Some applications of Hawkes point processes to high frequency finance

Tesi di Perfezionamento
in
Matematica per la Finanza

presentata da
Marcello Rambaldi

Supervisor

Prof. Fabrizio Lillo

Introduction

Advances in technology have deeply changed the way securities are traded. New technologies have enabled exchanges to automate most - if not all - the steps involved in trading. This has led on one side to dramatic cost reductions and on the other side has opened a full set of new possibility for market participants. In parallel to automation on the exchange side, also brokers, hedge funds, proprietary trading firms and other market participants have profited from the new technology for automating a variety of tasks, from optimization of order execution to whole trading strategies (Hendershott et al., 2011). So called algorithmic trading, that is the use of computer algorithms to automatically make trading decisions and handle the order submission, is nowadays deemed to account for the majority of trading in developed markets (Hendershott et al., 2011; Abergel et al., 2014; Chaboud et al., 2014).

With progressive market automation, also larger amount of data have started to be available. In particular, recent years have witnessed the increasing availability of high frequency financial data, i.e. data where each event is recorded separately and timestamped to a high precision, typically microsecond or even nanosecond in some cases[1]. This has opened new possibilities to exploit this information for research and trading purposes, and it was crucial for the development of algorithmic trading. New methods and techniques have been developed or borrowed from other disciplines, notably statistical physics, to handle and leverage these finer data. Several empirical studies (cfr. Cont (2001); Chakraborti et al. (2009)) contributed to the individuation of a set of statistical regularities and anomalies found in a large set of markets and assets and usually referred to as stylized-facts. More recently, advances in pattern recognition and statistical learning have been closely followed in particular by quantitative funds.

High frequency data display several peculiar properties, such as the irregular spacing in time, the discreteness of price changes, the bid-ask bounce as well as the presence of strong intraday seasonalities and persistent dynamics (Bauwens and Hautsch, 2009). Moreover, when order book data are available, one faces also the challenge of dealing with a multitude of event types, such as different kind of orders (limit orders, market orders, cancellations), at various price levels.

One of the key features of high-frequency financial data is the irregular time spacing between subsequent events. Since the seminal papers of Hasbrouck (1991) and Engle and Russell (1998), the informational content of events time spacing has been widely recognized. This has spurred the interest towards continuous time econometric models, i.e. point processes based models, as opposed to discrete models used in evenly spaced time series modeling (Bauwens and Hautsch, 2009). Among these models, Hawkes processes have recently risen in popularity (Bacry et al., 2015), and one of the reasons of their success is their straightforward application to cases where multiple event types are present. In this thesis we present some novel applications of Hawkes processes to the domain of high-

[1]Note that high precision does not always translate to the same amount of accuracy.

frequency financial data and propose an extension of the standard Hawkes framework that is designed to deal with local external perturbations that produce a sudden increase in the intensity of the process. All the methods and models that we present here are in principle amenable to applications in fields different from finance, whenever a point process description might be appropriate.

The rest of this thesis is organized as follows.

Chapter 1 Here we introduce in more detail the themes and research questions that we address in this thesis. We provide some context and we highlight the most relevant contributions.

Chapter 2 In this chapter we introduce more thoroughly Hawkes processes. We give the main definitions and introduce their most relevant properties. We also outline the two main estimation methods that we will use in application, namely Maximum Likelihood and the non parametric estimation developed by Bacry and Muzy (2016).

Chapter 3 This chapter contains the first original contribution of this thesis. We use the non parametric estimation method of Bacry and Muzy (2016) to study the high frequency dynamics on two very liquid future contracts traded on the Eurex exchange. We examine in particular the role played by the order size and we show that our approach can highlight several features of the high-frequency dynamics.

Chapter 4 Here we confront the problem of local non-stationarities due to external shocks, such as news announcements, that can greatly affect financial time series. We focus on large but localized variations in the intensity of order submission - that we name Intensity Bursts (IBs) - that can arise in correspondence of these events. We study this phenomenon on Foreign Exchange (FX) market data, which are especially affected by economic news announcements. We propose an extension of the Hawkes process that allows for the presence of externally driven IBs and we show that it is able to capture both the amplitude and time persistence of these IBs.

Chapter 5 In this final chapter we build on the work of Chapter 4 and we present a Hawkes process based procedure for IBs detection. While in the previous work the time and number of IBs were a priori fixed, here we devise an iterative procedure to identify them. One important feature of our procedure is that it tests the presence of an IB against a standard Hawkes process background. Since Hawkes processes naturally produce burst in the intensity, this allows to separate burst that cannot be explained solely by the Hawkes process from the others.

The last three chapters contain the original contributions of this thesis. Each of them is self-contained and in principle can be read separately.

Contents

List of publications

The following articles are part of this thesis:

- Rambaldi, M., P. Pennesi, and F. Lillo (2015). Modeling foreign exchange market activity around macroeconomic news: Hawkes-process approach. *Physical Review E 91*, 012819
- Rambaldi, M., E. Bacry, and F. Lillo (2016). The role of volume in order book dynamics: a multivariate Hawkes process analysis. *(submitted)*
- Rambaldi, M., V. Filimonov, and F. Lillo (2016). Detection of intensity bursts using Hawkes processes: an application to high frequency financial data. *(submitted)*

Chapter 1

Presented research

In this thesis we explore some applications of Hawkes point processes to the modeling of high frequency financial data. Hawkes processes were first introduced by Hawkes (1971) who was in search of a model for earthquake aftershocks occurrences (Linger, 2009). He proposed a multivariate point process model where the intensity function[1] has the form

$$\lambda_t^i = \mu^i + \sum_{j=1}^{D} \int_{-\infty}^{t} \phi^{ij}(t-s) \mathrm{d}N_s^j$$

where μ is a vector of constant background intensities (Poisson component) and $\phi(t)$ is a matrix of nonnegative functions called the *kernel* of the process, that is responsible for the characteristic self- and cross- exciting behavior of the process. Indeed the intensity of the Hawkes process depends on the past history of the process itself and the amplitude and persitence of the memory on past events is regulated by the kernels $\phi^{ij}(t)$. In Figure 1.1 a realization of a univariate Hawkes process with kernel function $\phi(t) = \alpha e^{-\beta t}$ is reported together with the corresponding intensity function to demonstrate the self-exciting behavior of the process. When an event occurs, the intensity has a positive jump and in turn this increases the likelihood of new events, creating a cascade effect. A more thorough introduction will be given in Chapter 2.

Hawkes processes have recently attracted a lot of attention in finance for several reasons. First, with the availability of data with increasingly fine time resolution, the time arrival of events has started to be recognized as an important source of information. This is in contrast with previous works where all analysis where conduced in event time as opposed to physical time. While this was a obliged choice when only evenly sampled data were available, and it often simplifies the analysis, now models that take into account the full time information are increasingly popular and Hawkes processes have been found to reproduce well the event time clustering typically found in financial order flow data. The second reason of Hawkes process success lies in the straightforward interpretation of model results. Indeed, as will be detailed in the next chapter, the kernels ϕ^{ij} directly specify how the occurrence of an event of type j influence the successive occurrence of a type i event. Moreover, the integral of the kernels is a measure of the average number of type j event triggered by a type i event. Furthermore, the relative importance of the feedback mechanism, as measured by the kernel integrals, and of the baseline intensities μ_i can be used to compare the different strengths of endogenous and exogenous factors

[1] The intensity function can be heuristically defined as $\lambda_t = \lim_{\Delta \to 0^+} \Delta^{-1} \mathbb{E}\left[N_{t+\Delta} - N_t | \mathcal{F}_t\right]$, where $N_t = \sum_{m \geq 1} \mathbb{1}_{t_m \leq t}$ denotes the counting process, $\mathbb{1}$ is the indicator function, and $\mathcal{F}_t$ represents the information available at time t.

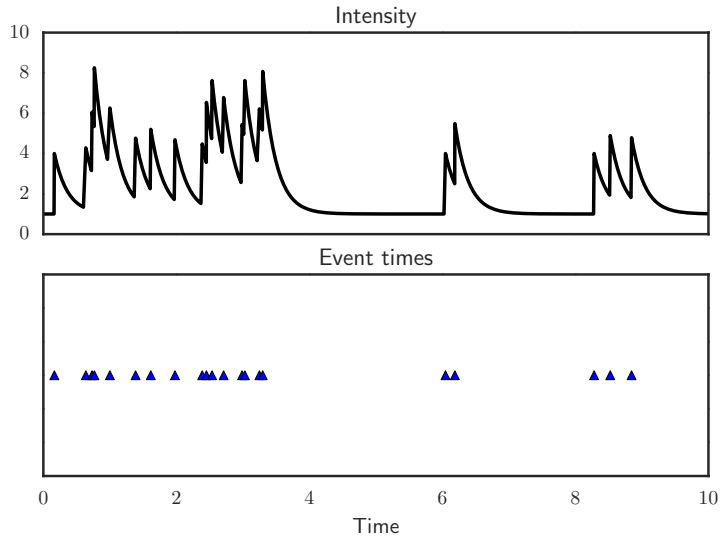

Figure 1.1: Example of a realization of a one dimensional Hawkes process. The blue triangles represent the arrival times of events, while the black line is the intensity λ_t of the process. Here we took the kernel $\phi(t)$ to be of the form $\phi(t) = \alpha e^{-\beta t}$, with $\alpha = 3$ and $\beta = 4$. The baseline intensity μ is equal to 1. This image illustrates the self-exciting behavior of the Hawkes process: every time one event arrive the intensity jumps thus increasing the likelihood of new events. This in turn generates events clustering as it is visible from the plot.

in the process dynamics. Finally, Hawkes processes lend themselves very naturally to multivariate applications. All these characteristics make Hawkes processes an appealing model when one's interest is in the study of mutual influence relationships between different types of events within a system. Additionally, the increasing availability of computational power along with the development of estimation techniques capable of dealing with large dimensions has made this kind of application even more appealing.

