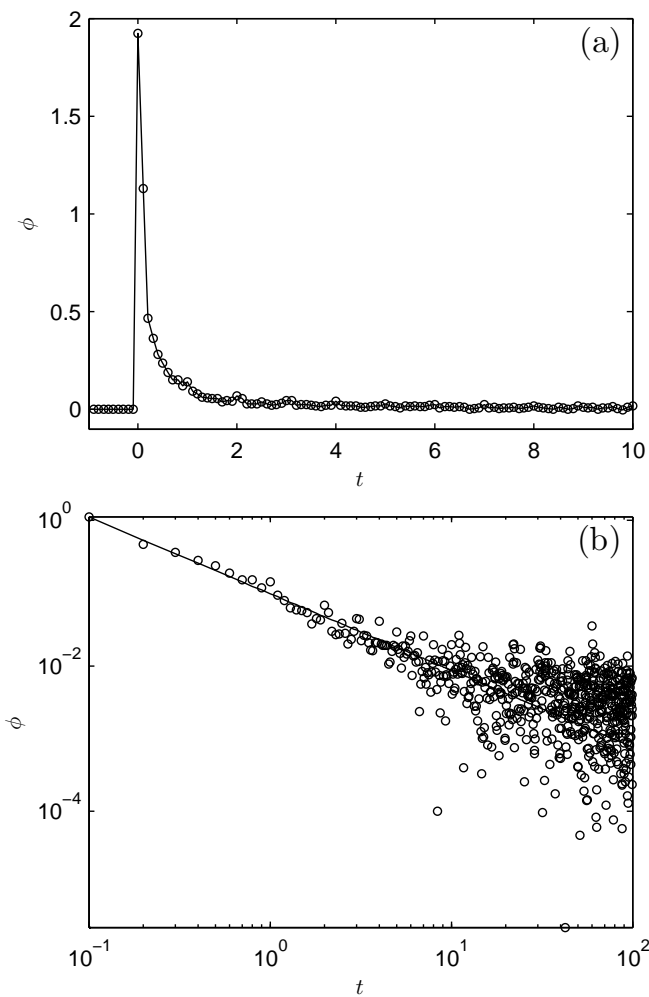
Δ	α	β
1	0.146276	-1.41515
0.5	0.117503	-1.30292
0.1	0.098624	-1.05329
0.05	0.092975	-1.03596
0.01	0.089227	-0.99899

Table 1. Results of the Power Law fit αt^{β}, applied to the estimated Hawkes kernel for the rate of incoming market orders of the Bund Futures. We show the results for different values of the parameter Δ. As Δ decreases, the estimations for both α and β stabilize and seem to indicate that the choice of $\Delta = 0.1$ is not far from the optimal value Δ^*.

N_t^-

$$X_t = N_t^+ - N_t^-.$$

We look at the two dimensional point process

$$N_t = \begin{pmatrix} N_t^+ \\ N_t^- \end{pmatrix}$$

As advocated in Ref. [2], 2 dimensional Hawkes processes are suited to model the so-defined N_t. A simple bisymmetric exponential kernel (with a null diagonal term) is able to reproduce remarkably the so-called *signature plot*. Moreover, in the case of 2 assets, each of them modeled by a 2-dimensional Hawkes and their interaction modeled by a simple symmetric cross term (leading to a 4-dimensional Hawkes process), these models have also been able to reproduce the so-called *Epps effect*.

The Signature plot and the Epps effect are two of the main stylized facts of financial time-series at a microstructure level. It is not the goal of this paper to go into more details about them, however, it is interesting to point out that the quantity $v_\tau^{(h)}$ on which our estimation procedure is based includes both the Signature plots (basically corresponding to the diagonal terms of the $h \to v_0^{(h)}$ matrix function) and the Epps effect (basically corresponding to the non diagonal terms of the same function).

Fig. 4. 1-dimensional non parametric estimation of ϕ for the point process of incoming market orders of the Bund futures. We used $\Delta = h = 0.1$ and $\tau_{\max} = 100$. (a) The estimation clearly displays a slow decay. (b) Log-Log plot the figure (a) above reveals that, in good approximation, the kernel can be considered as decaying as $t^{-\beta}$. A power-law least-square fit (solid line) provides the exponent $\beta = -1.05$.

6.2 Estimation in the case of the two-dimensional model

6.2.1 The two dimensional model

In this Section, we apply our estimation framework to the process that was initially introduced in [2] for modeling the changes in the mid-price of a given asset. Let X_t be the mid-price, we decompose it as the sum of the cumulative positive jumps N_t^+ and the cumulative negative jumps[5]

6.2.2 Validating hypothesis **H3**

In order to apply our estimation framework in dimension 2, we first need to check the hypothesis **H3**. This hypothesis is two-folded.

– First of all, on Fig. 5, we show for each single day an estimation of $\Lambda^1 = \Lambda^+$ versus $\Lambda^2 = \Lambda^-$. The plot shows that $\Lambda_1 \simeq \Lambda_2$ with a very small variation and therefore that we are within our assumption that $\Sigma = \bar{\lambda}\mathbb{I}$ with $\bar{\lambda} = \Lambda^1 = \Lambda^2$.
– Secondly, we need to show that the kernel matrix diagonalizes in a basis that is constant. Actually, Fig. 6 shows that $v_\tau^{(h)}$ (and consequently the kernel) is, in good approximation bisymmetric which implies that it diagonalizes on a constant basis. Indeed we see that $v_\tau^{(h),11}$ ($\circ$, in Fig. 6(a)) is very close to $v_\tau^{(h),22}$ ($\triangle$, in

---

[5] strictly speaking, the jumps have not always the same amplitude. Though for the Bund, most of them are 1-tick large it is not the case for the DAX data. If it is not the case N_t^+ (resp. N_t^-) just represent the point process (with jumps always equal to 1) with the same arrival time as the upward (resp. downward) jumps of X_t

Fig. 6(a)) for all $\tau > 0$. In the same way, we see that $v_\tau^{(h),12}$ ($\circ$, in Fig. 6(b)) is also (up to some estimation noise) equal to $v_\tau^{(h),21}$ ($\triangle$, in Fig. 6(b)) for all $\tau > 0$ (same results are obtained for $\tau < 0$).

Let us point out that the choice of the midpoint price series is motivated by two factors. First, if we choose the series of traded prices, we have, because of a non-zero spread value, an important "spurious" bouncing effect (oscillation between best bid and best ask prices) that is very hard to capture by our modeling approach (see [2]) and we get negative decay functions on the diagonal of ϕ: ϕ^{11} and ϕ^{22} (see Section 6.2.3 for a longer discussion on the influence of the bouncing effect on the kernel estimation). Second, if we choose the series of the last traded prices of buy orders only as in [2], the bouncing artifact disappears, however the symmetry of Λ^+ and Λ^- is naturally no longer verified. The series of midpoint prices has the advantage of having a reduced bouncing effect and of presenting identically distributed N^+ and N^- processes[6].