In this thesis we develop the Hawkes process framework along two main lines. The first leverages recent developments in non-parametric estimation of multivariate Hawkes processes to gain new insights on the high frequency order book dynamics and in particular on the role played by order size. The second line is instead concerned with the necessity to account for external localized perturbation of the dynamics such as those generated by news announcements. Before describing these points in more detail in the following, let us remark once again that although in this thesis the focus is on financial data, the methodologies proposed have in principle a much wider scope.

1.1 Analysis of order book dynamics

More than half of the world's financial market are nowadays operated fully electronically via a Limit Order Book (LOB) (Gould et al., 2013; Parlour and Seppi, 2008). Thus, understanding order book dynamics is of great importance for practitioners, researchers and regulators alike. Indeed, this understanding is needed to decide what actions to take in a given market situation (Harris and Hasbrouck, 1996), to design optimal order execution strategies (Obizhaeva and Wang, 2013), to better control market impact of one's trading activity (Eisler et al., 2012), as well as for the design of better electronic trading algorithms (Engle et al., 2006), and for monitoring market stability (Kirilenko et al., 2015).

Modeling LOBs is not an easy task. In fact, the interaction of a large number of hetero-

geneous agents gives rise to complicated global phenomena (Mitchell (2009); Gould et al. (2013)). Works investigating LOBs have taken a number of different approaches. One one side, perfect rationality models traditionally focus on the strategic behaviors of the single agents and from those derive the equilibrium relations (Kyle, 1985; Glosten and Milgrom, 1985; Parlour, 1998; Foucault, 1999; Almgren and Chriss, 2001; Roşu, 2009). On another extreme, zero-intelligence models (Bouchaud et al., 2002; Farmer et al., 2004, 2005; Cont et al., 2010; Toke, 2011) model the aggregate order flow with a given stochastic process, disregarding any strategic behavior. While both approaches are somewhat extreme and often too simplified, they can nevertheless provide insights on trading strategies and on the origin of statistical regularities observed in real data. Zero-intelligence models have also the advantage that the parameters of the stochastic processes governing the order flow can be estimated from real data and thus the statistical properties of these models can be compared with those of real LOBs. In contrast, perfect rationality models are often build on a series of assumptions necessary to determine the value of unobserved parameters, making it difficult to directly compare them in a statistically sound manner to real LOBs.

Most of the earlier zero-intelligence models assume that order flows are driven by Poisson processes. Although this assumption is in contrast with empirical observations, its simplicity makes the models tractable, thus allowing to derive testable formulas. Indeed, one of the empirical evidences on which most authors agree is that the order flow appears to depend both on its past history and on the present state of the book (Biais et al., 1995; Ellul et al., 2003; Hall and Hautsch, 2006; Hollifield et al., 2004; Lo and Sapp, 2010). Hence, a series of models that try to allow for dependence in the order flow were proposed. Zhao (2010) and Toke (2011) replaced Poisson processes with Hawkes processes to account for the empirically observed clustering of order arrivals. Abergel and Jedidi (2015) have studied the long time behavior of a Hawkes process driven LOB, building on their previous work (Abergel and Jedidi, 2013) where order flows were assumed to be driven by Poisson processes instead. They show that this Hawkes-driven order book is ergodic and leads to a diffusive behavior of the price at large time scales. Huang et al. (2015), while keeping the assumption of Poisson driven order flow, condition the Poisson rate on the state of the book and derive the limiting behavior of their model. Toke and Yoshida (2016) also recently proposed an order book simulator where order flow intensities depends on the available volume in the books and also on the spread.

All these works aim at the still open challenge of devising tractable yet sufficiently rich models that are able to closely reproduce the observed LOBs behavior. This is of great interest not only to get a deeper understanding of markets, but also for practical reasons such as building reliable market simulators as well as tools for transaction cost analysis and risk management. One further simplification that it is usually made in most zero intelligence models is to consider the order size as constant or as a *iid* random variable.

Our contribution. In the work presented in Chapter 3, we abandon this simplification and we show how multivariate Hawkes processes in combination with a recent non-parametric estimation procedure (outlined in Section 2.3.2) can be successfully used to investigate the role of the order size in the high frequency order book dynamics. Hawkes processes are naturally suited to describe mutual interaction between different event types (e.g. limit orders, market orders, cancellations). We further incorporate dependence on the order size by increasing the model dimension. Doing so has the advantage that it allows to analyze in a fairly general way the dependence on a mark, in this case the order size, with the already available tools developed for multivariate point process. Moreover, our approach can be easily generalized to variables other than the order size. This is of

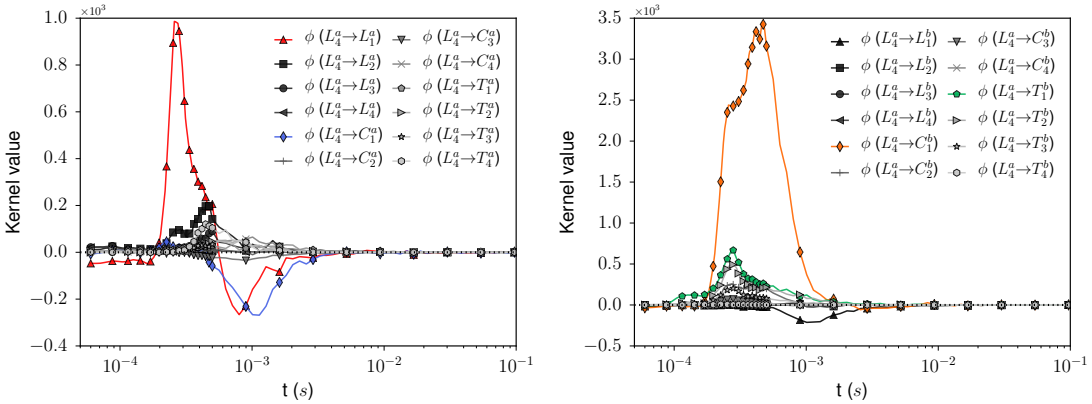

Figure 1.2: Example of the application of Hawkes processes and non-parametric estimation to high frequency data on the DAX future contract (from the work presented in Chapter 3). This figure shows estimated Hawkes kernels that describe the exciting effect of a large limit order being posted on the ask side on other order types. Order types are denoted L for limits, C for cancellations and T for marketable orders (trades). The side of the order is labeled a for ask and b for bid. Finally, order sizes are divided into four groups with 1 being the smallest size and 4 the largest one. The figure on the left shows the effect of a large limit order on the ask side, the one the right shows the effect on the bid side.

great importance as one of the challenges of LOBs modeling is precisely in understanding the role played by different variables in order to build parsimonious yet realistic models.

Our approach shows that mutual influences between orders are size-dependent, and that the order size is a relevant variable that needs to be taken into account for a faithful description of the order book dynamics. We are also able to identify several characteristics of the high-frequency dynamics, such as the typical reaction times as well as signatures of some common trading strategies like order splitting and high-frequency market making.

This is made possible by taking a non-parametric approach, that allows us to significantly limit assumptions on the shape of the kernels. In Figure 1.2 we anticipate some of the results of Chapter 3 and we plot the estimated Hawkes kernels that describe the effect of a large limit order arriving at the best ask. Each kernel $\phi(X \to Y)(t)$ determines the increase in the intensity of type Y events after t seconds from the occurrence of a type X event. By comparing the different kernels, we note a very strong and time-localized influence on smaller limit orders on same side (red triangles) and also a analogue effect on cancellation (orange diamonds) and trades (green pentagons) on the opposite side. It has been often argued that a large volume order carries more information. Following this reasoning, a large limit order on the ask may thus shift perceived fair price towards that side, thus triggering other limits on the same side and rush to catch the outstanding limits on the opposite side before they are canceled. We also remark that the estimated kernels are not positive on the entire domain, thus indicating that inhibition effects are also present.

In fact, we show that the use of the recent non-parametric estimation method developed by (Bacry and Muzy, 2016) is suitable also to analyze systems where moderate inhibition effects exist. This is possible despite the fact that standard Hawkes specification does not allow for inhibition. We furthermore provide a new result relating the presence of inhibition effects with negative kernels obtained with non-parametric estimation in Section 3.2.

1.2 Local non-stationarities

Several systems are characterized by the presence of discontinuities or "large" variations in some variables taking place in a short time interval, variously referred to as anomalies, outliers, jumps etc. depending on the application domain (Chandola et al., 2009). Financial markets are an emblematic example of such a system. Indeed, financial markets are subject to both internal feedback mechanisms and to external perturbations (political decisions, breaking news, regulation changes, evolving technology, etc.). Both these forces can determine sudden variation in some variables - most notably prices - as well as more persistent shift in their dynamics.

The detection and modelization of such "anomalies" is important for several reasons. Correct pricing of derivatives and other products requires that the underling stochastic model of price and volatility dynamics correctly deals with the presence of jumps. Dealing with surprise elements is also of fundamental importance for risk managers, as these events can greatly affect performances. Regulators are interested in detecting market abuse (e.g. insider trading cases) and monitor systemic risk. Finally, at a more fundamental level, jumps are related to the way financial markets process information. Discerning between exogenously triggered jumps and endogenous one is thus of both practical and theoretical importance as it related to the old yet always actual question on what causes prices to move (Shiller, 1981).