Fig. 6. Estimation of $v_\tau^{(h)}$ for $\tau > 0$ on 75 days (from 9am to 11am) in the framework of the 2-dimensional model for mid-price of the Bund futures. We see that the matrix $v_\tau^{(h)}$ is, in good approximation, bisymmetric. Same results would be obtained for $\tau < 0$. That completes (with Fig. 5) the validation of hypothesis **H3**. (a) $v_\tau^{(h),11}$ ($\circ$) and $v_\tau^{(h),22}$ ($\triangle$). (b) $v_\tau^{(h),12}$ ($\circ$) and $v_\tau^{(h),21}$ ($\triangle$).
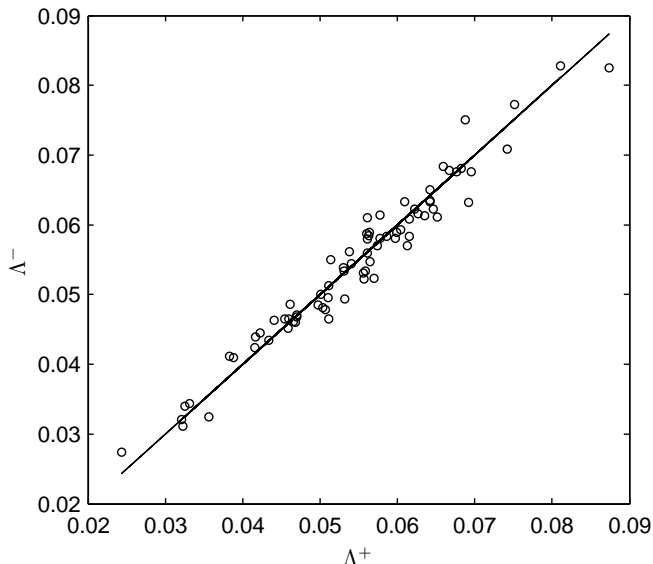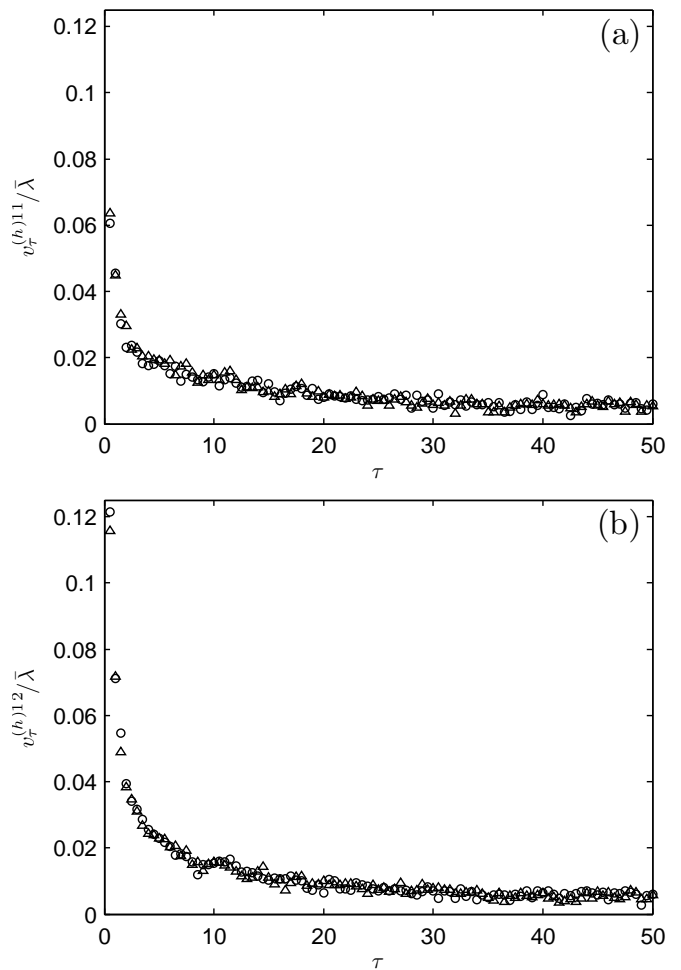

Fig. 5. Estimated $\Lambda^1 = \Lambda^+$ versus $\Lambda^2 = \Lambda^-$ for every single day (75 total) in the 2-dimensional model for mid-price of the Bund futures. Each dot represents a single day. The solid line corresponds to $\Lambda^+ = \Lambda^-$. We see that, in good approximation $\Lambda^+ \simeq \Lambda^-$ validating the first part of hypothesis **H3**.

6.2.3 Kernel estimation

We use the framework developed in Section 4.4.2 in order to estimate ϕ.

---

[6]   Still, when actually going through the estimation process, for stability reasons, we chose $\bar{\lambda}$ to be the average of the estimated Λ^+ and Λ^- and, similarly, $v_\tau^{(h),11}$ and $v_\tau^{(h),22}$ have been averaged as well as $v_\tau^{(h),12}$ and $v_\tau^{(h),21}$

The result of the non parametric estimation is shown in figure 7. We immediately notice that the diagonal term $\phi^{(d)} = \phi^{11}$ is an order of magnitude smaller than the non diagonal one $\phi^{(a)} = \phi^{12}$ and can be considered as being zero. This means that N^+ and N^- are not self exciting and exclusively mutually exciting. The log-log plot in Fig. 8 of the anti-diagonal term $\phi^{(a)} = \phi^{12}$ displays a power-law behavior αt^β with $\alpha \simeq 0.095$ and $\beta \simeq -0.99$ which is unsurprisingly close to the values we found for the 1-dimensional model in Section 6.1.