Pronounced variations in the rate of order arrivals (or price changes) are also often observed in financial markets. The rate of trading is indeed highly non-homogeneous. At microscopic scale, bursts of activity are intervaled by windows of lower activity, and this behavior is not well reproduced by Poisson models. At the daily level, it is well known that activity follows a roughly u-shaped pattern dictated by typical working hours and some scheduled event (such as the spike observed in European markets in correspondence of the opening of North-American ones).

Hawkes processes are increasingly used in finance to model the arrival of orders as they reproduce well the burstiness of intensity observed at short time scales thanks to their self-exciting specification. However, the standard and most widely used form of Hawkes processes considers a constant baseline intensity μ. This specification cannot therefore reproduce typical activity patterns such as the u-shaped daily pattern. This obstacle can be overcome either by introducing varying baseline intensities (typically piece wise constant or linearly varying) or by rescaling the original durations.

Other types of non-stationarities that have localized yet dramatic effect on the dynamics, such as external perturbation like news releases, have instead been largely ignored in these models as well as in other popular frameworks such as the Autoregressive Conditional Duration models.

Our contribution I. In the work presented in Chapter 4 we propose a Hawkes process featuring external shocks,

$$\lambda_t = \mu + \int_{-\infty}^{t} \phi(t-s)\mathrm{d}N_s + \int_{-\infty}^{t} \phi_N(t-s)\mathrm{d}N_s^{\mathrm{ext}}$$

where $\phi(t)$ is the endogenous self-exciting kernel and $\phi_N(t)$ is an exogenous kernel that accounts for the external perturbation. We show that this model can describe nicely the observed activity pattern in FX markets around economic announcements. Since these announcements are scheduled in advance and their release time is known to market participants, we also explore anticipation effects that could arise. Indeed, we find that the

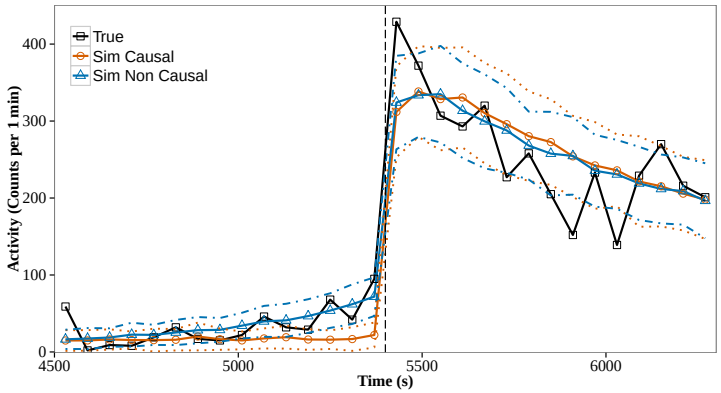

Figure 1.3: Actual and simulated activity from our calibrated model measured as events per 1 min for EUR/USD on Friday April 06, when the figure on US change in non-farm payrolls was released at 13:30. The dashed vertical line corresponds to the time of the announcement. The figure is a zoom on the 30 minute window around the news, while calibration was performed on a 3 hours window centered around the release time. Results from simulation of the models with and without the non causal term are compared. The dotted and dashed-dotted lines are one standard deviation confidence intervals obtained from simulations of the calibrated model of the causal-only and non-causal model, respectively. The anticipation effect is clearly visible on the actual data and is nicely captured by our model.

description of the activity pattern can be further improved by allowing also a non-causal, i.e. anticipating term.

In Figure 1.3 we show simulations from our calibrated model with and without the non-causal term in the case of the release of the latest figure on the variation in the number of paid workers in the US (Non-Farm Payrolls). This is a highly anticipated release that has often a significant impact on the market. We note that the model captures nicely the behavior of the mid-quote change intensity and that the addition of a non-causal term is a valuable addition in this case, reflecting the fact that operators position themselves before the release.

Our contribution II. In the work discussed in Chapter 5 we build on our previous work and we propose an iterative procedure that allows us to detect both the number and the locations of these sudden and dramatic increases in the intensity - that we named Intensity Bursts (IBs) - which are not compatible with the standard Hawkes dynamics. This overcomes two limitations that were imposed in the previous work, namely that the number and arrival time of the IB events were assumed to be known, i.e. the process N_t^{ext} was assumed to be known in advance. The new procedure thus yields a better fit to the data as compared to the previous model as the IB arrival time is now a model parameter that can be finely tuned. More importantly, it widens the scope of the model, as it can now be used as a burst detection tool. To demonstrate this, in Figure 1.4 we show the results of our procedure in a significant case. We applied our model to the intensity of mid-price changes in the EUR/USD cross during the testimony of the then Federal Reserve chairman Ben Bernake before the US House Budget Committee. Our procedure detects two IBs, one a few minutes before the scheduled start of the testimony and another during one of Mr Bernake's answers when he made a comment interpreted as dovish by the markets. Our procedure allows to precisely identify the occurrence time of the IB and also quantify its impact through model parameters.

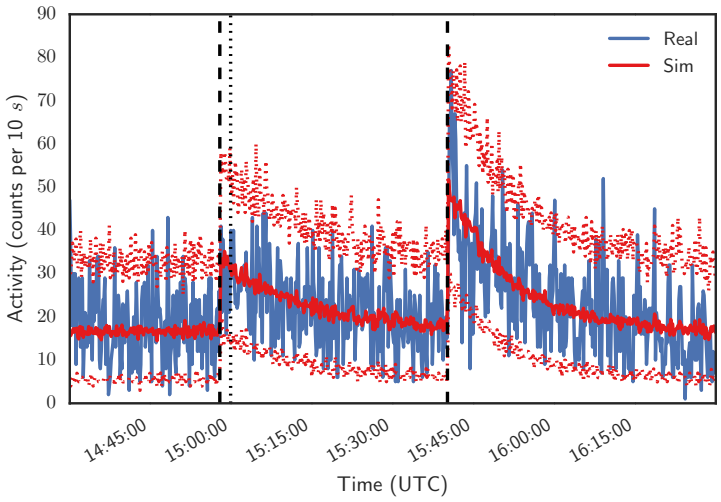

Figure 1.4: Market activity measured in mid price changes per unit time on the EUR/USD cross on 02 February 2012 during Federal Reserve then chairman Ben Bernake testimony before US House Budget Committee. The testimony was scheduled to start at 15:00 UTC (dotted line) and the blue line represents the actual marked activity. The red line is the average from 100 simulations from the model of Chapter 5, where both the time and number of intensity burst as well as their amplitude and persistence were estimated from the data. The red dotted lines represent the 95% confidence intervals obtained from the simulations. The dashed vertical lines indicate the time locations of the bursts detected by our model. We note that activity is found to increase significantly a first time a few minutes before the beginning of the testimony. Around the time marked by the second dashed line Bernake commented on the need to avoid too drastic spending cuts ("[...] *I think that as long as there's a credible, strong plan over time..and we move into that plan that we'll achieve most of the objectives of fiscal sustainability but we need to at least avoid doing harm.*[...]"[2].

Furthermore, the use of standard Hawkes process as base model consents to separate genuine external perturbations to the system from burst generated by internal feedback mechanisms which are already explained by Hawkes processes. We tested our procedure on FX market data and we found a large number of IBs. Some of them are specific to a single currency cross, while a significant part is found simultaneously on two or three crosses. Some of these bursts are found to be temporally related to price jumps and macroeconomic announcements. Interestingly, a significant fraction of IBs appears instead to be unrelated with price jumps and news, opening the question of the possible market events associated to them.

Finally, these works also contribute to the debate over the degree of endogenous vs exogenous influence on financial markets. In some of the recent literature (Filimonov and Sornette (2012); Hardiman et al. (2013)), the cluster structure of Hawkes processes has been used to assess the relative importance of endogenous and exogenous mechanisms. Our works show that properly accounting for localized external shocks is extremely important if one wishes to use Hawkes processes for such purpose. Indeed, we show that calibrating standard Hawkes models on data where localized external perturbations are present leads to severe overestimation of the branching ratio, which is the main quantity of interest in this kind of studies.

[2]See e.g. `http://www.reuters.com/article/us-usa-fed-bernanke-qa-idUSTRE8111K020120202` or `https://www.c-span.org/video/?304151-1/2012-economic-outlook` for the full testimony video.

Chapter 2

Hawkes processes

In this chapter we introduce the family of Hawkes point processes which constitutes the main tool used this thesis and we review its main properties. The main references for this chapter are Daley and Vere-Jones (2003, 2008); Cont and Tankov (2004); Bauwens and Hautsch (2009), and Bacry et al. (2015).

2.1 Point processes

A Point process on the real line is a stochastic process with some distinctive characteristics.

Definition 2.1.1. *Lets consider an interval I on the real line, without loss of generality we take $I = [0, T]$, $T \subset \mathbb{R}_0^+$. A simple, non-explosive point process on the interval $[0, T]$ is a sequence $\{t_m\}_{m\in\mathbb{N}}$ of random variables, $t_m \in T$, such that*

I. $\Pr(0 < t_1 < t_2 < ...) = 1$

II. $\Pr(t_m < t_{m+1}, t_m < \infty) = \Pr(t_m < \infty)$ *for* $m \geq 0$

III. $\Pr(\#\, t_m \in T < \infty) = 1$, $\forall\, T \subset \mathbb{R}$

A point process is thus an increasing sequence of real-valued random variables. Furthermore, we call it *simple* if the realization are distinct as long as they are finite. i.e, no simultaneous events occurs (property (II)). The last property assures that the number of events in every finite interval is always finite, i.e. the point process is non-explosive.