Figure 9 show the estimation of the bisymmetric kernel on Dax Futures time-series. Here, the diagonal kernel is of the same order (it can even be greater) than the anti-diagonal term. This property, that was not observed for the Bund, can be interpreted using tick size considerations. Indeed, the tick size of Dax Futures is well known to be very small in the sense that the agents trading on this
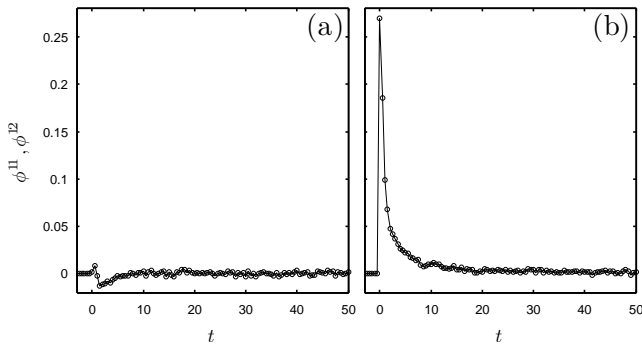
Fig. 7. Non parametric estimation of the bisymmetric kernel ϕ for the mid-price of the Bund futures using 75 days (between 9am to 11am). We used $\Delta = h = 0.5$ and $\tau_{\max} = 200$. The diagonal term $\phi^{(d)} = \phi^{11}$ appears to be negligible compared to the anti-diagonal term $\phi^{(a)} = \phi^{12}$ confirming that the N_+ and N_- are mutually but not self exciting (a) Estimation of the diagonal term $\hat{\phi}_t^{(d)}$. (b) Estimation of the anti-diagonal term $\hat{\phi}_t^{(a)}$.
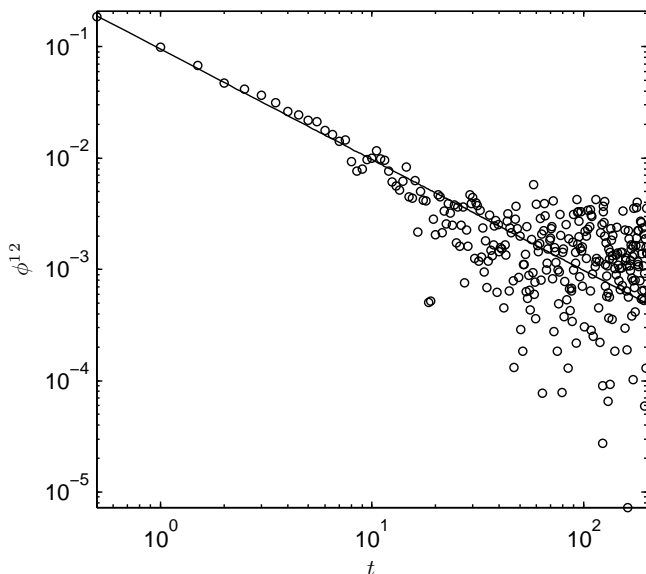

Fig. 8. Log-Log plot of the anti-diagonal kernel $\phi^{(a)} = \phi^{12}$ (estimated for the Bund futures) shown on Fig. 7(b). It shows that, in good approximation, this term can be considered as decaying as $t^{-\beta}$ A power law fit (solid line) gives the exponent $\beta \simeq -1$.

asset only care for moves that are greater than 1 tick (or equivalently for successive moves of 1 tick in the same direction). This translates into the fact that, at the scale of 1-tick, there is hardly no bouncing effect (negative autocorrelation of the returns). On the contrary, on the Bund, the tick size is well known to be "too big", i.e., when the price goes up by 1-tick, most of the time the next move is a downward move, i.e., the bouncing effect is very strong. In the framework of our model, it is clear that the strongest the bouncing effect the smaller the anti-diagonal term. Actually, looking carefully at the Fig. 7(a), the bouncing