The waiting times

$$x_m = t_m - t_{m-1} \tag{2.1}$$

are called the *durations* of the point process $\{t_m\}_{m\in\mathbb{N}}$. It is clear that the starting point t_0 and the durations $\{x_m\}_{\{m=1,2,...\}}$ completely specify the path of the point process $\{t_m\}$. Their knowledge is thus equivalent to that of the times of the realizations of the process.

To every point process it is associated a counting process N_t defined by:

Definition 2.1.2. *Let $\{t_m\}_{i\in\mathbb{N}}$ be a simple point process, the associated counting process N_t is the right-continuous function given by $N_t = \sum_{m\geq 1} \mathbb{1}_{t_m \leq t}$. Where $\mathbb{1}$ is the indicator function: $\mathbb{1}_A = 1$ if $t \in A$ and $\mathbb{1}_A = 0$ if $t \notin A$. N_t is itself a stochastic cadlag[1] process with piecewise constant trajectories and unit increments.*

[1] A *cadlag* function (from the french "continue à droite, limite à gauche) is function defined on the real numbers (or a subset of them) that is everywhere right-continuous and for which the left limit $f(t-)$ exists $\forall t$.

N_t is the total number of events occurred up to time t. Since there is a one-to-one correspondence between a point process and the associated counting process, a point process can be uniquely represented by a counting process. In fact, the notation N_t is often used to indicate a point process.

From Daley and Vere-Jones (2003) a point process is stationary according to the following definition:

Definition 2.1.3. *A point process is stationary when for every $r = 1, 2, ...$ and all bounded Borel subsets $A_i, ..., A_r$ of the real line, the joint distribution of*

$$\{N(A_1 + t), ..., N(A_r + t)\}$$

does not depend on t $(-\infty < t < \infty)$.

2.1.1 Marked and multivariate point processes

Often, it is desirable to associate another random variable to each realization of a point process. For example, if we are observing the times at which transactions on a certain asset take place, we can associate to each of them the price paid by the buyer. In this case the process is said a *marked* point process.

Definition 2.1.4. *A marked point process in the interval $T \subset \mathbb{R}$ and with marks in the space $\mathcal{K}$ is a point process $\{(t_i, k_i)\}$ on $T \times \mathcal{K}$ with the additional property that the ground process, i.e. the process without marks, is itself a point process on T.*

We observe that the marks need not to be scalars, they can as well be vector-valued. For instance, with reference to the previous example, we can observe the price paid P together with the amount traded V, the realization of the point process have thus the form (t_i, P_i, V_i).

In many cases it is also interesting to study the interdependence of different point processes, this lead to the concept of *multivariate* point process, which is a special case of a marked point process:

Definition 2.1.5. *A multivariate point process is a marked point process $\{(t_i, W_i)\}$ with marks in a finite set $\mathcal{M} = \{1, ..., M\}$.*

In this case the marks are labels of the type of event, for example if we are observing the transactions on two different stocks, say A and B, we have $\mathcal{M} = \{\mathcal{A}, \mathcal{B}\}$. In the multivariate case we can identify the individual processes as $N_t^i = \sum_{i \geq 1} \mathbb{1}_{\{t_i \leq t\}} \mathbb{1}_{\{W_i = m\}}$ and the total process as $N_t = \sum_{i=1}^{M} N_t^i$. Although a multivariate point process is a special case of a marked point process, there is a clear conceptual difference between the finite-valued marks case and the infinite-valued case that justifies the distinction. In the first case we are interested in the relations between different point processes, in the second case we are dealing with a single component to which we associate a (possibly vector-valued) random variable. Indeed, we can have multivariate point process with associated marks. Continuing with the two stocks example, we can associate to each realization t_i^A or t_j^B the corresponding price paid P_i^A and P_j^B and get a multivariate marked point process.

2.1.2 Filtrations and History

If we interpret the index t of a stochastic process as a time variable, we introduce a dynamic aspect which require us to properly define the notions of information and predictability

(Cont and Tankov, 2004). As time passes, more information (e.g the realizations of the process) is revealed. Quantities that are random at time $t = 0$ may become predictable as more information is released. To take into account this dynamic aspect, the concept of *filtration* is introduced:

Definition 2.1.6. *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a filtration or information flow on $(\Omega, \mathcal{F}, \mathbb{P})$ is an increasing family of σ-algebras*[2]

$$(\mathcal{F}_t) \; : \; \forall\, t \geq s \geq 0, \; \mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F} \tag{2.2}$$

$\mathcal{F}_t$ is interpreted as the information known at time t, which increase with time. In the case when the past realizations of the process are the only information available, the filtration is called the *natural filtration* or *history* and denoted $\mathcal{H}_t$. Given $\mathcal{F}_t$ it is possible to distinguish quantities that are known from those that are still random at t.

A stochastic process X_t is said to be $\mathcal{F}_t$-*adapted* or *nonanticipating* if its random value is revealed by $\mathcal{F}_t$ (i.e. it is $\mathcal{F}_t$-measurable). A $\mathcal{F}_t$-*predictable* process is a process for which its value at t is known given $\mathcal{F}_{t^-}$, that is, it is $\mathcal{F}_{t^-}$-measurable. A predictable process is also adapted, but the contrary is not true. In particular an adapted left-continuous process is predictable.

2.1.3 Compensator and Conditional Intensity

In martingale-based point process theory, an important concept is that of *compensator*. This, in turn, lead to the introduction of the conditional intensity function, which is of crucial importance in the definition of many point processes including Hawkes processes. In the exposition, we will follow mainly Bauwens and Hautsch (2009). Before introducing these notions we first recall the definition of *martingale*:

Definition 2.1.7. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a filtered probability space with filtration $\mathcal{F}_t$. A stochastic right-continuous process X_t on $[0, T]$ is an $\mathcal{F}_t$-martingale if*

I. X_t is nonanticipating with respect to $\mathcal{F}_t$

II. $\forall\, t \in [0, T]$, $\mathbb{E}[X_t] < \infty$

III. $\forall\, s > t$, $\mathbb{E}[X_s | \mathcal{F}_t] = X_t$

If property (III) is substituted by

IV. $\forall\, s > t$, $\mathbb{E}[X_s | \mathcal{F}_t] \geq X_t$

then the process is said a submartingale.

From definition 2.1.7, it is clear that a non-explosive counting process N_t is a submartingale. As such, it can be decomposed in a zero-mean martingale M_t and a unique $\mathcal{F}_t$-predictable process, L_t, called the compensator of N_t, namely

$$N_t = M_t + L_t \tag{2.3}$$

[2] A σ-algebra $\mathcal{F}$ on a space Ω is a family of subset of Ω such that:

- $\Omega \in \mathcal{F}$
- If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$
- If $\{A_i\}_{i \in \mathbb{N}} \in \mathcal{F}$ is a numerable sequence of disjoint subsets of Ω, then $A = \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

This decomposition is known as the Doob-Meyer decomposition. Now, since

$$\mathbb{E}\left[M_t|\mathcal{F}_t\right] = 0, \tag{2.4}$$

it follows that

$$\mathbb{E}\left[N_t|\mathcal{F}_t\right] = \mathbb{E}\left[L_t|\mathcal{F}_t\right]. \tag{2.5}$$

The compensator is then identifiable as the local conditional mean of N_t. Let λ_t be a scalar, positive, $\mathcal{F}_t$-predictable process. Then, λ_t is called the $\mathcal{F}_t$-*conditional intensity* of N_t if

$$L_t = \int_0^t \lambda_u \mathrm{d}u. \tag{2.6}$$

From (2.5), λ_t can also be defined by the relation

$$\mathbb{E}\left[N_s - N_t|\mathcal{F}_t\right] = \mathbb{E}\left[\int_t^s \lambda_u \mathrm{d}u \mid \mathcal{F}_t\right] \tag{2.7}$$

that must hold for all t, s such that $0 \leq t \leq s$. It is possible to obtain a more intuitive interpretation of λ_t letting $s \to t$ in the previous equation

$$\lambda_{t^+} = \lim_{s \to t^+} \lambda_s = \lim_{s \to t^+} \frac{1}{s-t}\mathbb{E}\left[N_s - N_t \mid \mathcal{F}_t\right] \tag{2.8}$$

Then, $\lambda(t|\mathcal{F}_t)$ is interpreted as the instantaneous expected arrival rate conditional on the available information. In case the point process N_t is stationary, then $\Lambda := \mathbb{E}\left[\mathrm{d}N_t\right]/\mathrm{d}t = \mathbb{E}\left[\lambda_t\right]$ is constant.

For a multivariate point process we have a conditional intensity function λ_t^i for each subprocess, and the total intensity of the ground process is $\lambda_t = \sum_{i=1}^M \lambda_t^i$.

2.1.4 Representations of a Point Process

We conclude this section by examining the problem of how to specify the properties of a point process, a question that is of great importance to build point process models of real phenomena. We have seen that to a point process we can uniquely associate a counting process, the process of its durations and a conditional intensity function. Thus, to specify the characteristics of a point process we can set the properties of one of the aforementioned associated processes, this lead to the *counting representation*, the *duration representation* and the *intensity representation* respectively.

The well known Poisson process can be defined following each of the possible representations. In fact, for a Poisson process we have:

Counting representation The Poisson process is the one for which the counting process has independent increments and the number of events in each time interval τ follows a Poisson distribution

$$\Pr\left(N_{t+\tau} - N_t = k\right) = e^{-\lambda\tau}\frac{(\lambda\tau)^k}{k!}.$$

Duration representation Let $\{x_i\}_{i\geq 1}$ be independent exponentially distributed random variables with parameter λ, the Poisson process $\{t_m\}$ is then defined by $t_m = \sum_{k=1}^m x_k$. That is, a process with independent exponential durations is a Poisson process.

Intensity representation The Poisson process is defined as the point process with constant conditional intensity $\lambda_t = \lambda$.

The discrete nature of duration models cause them to be unsuitable when the information set has to be updated within a duration spell as is the case for multivariate processes, for example because of an event in other related point process has happened. On the opposite side, intensity based models lend themselves perfectly to the multivariate case, as the total intensity is simply the sum of each component intensity and it is easy to make each component depend on the realizations of other-type events.

2.2 The Hawkes process

Hawkes processes belong to the class of self-exciting processes, characterized by the fact that the intensity is determined by a a weighted function of the time distance to previous points of the process (Bauwens and Hautsch, 2009). They were first introduced by Hawkes (1971) in the field of seismology. Since then, Hawkes processes have found application in very diverse fields such as seismology (Ogata, 1988), genomics (Reynaud-Bouret et al., 2010), neurophysiology (Pierre Bremaud (1996); Chornoboy et al. (1988)) as well as in works on the spread of crime and violence (Lewis et al., 2012; Mohler et al., 2011) and on social network dynamics (Crane and Sornette, 2008; Blundell et al., 2012; Zhou et al., 2013).

2.2.1 Definition

Definition 2.2.1. *A D-variate Hawkes process is a point process whose conditional intensity vector* $\lambda_t = \{\lambda_t^i\}_{i=1,\dots,D}$ *has the form*

$$\lambda_t^i = \mu^i + \sum_{j=1}^{D} \int \phi(t-s)\mathrm{d}N_s \tag{2.9}$$

where $\mu = \{\mu^i\}_{i=1,\dots,D}$ *is a vector of baseline intensities and* $\Phi(t) = \{\phi^{ij}(t)\}_{i,j=1,\dots,D}$ *is a matrix valued function called the* kernel *of the process. Furthermore, the components* $\phi^{ij}(t)$ *are such that*

- $\phi^{ij}(t) \geq 0 \; \forall \; 1 \leq i,j \leq D$;
- $\phi^{ij}(t) = 0 \; \forall \; t < 0$;
- $\int |\phi^{ij}(t)|\mathrm{d}t < \infty$

The linear structure of the intensity (2.9) allows one to map the Hawkes process exactly onto a cluster process (Hawkes and Oakes, 1974), where the process consists of a superposition of random clusters, each of which starts with a single *immigrant*, generated from a homogeneous Poisson process with intensity μ. Specifically the following is an equivalent definition of a Hawkes process.

Definition 2.2.2. *Let D denote a positive integer and let* $[0,T]$ *with* $0 < T \leq \infty$ *be a time interval. Then, the sequence of events* $\{(t_m, k_m)\}$ *generated according to the following procedure is equivalent to the process* (2.9).

1. *For each component* $1 \leq i \leq D$, *the points that form the cluster centers (*immigrants*) are generated by Poisson processes with rate* μ^i.

2. *For each component* $1 \leq j \leq D$*, each immigrant event* $(t_m^{(0)}, j)$ *generates a cluster* C_m *which consists of events of generation* $g = 0, 1, \ldots$ *that obey the following branching structure:*

   (a) $(t_m^{(0)}, j)$ *belongs to generation 0.*

   (b) *Recursively, for each component* j *the event* $(t_k^{(g)}, j)$ *of generation* g *gives rise to* D *independent non-homogeneous Poisson process with intensity function* $\phi^{ij}(t)$ *each with offspring in generation* $g + 1$ *of component* i*.*

   (c) *The cluster* $C_{m,j}$ *is the union of all the offspring of* $(t_m^{(0)}, j)$*.*

3. *The union of all clusters forms the Hawkes point process* N*.*

Although the Hawkes process is well defined for each $\Phi(t)$ that respects the conditions in Definition 2.2.1, some constraint has to be imposed on Φ for the process to be stable. The following Proposition (Hawkes, 1971) gives the condition for stationarity.

Proposition 2.2.1. *(Stationarity) If the kernel matrix* $\Phi(t)$ *is such that the spectral radius of the matrix* $\|\Phi\| = \{\|\phi^{ij}\|$ *is smaller than one, then the associated counting process has stationary increments and the intensity* λ_t *is asymptotically stationary.*

2.2.2 Properties

Thanks to the linear nature of (2.9), many properties can be derived analytically. All the properties below are derived under the assumption that the stationarity condition of Proposition 2.2.1 is respected.

First and second order properties are of particular relevance for the Hawkes process as it turns out that the knowledge of these is sufficient for the compete characterization of the process (Bacry and Muzy, 2016).

These properties are conveniently expressed in terms of the matrix $\Psi(t)$ defied as

Definition 2.2.3. *For a Hawkes process with stationary increments, the function* $\Psi(t)$ *is defined as the causal solution of*

$$\Phi(t) + \Psi(t) * \Phi(t) = \Psi(t) \tag{2.10}$$

which is given (Bacry and Muzy, 2016) by

$$\Psi(t) = \sum_{n=0}^{\infty} \Phi(t)^{(*n)} \tag{2.11}$$

where $\Phi(t)^{(*n)}$ *denotes convolution with itself* n *times. In Laplace domain* (2.11) *reads*

$$\hat{\Psi}(z) = (\mathbb{I} - \hat{\Psi}(z))^{-1} - \mathbb{I} \tag{2.12}$$

where the two sided definition of the Laplace transform

$$\hat{f}(z) = \int_{-\infty}^{+\infty} f(t) e^{zt} \mathrm{d}t \tag{2.13}$$

was used.

We can now state first and second order properties, we refer to Bacry and Muzy (2016) for the proofs.

Proposition 2.2.2. *Given a stationary Hawkes process with intensity* λ_t

- *First order: the average intensity* $\Lambda = \mathbb{E}\left[\mathrm{d}N_t\right]/\mathrm{d}t$ *is given by*

$$\Lambda = (\mathbb{I} - \|\Phi\|)^{-1}\mu = (\mathbb{I} + \hat{\Psi}(0))\mu \tag{2.14}$$

- *Second order: the infinitesimal covariance matrix*

$$\nu^{ij}(t - t') = \frac{\mathbb{E}\left[\mathrm{d}N_t^i \mathrm{d}N_{t'}^j\right] - \mathbb{E}\left[\mathrm{d}N_t^i\right]\mathbb{E}\left[\mathrm{d}N_{t'}^j\right]}{\mathrm{d}t\mathrm{d}t'} \tag{2.15}$$

is given in Laplace domain by

$$\hat{\nu}^{ij}(z) = (\mathbb{I} - \hat{\Psi}(-z))\Sigma(\mathbb{I} - \hat{\Psi}^T(z)) \tag{2.16}$$

where Σ *is a diagonal matrix with entries* $\Sigma^{ii} = \Lambda^i$.

As mentioned before, the knowledge of first and second order properties is enough to fully characterize a Hawkes process. Before presenting this result, which was proved in Bacry and Muzy (2016), we introduce an equivalent specification of the second order properties.

Definition 2.2.4. *Let* $\nu^{ij}(t)$ *represent the covariance density of a Hawkes process. The conditional law matrix* $g(t)$ *is defined as*

$$g^{ij}(t) = \frac{\mathbb{E}\left[\mathrm{d}N_t^i | \mathrm{d}N_0^j = 1\right]}{\mathrm{d}t} - \Lambda^i, \ \forall\, t > 0 \tag{2.17}$$

and it is related to the covariance density through the relation

$$\nu(t) = \Sigma g^T(t). \tag{2.18}$$

We can now state the result of Bacry and Muzy (2016).

Theorem 1. *Given a Hawkes process defined by* (2.9) *which satisfies the stationarity condition, the matrix function* $\chi(t) = \Phi(t)$ *is the unique solution of the Wiener-Hopf equation*

$$g(t) = \chi(t) + \chi(t) * g(t) \ \ \forall\, t > 0 \tag{2.19}$$

such that the components of $\chi(t)$ *are causal and* $\chi^{ij} \in L^i$ *for all* i, j.

Thus, for given Λ and $g(t)$, there exists at most one Hawkes process compatible with them. Note however that such a process is not always guaranteed to exist. This result also represent the starting point for nonparameric estimation of the interaction kernel as detailed in Section 2.3.2 below.

Although in this thesis we will not make use of them, it is worth noting that expression for higher order properties also exist. Indeed, Jovanović et al. (2015) have derived a combinatorial procedure for obtaining cumulants of arbitrary order of a Hawkes process.

2.2.3 Interpretation

One of the key strengths of Hawkes processes is their convenient interpretability. Following the culstering representation of Definition 2.2.2, one can express the total number of events of type i, N_t^i as

$$N_t^i = N_t^{i\leftarrow 0} + \sum_{j=1}^{D} N_t^{i\leftarrow j} = N_t^{i\leftarrow 0} + \sum_{j=1}^{D} N_t^{i\leftarrow j*} \tag{2.20}$$

where

- $N_t^{i\leftarrow 0}$ are immigrants events (exogenous)
- $N_t^{i\leftarrow j}$ are events of type i directly triggered by a type j event
- $N_t^{i\leftarrow j*}$ are events of type i directly or indirectly triggered by a type j event. I.e. the oldest ancestor is a type j event.

These quantities are related to the process kernel and baseline intensity by the following proposition .