effect on the Bund is so strong that, though much smaller than the anti-diagonal kernel, the estimated diagonal kernel seems to become significantly negative (compared to the estimation noise) which defies the assumption of our model. Indeed, if ϕ^{11} becomes negative, there is a priori, in theory, no guarantee that λ_t remains positive, and a negative rate of arrival is unacceptable. Of course, in practice, since these negative values are much smaller than the anti-diagonal kernel, the probabilities for λ_t to become negative are clearly negligible. The bouncing effect is most apparent on the series of trade prices (bouncing between best ask and best bid prices) and while it is heavily dampened when we use the series of midpoint prices, we see that it is still too strong to be fully captured by our model (see the discussions in [2] and in [9]).
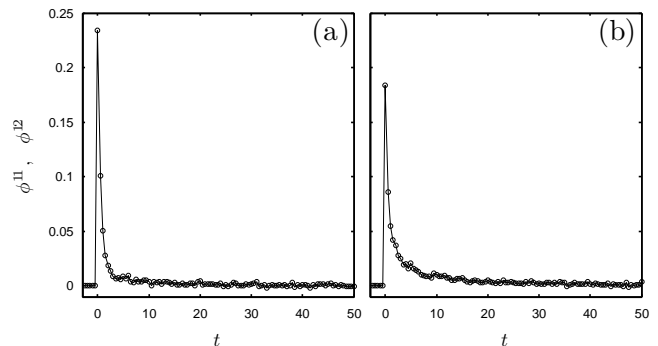

Fig. 9. Non parametric estimation of the bisymmetric kernel ϕ for the mid-price of the Dax futures using 75 days (between 9am to 11am). We used $\Delta = h = 0.5$ and $\tau_{\max} = 200$. Contrarily to the Bund kernel (Fig. 7), the diagonal term $\phi^{(d)} = \phi^{11}$ appears to be of the same order as the anti-diagonal term $\phi^{(a)} = \phi^{12}$. This is due to the tick size of the Dax which is much smaller than the tick size of the Bund. (a) Estimation of the diagonal term $\hat{\phi}_t^{(d)}$ $\beta \simeq -1.2$. (b) Estimation of the anti-diagonal term $\hat{\phi}_t^{(a)}$ $\beta \simeq -0.8$.

7 Conclusion and prospects

In this paper we have introduced a non parametric method to estimate the shape of the self-exciting kernels for symmetric Hawkes processes. Our method can be implemented very easily since it relies on the computation of the empirical covariance matrix and mainly uses Fourier transforms. As illustrated on specific numerical examples (1D and 2D), it provides reliable results for series of about 10^5 events long. This method can be very helpful to get a precise idea of the kernel functional shape before proceeding to a classical parametric (e.g. maximum likelihood) estimation. In a future work, we will consider a natural extension of the method in order to account for random marks associated with each events as e.g. in the ETAS model for earthquakes [23,15].

As far as financial time series are concerned, we have shown that, for both Bund and Dax Futures high frequency data, the Hawkes kernels are slowly (power law)

decaying. Even if this observation has to be confirmed by further studies, it is noteworthy that the values of the kernel power-law exponent we found ($\beta \simeq -1$) corresponds to the onset of stationarity. Our findings can be of great interest since they allow one to finely describe, at high frequency, the market activity as a long-memory self-exciting process. This can notably allow one to improve the results of Ref. [2] where the authors restrict themselves to exponential kernels to reproduce noise microstructure main features using Hawkes processes. On a more general ground, one can hope that our findings will be helpful to bridge the gap between a theory of price variations at the microstructure level and the standard coarse scale models with long range correlated volatility.

References

1. L. Adamopoulos. Cluster models for earthquakes: Regional comparisons. *Journal of the International Association for Mathematical Geology*, 8(4):463–475, August 1976.

2. E. Bacry, S. Delattre, M. Hoffmann, and J. F. Muzy. Modeling microstructure noise with mutually exciting point processes. *Accepted for publication in Quantitative Finance*, 2011.

3. E. Bacry, S. Delattre, M. Hoffmann, and J. F Muzy. Scaling limits for hawkes processes and financial data modelling. *preprint*, 2011.

4. M. S. Bartlett. The spectral analysis of point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 25(2):264–296, January 1963. ArticleType: research-article / Full publication date: 1963 / Copyright 1963 Royal Statistical Society.