Proposition 2.2.3. *Given a stationary D-variate Hawkes process with counting process N_t , the following relations hold*

$$\mathbb{E}\left[\mathrm{d}N_t^{i\leftarrow 0}\right] = \mu^i \mathrm{d}t \tag{2.21}$$

$$\mathbb{E}\left[\mathrm{d}N_t^{i\leftarrow j}\right] = \hat{\phi}^{ij}(0)\Lambda^j \mathrm{d}t \tag{2.22}$$

$$\mathbb{E}\left[\mathrm{d}N_t^{i\leftarrow j*}\right] = \hat{\psi}^{ij}(0)\mu^j \mathrm{d}t \tag{2.23}$$

Hence, these quantities can be used in application to estimate the average fraction of events of a certain type triggered directly or indirectly by events of another type.

2.3 Estimation methods

The availability of estimation methods is of paramount importance for practical application of a model. Several parametric and non-parametric estimation techniques have been developed for Hawkes process. In the next two sections we describe the Maximum Likelihood method for Hawkes processes and the non-parametric estimation technique of Bacry and Muzy (2016), which are the two estimation approach we took in the works presented in this thesis.

2.3.1 Parametric estimation: Maximum Likelihood

For point process models that are specified via the conditional intensity, the likelihood function has a fairly simple expression (Rubin, 1972). This makes maximum likelihood inference a popular choice for these models.

Suppose we are observing a point process in the interval $[0, T]$. Following Bowsher (2002), we denote $\lambda_t^* = \lambda(t|\mathcal{H}_{(-\infty,t]})$ the *complete* conditional intensity of the process, that is conditioned on the whole past history. We indicate with $\lambda_t = \lambda(t|\mathcal{H}_{(0,t]})$ the intensity conditional on the sole information from the observation interval $[0, t]$. Then, from Ogata (1978), the exact log-likelihood on the interval $[0, T]$ for a simple univariate stationary point process with intensity λ_t is

$$\ln \mathcal{L}\left(\theta; N_t\right) = -\int_0^T (\lambda_{\boldsymbol{\theta}}(s)\mathrm{d}s + \int_0^T \ln \lambda_{\boldsymbol{\theta}}(s)\mathrm{d}N_s, \tag{2.24}$$

where $\boldsymbol{\theta}$ denotes the parameters vector. Ogata (1978) defines also a theoretical log-likelihood under the information from the infinite past

$$\ln \mathcal{L}^*\left(\boldsymbol{\theta}; N_t\right) = -\int_0^T \lambda^*_{\boldsymbol{\theta}}(s)\mathrm{d}s + \int_0^T \ln \lambda^*_{\boldsymbol{\theta}}(s)\mathrm{d}N_s. \tag{2.25}$$

Ogata (1978) proves, under assumptions that are verified for Hawkes processes, some proprieties of the maximum likelihood estimator $\hat{\theta}^T$ obtained by maximizing (2.24) with respect to θ. In particular, $\hat{\theta}^T$ is found to be:

- *Consistent*, i.e. $\hat{\theta}^T$ converges to the true value θ_0 as $T \to \infty$:

$$\forall \epsilon > 0, \quad \lim_{T\to\infty} P\left[|\hat{\theta^T} - \theta_0| > \epsilon\right] = 0 \tag{2.26}$$

- *Asymptotically normal*, i.e.:

$$\sqrt{T}(\hat{\theta}^T - \theta_0) \to \mathcal{N}(0, I^{-1}(\theta_0)) \tag{2.27}$$

  with

$$I^{-1}(\theta) = \mathbb{E}\left[\frac{1}{\lambda^*_{\theta_0}(t)}\frac{\partial\lambda^*_{\theta_0}(t)}{\partial\theta_i}\frac{\partial\lambda^*_{\theta_0}(t)}{\partial\theta_j}\right] = -\mathbb{E}\left[\frac{1}{T}\frac{\partial^2 \ln \mathcal{L}^*_T(\theta_0)}{\partial\theta_i\partial\theta_j}\right] \tag{2.28}$$

- *Asymptotically efficient*, the following asymptotic relation exists between the Hessian of the likelihood function and the information matrix $I(\theta_0)$:

$$-\mathbb{E}\left[\frac{1}{T}\frac{\partial^2 \ln \mathcal{L}_T(\theta_0)}{\partial\theta_i\partial\theta_j}\right] \to I_{ij}(\theta_0) \tag{2.29}$$

  so the variance of the estimator reaches asymptotically the lower bound $I(\theta_0)^{-1}$.

These are essentially the standard asymptotic properties of maximum likelihood estimators, however there is an important difference: it is the complete intensity $\lambda^*(t)$ that enters (2.28), and not λ_t. Hence is not true in general that

$$I_{ij}(\theta) = -\mathbb{E}\left[\frac{1}{T}\frac{\partial^2 \ln \mathcal{L}_T(\theta_0)}{\partial\theta_i\partial\theta_j}\right],$$

albeit the convergence result (2.29) supports the use of the Hessian matrix

$$\frac{1}{T}\frac{\partial^2 \ln \mathcal{L}_T(\hat{\theta})}{\partial\theta_i\partial\theta_j}$$

to estimate $I_{ij}(\theta_0)$ (Bowsher, 2002).

As for the multivariate case, the log-likelihood can be computed as the sum of the likelihood of each coordinate:

$$\ln \mathcal{L}(\boldsymbol{\theta}; \{t_m\}_{m=1,\dots,N(T)}) = \sum_{i=1}^{D} \ln \mathcal{L}^i(\boldsymbol{\theta}; \{t_m\}) = -\sum_{i=1}^{D}\int_0^T \lambda^i_t \mathrm{d}T + \sum_{i=1}^{D}\int_0^T \ln \lambda^i_t \mathrm{d}N^i_t \tag{2.30}$$

In the particular case that the components of the intensity λ_t^i have no parameters in common, the maximization can be performed independently for each component. Differently from the simple univariate case, however, results on the proprieties of the ML estimators in the multivariate case do not exist at present.

In the case of Hawkes processes, where the intensity depends on the previous history of the process, the last term on the right hand side hides a double summation that makes the computational complexity of the expression grow as $\mathcal{O}(DN^2)$, where N is the number of events in the interval. This constitutes an obstacle to applications on large samples. However, for the particular case of a Hawkes process with kernel given by a sum of P exponentials of the form

$$\phi(t) = \sum_{q=1}^{P} \alpha_q e^{-\beta_q t}, \tag{2.31}$$

Ogata (1981) shown that the log likelihood can be computed recursively reducing computational complexity to $\mathcal{O}(DN)$. In fact, we have

$$\begin{aligned} \ln \mathcal{L} = &- \mu T - \sum_{q=1}^{P} \sum_{t_m} \frac{\alpha_q}{\beta_q} \left(1 - e^{-\beta_q (T - t_m)}\right) \\ &+ \sum_{i=1}^{N} \ln \left(\mu + \sum_{j=1}^{P} \sum_{k=1}^{i-1} \alpha_j e^{-\beta_q (t_m - t_k)} \right), \end{aligned} \tag{2.32}$$

where N denotes the number of events in the interval $[0, T]$. Now, setting

$$\begin{aligned} R_q(m) &= \sum_{k=1}^{i-1} e^{-\beta_q (t_m - t_k)} \\ &= e^{-\beta_q (t_m - t_{m-1})} \sum_{k=1}^{m-1} e^{-\beta_q (t_{m-1} - t_k)} \\ &= e^{-\beta_q (t_m - t_{m-1})} \left(1 + \sum_{k=1}^{i-2} e^{-\beta_q (t_{m-1} - t_k)}\right) \\ &= e^{-\beta_q (t_m - t_{m-1})} \left(1 + R_j(m-1)\right) \end{aligned} \tag{2.33}$$

the log likelihood expression (2.32) can be rewritten as:

$$\begin{aligned} \ln \mathcal{L} = &- \mu T - \sum_{q=1}^{P} \sum_{t_m} \frac{\alpha_q}{\beta_q} \left(1 - e^{-\beta_q (T - t_m)}\right) \\ &+ \sum_{m=1}^{N} \ln \left(\mu + \sum_{q=1}^{P} \alpha_q R_q(m) \right) \end{aligned} \tag{2.34}$$

with $R_j(1) = 0$, $\forall j$.

In the case where the point process is a multivariate Hawkes process, analogous expression and recursive formulae for the log-likelihood can be obtained. The drastic computational cost reduction allowed by the choice of exponential kernel is the main reason of their widespread use in parametric application of Hawkes processes.

We conclude by noting that another parametric approach using the Generalized Method of Moments was recently proposed by Da Fonseca and Zaatour (2014). Also anapproach

based on a least square cost function has been used in some applications (Reynaud-Bouret et al., 2013).

2.3.2 Non parametric estimation using Wiener-Hopf equation

In Bacry and Muzy (2016) and Bacry et al. (2016) a novel non-parametric estimation method for multivariate Hawkes process was presented. It has been shown that the method works well when large amounts of data are available and the kernel is non-localized, which is typically the case in high-frequency finance applications. This methodology leverages the fact, highlighted in Section 2.2.2, that a stationary multivariate Hawkes process is completely specified by its first- and second-order properties. In fact, the kernel Φ and the baseline intensity μ are related to the average intensity Λ and to the conditional law $g(t)$ through equations (2.14) and (2.19).

Hence, the method proposed by Bacry and Muzy (2016) goes as follows.