5. M. S. Bartlett. The spectral analysis of Two-Dimensional point processes. *Biometrika*, 51(3/4):299–311, December 1964. ArticleType: research-article / Full publication date: Dec., 1964 / Copyright 1964 Biometrika Trust.

6. L. Bauwens and N. Hautsch. Modelling financial high frequency data using point processes. In T. Mikosch, J-P. Kreiss, R. A. Davis, and T. G. Andersen, editors, *Handbook of Financial Time Series*, pages 953–979. Springer Berlin Heidelberg, 2009.

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

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

9. K. A. Dayri. *Market Microstructure and Modeling of the Trading Flow*. PhD thesis, École Polytechnique, December 2011.

10. P. J. Brantingham E. Lewis, G. Mohler and A. Bertozzi. Self-exciting point process models of the insurgency in iraq. 2011.

11. P. Embrechts, J. T. Liniger, and L. Lu. Multivariate hawkes processes: an application to financial data. *Journal of Applied Probability*, 48:367–378, August 2011.

12. K. Giesecke and L. R. Goldberg. A top down approach to Multi-Name credit. *SSRN eLibrary*, 2007.

13. A. Hawkes. Point spectra of some mutually exciting point processes. *Biometrika*, 58:83–90, April 1971.

14. A. Hawkes. Spectra of some Self-Exciting and mutually exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 33-3:438–443, April 1971.

15. A. Helmstetter and D. Sornette. Subcritical and supercritical regimes in epidemic models of earthquake aftershocks. *J. Geophys. Res.*, 107:2237, 2002.

16. P. Hewlett. Clustering of order arrivals, price impact and trade path optimisation. In *Workshop on Financial Modeling with Jump processes*, 2006.

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

18. E. Lewis and G. Mohler. A nonparametric EM algorithm for a multiscale hawkes process. Joint Satistical Meetings 2011, 2011.

19. J. T. Liniger. *Multivariate Hawkes Processes*. PhD thesis, ETH Zurich, 2009.

20. D. Marsan and O. Lenglin. Extending earthquakes' reach through cascading. *Science*, 319(5866):1076 –1079, February 2008.

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

22. Y. Ogata. On lewis' simulation method for point processes. *Ieee Transactions On Information Theory*, 27:23–31, January 1981.

23. Y. Ogata. Seismicity analysis through point-process modeling: A review. *Pure and Applied Geophysics*, 155(2-4):471–507, August 1999.

24. Y. Ogata and H. Akaike. On linear intensity models for mixed doubly stochastic poisson and self- exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(1):102–107, January 1982. ArticleType: research-article / Full publication date: 1982 / Copyright 1982 Royal Statistical Society.

25. A. V. Oppenheim, R. W. Schafer, and J. R. Buck. *Discrete-time signal processing*. Prentice-Hall signal processing series. Prentice Hall, 1999.

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

27. R.E.A.C. Paley and N. Wiener. *Fourier transforms in the complex domain*. Colloquium Publications - American Mathematical Society. American Mathematical Society, 1934.

28. P. Reynaud-Bouret and S. Schbath. Adaptive estimation for hawkes processes; application to genome analysis. *The Annals of Statistics*, 38(5):2781–2822, October 2010. Zentralblatt MATH identifier: 1200.62135; Mathematical Reviews number (MathSciNet): MR2722456.

29. I. M. Toke. Econophysics of order-driven markets. chapter "Market making" behaviour in an order book model and its impact on the bid-ask spread. Springer Verlag, 2011.

30. D. Vere-Jones. Stochastic models for earthquake occurrence. *Journal of the Royal Statistical Society. Series B (Methodological)*, 32(1):1–62, January 1970. ArticleType: research-article / Full publication date: 1970 / Copyright 1970 Royal Statistical Society.

31. D. Vere-Jones and T. Ozaki. Some examples of statistical estimation applied to earthquake data. *Annals of the Institute of Statistical Mathematics*, 34(1):189–207, December 1982.

32. J. Zhuang, Y. Ogata, and D. Vere-Jones. Stochastic declustering of Space-Time earthquake occurrences. *Journal of the American Statistical Association*, 97(458):pp. 369–380, 2002.