1. Empirical estimation of the average intensity Λ and of the conditional laws $g(t)$. Estimation is performed by replacing expectations in the definitions with empirical averages. The estimation of Λ is thus straightforward as it is simply given by

$$\Lambda^i = \frac{N_T^i}{T}$$

where N_T^i is the total number of type i event observed in the time interval T. Estimation of g is akin to the estimation of a conditional probability density, though it does not integrate to 1. Therefore kernel density estimation techniques can be used:

$$\hat{g}_*^{ij} = \hat{g}^{ij}(t) + \Lambda^i = \frac{1}{Jh} \sum_{m=1}^{J} \int_0^T \mathrm{d}N_u^i K\left(\frac{u - t_m^j - t}{h}\right) \tag{2.35}$$

where J is the number of events of type j in the observation interval T and K is a suitably chosen kernel function. This requires to set the bandwidth parameter h in advance. In Bacry and Muzy (2016) the choice of the bandwidth parameter as well as the estimation error is discussed. For financial data, which often present long-range dependencies, Bacry et al. (2016) propose a different scheme for the estimation of g which relies on a lin-log grid. More precisely, a time grid $\{T_l\}$ is chosen such that the approximations

$$g_*^{ij}(\frac{T_l + T_{l+1}}{2}) \approx \frac{1}{T_{l+1} - T_l} \int_{T_l}^{T_{l+1}} g_*^{ij}(s)\mathrm{d}s$$

$$\int_{T_l}^{T_{l+1}} g_*^{ij}(s)\mathrm{d}s \approx \frac{1}{J} \sum_{k=1}^{J} \sum_{k'=1}^{J} \mathbb{1}_{t_k^i - t_{k'}^j \in [T_l, T_{l+1}]}$$

hold. Then, $g_*^{ij}(\frac{T_l+T_{l+1}}{2})$ is approximated by

$$\hat{g}_*^{ij}(\frac{T_l + T_{l+1}}{2}) = \frac{1}{T_{l+1} - T_l} \frac{1}{J} \sum_{k=1}^{J} \sum_{k'=1}^{J} \mathbb{1}_{t_k^i - t_{k'}^j \in [T_l, T_{l+1}]} \tag{2.36}$$

and finally a linear interpolation is taken between the points $\frac{T_l+T_{l+1}}{2}$. The choice of the grid for typical financial applications is discussed in depth in Bacry et al. (2016).

2. Numerically solve the Wiener-Hopf system (2.19) using the empirical estimates in place of Λ and g. In Bacry and Muzy (2016) the authors propose the use of the Nystrom method. Fixing a set of $\{t_k\}_{1\leq k\leq Q}$ quadrature points along with a set of quadrature weights $\{w_k\}_{1\leq k\leq Q}$, each component of the system (2.19) is replaced by its discretized version

$$\hat{g}^{ij}(t_n) = \phi^{ij}(t_n) + \sum_{l=1}^{D}\sum_{k=1}^{Q} w_k \hat{g}^{il}(t_n - t_k)\phi^{lj}(t_k), \quad \forall\, n \in [0, Q] \tag{2.37}$$

by inverting the so obtained QD^2 linear system one recovers the values of the kernels at the quadrature points. Values for other points can then be obtained for example by linear interpolation. Bacry and Muzy (2016) show that Gaussian quadrature works well for a variety of cases. When power-law kernels are present, Bacry et al. (2016) suggest the use of a different quadrature scheme in the same spirit of the grid used for g.

3. Finally the values of μ can be recovered from (2.14) using the empirical estimated of Λ and the numerical integrals of the kernels obtained from the previous step.

We remark that (2.19) has at most one solution even if $g(t)$ is not generated by a Hawkes process. More precisely the following result holds (cf. Jaisson (2015) p. 142)

Theorem 2. *If $g \in L^1$ is the conditional law of a point process, then Equation* (2.19) *has one and only one solution in L^1.*

Hence, it is reasonable to look for a solution of (2.19) when $g(t)$ is estimated on empirical data, which is the case of interest in applications.

One of the work presented in this thesis (Chapter 3) makes use of this method and demonstrates its usefulness. There we discuss more in depth the procedure as well as some consequence of the result 2 with respect to inhibition effects and negative valued kernels.

We conclude by mentioning that the one presented here is not the only non-parametric method available for Hakes processes estimation. Indeed, the first method, based on a probabilistic interpretation of the kernel, was proposed by Lewis and Mohler (2011). Reynaud-Bouret et al. (2010) proposed instead a method based on the mimimization of a contrast function. This method is actually shown Bacry et al. (2015) to be very closely related to the one presented here.

2.4 Hawkes processes in finance

Several are the works that feature the use of Hawkes point processes found in the financial markets literature. Areas of application include volatility modeling, high frequency price models, price impact and optimal execution as well as order book models. Here we propose a brief, non-exhaustive overview and we refer to Bauwens and Hautsch (2009) and Bacry et al. (2015) for a thorough review.

One of the first applications of Hawkes processes to financial data is that of Bowsher (2007), who proposed to use Hawkes processes to model volatility clustering, building on the observation that volatility at the transaction level can be related to the number of events (trades, mid-price changes) that occur in a given time interval. He calibrated a one dimensional Hawkes process on intraday equity data, and he also accounted for intraday patterns and overnight effects by introducing additional parameters in the model.

In references (Filimonov and Sornette (2012); Hardiman et al. (2013); Hardiman and Bouchaud (2014); Filimonov and Sornette (2015)), authors used univariate Hawkes processes to assess the level of endogeneity and the stability of markets. Filimonov and Sornette (2012) noted that the branching ratio $n = \int_0^\infty \phi(t)\mathrm{d}t$ could be used as a measure of the market "reflexivity", that is the prevalence of internal feedback mechanism over the flow of external information. In fact, in a univariate Hawkes process, n can be directly interpreted as the fraction of events due to the self-exciting (endogenous) mechanism. The works of the two groups diverge on several methodological points and therefore do not fully agree on the conclusions, in particular on the degree to which financial market are close to critical ($n \approx 1$). In their first work on the topic, Filimonov and Sornette found that markets have become increasingly "critical" in the last decade. Hardiman *et al.* contested this claim and instead asserted that the degree of endogeneity has remained relatively stable. In their more recent work, Filimonov and Sornette pointed out several pitfalls in the estimation procedure of Hawkes processes that can lead to biased results.

Bacry et al. (2013) devised a two dimensional Hawkes model to describe up and down price variations. Within this framework they derived a close form expression for the signature plot[3] and estimating the model on high frequency data. They showed that it reproduces well the scaling behavior of the signature plot, which is known to strongly increase for $\tau \to 0$. Jaisson et al. (2015) showed that a modified version of the model of Bacry et al. (2013) converges at large scale towards a Heston price model, thus suggesting a path towards a unified model of the price that could describe both microstructural (mean-reverting) and diffusive (volatility clustering) stylized facts. Starting from the work of Bacry et al. (2013), Zheng et al. (2014) further refined the model by introducing bid-ask dependence. They achieved this by imposing some constraint on the Hawkes intensities in order to prevent the ask price to fall below the bid price. They developed a theoretical framework for such constrained Hawkes processes and proved several properties including a diffusive limit for the mid-price. By calibrating their model to financial data, they showed that it reproduces some features of the signature plot.

Hawkes-based market impact (Bouchaud, 2010) models have also been proposed. Bacry and Muzy (2014) presented a four-dimensional Hawkes model where two dimensions account for upward and downward mid-price moves and the remaining two model the buying and selling market order flow. The authors calibrated the model on high-frequency data using the non-parametric technique described in Section 2.3.2 and, among other things, showed that the impact of market orders on the price is mainly diagonal, i.e. buy (sell) market orders mainly triggers upward (downward) price moves. In Bacry et al. (2015) the authors introduced a class of models, named Hawkes Impact Models, where the intensities of upward and downward price moves are modeled as:

$$\lambda_t^+ = \mu + \int_{-\infty}^t \phi(s)\mathrm{d}N_s^- + \int_{t_0}^t f(r_s)g^+(s-t_0)\mathrm{d}s$$
$$\lambda_t^- = \mu + \int_{-\infty}^t \phi(s)\mathrm{d}N_s^+ + \int_{t_0}^t f(r_s)g^-(s-t_0)\mathrm{d}s$$

with N_t^+ (N_t^-) indicating the total number of upward (downward) price moves up to time t, and where r_t is the rate of (buying) orders from a meta-order strategy that starts at

[3] The signature plot is defined as the quadratic variation of the mid-price P_t at a scale τ

$$C(\tau) = \frac{1}{T}\sum_{i=0}^{T/\tau}(P_{(i+1)\tau} - P_{i\tau})^2.$$

t_0. Furthermore, $f(r_t)\mathrm{d}t$ specifies the infinitesimal impact of a buy order of volume $r_t\mathrm{d}t$, and g^+ and g^- are the impact kernel and the cross impact kernel respectively. Moreover the mean reverting behavior of the high-frequency price is encoded by the anti-diagonal Hawkes matrix ($\phi^{++} = \phi^{--} = 0$, and $\phi^{+-} = \phi^{-+} = \phi$). Within this model, the authors derived analytical formulae for the market impact curve and demonstrated that this simple model can reproduce the typical shape of the impact curve as measured from real meta-orders data.

Hewlett (2006) employed Hawkes processes to study the problem of optimal liquidation strategies, i.e. how to optimally execute a large order in order to minimize its impact on prices and thus the cost paid by the sender. He modeled the occurrence of buy and sell market orders on FX markets as a bivariate Hawkes process. He found self-excitation to be preponderant, with cross excitation between buy and sell orders being negligible. The author derived the expression of the expected future trade imbalance within the model. This, within a linear price impact model, allows one to determine the expected future price returns and the associated risk. By maximizing a mean-variance utility objective function, he demonstrated that it is possible to devise an optimal liquidation strategy. More recently, also Alfonsi and Blanc (2016) used Hawkes processes in the domain of optimal execution. They provided closed-form expressions for liquidation strategies when the flow of market orders is modeled by a two dimensional Hawkes process featuring a symmetric exponential kernel matrix.

Hawkes processes have also attracted interest as models for the order book dynamics. Large (2007) proposed to measure order book resiliency, i.e. the degree to which liquidity return after been consumed by a large trade, using the response kernel $G^{ij}(t)$ defined as

$$G^{ij}(t)\mathrm{d}t = \mathbb{E}\left[\mathrm{d}N_t^i|\mathcal{F}_o, \mathrm{d}N_0^j = 1\right] - \mathbb{E}\left[\mathrm{d}N_t^i|\mathcal{F}_0\right].$$

Large proved that G^{ij} is related to the Hawkes kernel matrix through the relation (2.10) that gives the definition of $\Psi(t)$. He calibrated a 10 dimensional Hawkes process with exponential kernels to level-1 order book data (market an limit orders that move the mid price, market an limit orders that do not move the mid price and cancellations) from the London Stock Exchange and he found that in over 60 percent of the cases the order book does not replenish after a large trade. He also showed that, when replenishment does occur, it is rather fast.

Bacry et al. (2016) performed a similar study of level-1 order book dynamics. They investigated mutual interactions between orders that move the mid price and limit, market and cancel orders that do not with a 8-dimensional Hawkes process. Differently from previous studies, they used the non-parametric estimation method exposed in Section 2.3.2 to perform the estimation. This allowed them to relax the constraint imposed by a particular choice of the kernel shape. They found that events mainly trigger other events of the same type, i.e. the kernel norm matrix has strong diagonal terms. One exception is constituted by orders that move the price, indeed the authors confirmed the observation that a price change at time t mostly triggers future price changes in the opposite direction.

A full Hawkes-driven order book model was proposed by Toke (2011), who generalized the zero intelligence, Poisson driven model of Smith et al. (2003). He found that the use of Hawkes processes instead of homogeneous Poisson processes greatly improve the fit to empirical data. Besides, he evidenced a market making behavior where new limit orders are submitted following the arrival of market orders across several markets. His study suggests that in a first approximation marker makers react to liquidity takers, whereas these act in a more independent manner. Muni Toke and Pomponio (2011) studied the dynamics of market orders that hit levels beyond the best available quote (trade-troughs).

They modeled the occurrence of trade-trough with a bivariate Hawkes process and found confirmation that, as for the full market order flow, the main effect is the the self-exciting one. In a theoretical work, Abergel and Jedidi (2015) investigated the scaling limits of a Hawkes-driven order book model. By leveraging the fact that Hawkes processes with exponential kernels possess the Markov property, they show that such Hawkes-driven order book is ergodic and that the price at long time scales is diffusive.

We finally mention the works of Bormetti et al. (2015), who studied the simultaneous occurrence of price jumps in a pool of 20 Italian stocks. They model the time arrivals of these co-jumps with a one dimensional Hawkes process and they assume that each stock price has then a probability p_i to jump each time a co-jump occurs. Finally, for each stock, a different one dimensional Hawkes process is used to describe the occurrence of idiosyncratic price jumps. The authors calibrated their model on the empirical data and showed that it nicely captures both the time clustering of jumps and the occurrence of synchronized co-jumps.

Chapter 3

The role of volume in orderbook dynamics

3.1 Introduction

Modeling the order book is a complicated task. Indeed, even in the simplest setting, multiple types of orders (limit orders, market orders, cancellations) arrive on the market at random times, and each one has a label that specifies the quantity to be negotiated (called the volume) and the price. Several are the challenges for a mathematical description of this system. First, the arrival times of the orders are not well described by a Poisson process (see for example Chakraborti et al. (2011) and references therein). In fact, the time durations between events are not independent but display strong correlations. Moreover, also the sequence of volumes presents nontrivial correlations (Gould et al., 2013). Furthermore, the interplay between them is relevant and need to be considered for a deep understanding of the system. Finally, the existence of different types of events that could influence each other increases further the complexity of the problem.

Understanding the order book dynamics is of great importance for several reasons ranging from optimization of execution strategies, to market design and regulation as well as for development of trading strategies and their risk management. Different approaches have been developed to gain a better understanding of order book dynamics. On stream of literature (Glosten and Milgrom, 1985; Kyle, 1985; Roşu, 2009) does so by modeling the behavior of individual traders. Another stream (Challet and Stinchcombe, 2001; Cont et al., 2010; Abergel and Jedidi, 2015) models the order flow with a random process and explores the resulting dynamics. Others (Hasbrouck, 1988; Biais et al., 1995; Bessembinder et al., 2009) focus on a more empirical analysis of the market data. We refer to Parlour and Seppi (2008) and Gould et al. (2013) for a more comprehensive review of the literature.

In this work we are interested in exploring the effects of order size on the high-frequency order book dynamics. We accomplish this goal by using the non-parametric estimation procedure for multivariate self-exciting Hawkes point processes developed by (Bacry and Muzy, 2016), which was introduced in Section 2.3.2. Two are the main contributions of our work. First, we show that the order size is an important variable in the modeling of the order book dynamics. Indeed, we show that the impact on the dynamics of an order depends strongly on its size and in particular we find that large orders have a stronger and more persistent effect than small ones. This is in accordance with the notion that larger orders carry more information as was noted also in the pioneering works of Easley and O'hara (1987) and Hasbrouck (1988). Our work relates also to the work of Biais et al.

(1995), who first observed a "diagonal effect", namely that orders of a specific type are more likely to be observed just after orders of the same type. By including volume into the analysis, we find that this effect is markedly stronger for same-type same-size orders. Moreover, we find that size and time dependence are not easily separable. In fact, our study provides evidence that simple multiplicative models that separate size dependence effects from time dependence ones are not fit to describe faithfully empirical observations. Second, we demonstrate the effectiveness of the nonparametric Hawkes processes approach as a tool to investigate the high-frequency dynamics on modern electronic order-driven markets. Thanks to this approach, we are able to identify typical reaction times, order size effects, as well as signatures of several trading strategies.

In the following we discuss the advantages of the Hawkes processes framework over competing approaches for the analysis on the system at hand. To frame the problem of our interest, let us first consider just one type of event, say the occurrence of a transaction. From a statistical perspective, the observation of transactions and the corresponding volumes represents the realization of a *marked* point process (t_i, v_i). Since we are interested in the effects of order size on the dynamics and vice-versa, we cannot model the volume as an independent mark process, but instead we need to consider their interdependence. This two-way feedback mechanism between marks and the underlying point process represents a challenge for modelers.

Some models for this system have been proposed in the Autoregressive Conditional Duration framework (Engle and Russell, 1998; Bauwens et al., 2004; Pacurar, 2008). These models describe the point process using a duration representation. The i-th duration d_i of a point process is defined as $d_i = t_i - t_{i-1}$. ACD models can be summarized as follows

$$\begin{aligned} d_i &= \psi_i \epsilon_i, \qquad \epsilon_i \sim \text{i.i.d.}(1, \sigma_\epsilon^2) \\ \psi_i &\equiv \mathbb{E}\left[d_i | \mathcal{F}_i; \theta_d\right] \end{aligned} \tag{3.1}$$

where θ_d is a set of parameters to be estimated. A parametric form has to be specified for ψ. When marks enter the picture, indicating with d_i the i-th duration and with x_i the relative mark, the data generating process can be written

$$(d_t, x_i) \sim f(d_i, x_i | \mathcal{F}_i) \tag{3.2}$$

where f is the joint distribution of the i-th duration and mark conditional on all the available information $\mathcal{F}_i$. As was pointed out in Engle (2000), this joint density can be rewritten as

$$(d_i, x_i) \sim f(d_i, x_i | \mathcal{F}_i) = g(d_i | \mathcal{F}_i) \cdot h(x_i | d_i, \mathcal{F}_i) \tag{3.3}$$

A possible way to construct a model is thus to specify f or g and h as well as the distribution of the error terms. This approach has been followed for example in Manganelli (2005). In this stream of literature, the event process is described using Engle's and Russels' ACD models. Then, a model has to be specified for each mark's conditional distribution.

These models allow for a lot of flexibility, however, they also presents some significant limitations. In fact, parameters do not have a clear interpretation so that the excitement structure is not easy to recover. More important, these models are strongly parametric and the choice of the error distribution can have a large influence on the model performance (Allen et al., 2009). More generally, a strong structure is imposed *a priori* with the selection of the functional form for g and h. Finally, a severe limitation of ACD models for our purposes is that duration based approaches are not easily generalized to multiple dimensions. Therefore, it is hard to extend these models to describe the whole order book dynamics.

Intensity-based approaches are on the contrary much more amenable to extension to multiple dimensions. Among intensity based models, Hawkes self-exciting processes (Hawkes, 1971) have been applied successfully in finance to model the irregular arrival in time of a number of event types (trades, quotes, etc.), see for instance Bowsher (2007); Bacry et al. (2013); Filimonov and Sornette (2015); Bormetti et al. (2015); Hardiman et al. (2013); Bauwens and Hautsch (2009); Rambaldi et al. (2015) and Bacry et al. (2015) for a recent review.

A simple way to incorporate volumes v in the dynamics is to assume that they are independent from the underlying point process and that the dependence on the volume factorizes,

$$\lambda_t = \mu + \int_{-\infty}^{t} f(v_s)\phi(t-s)\mathrm{d}N_s, \tag{3.4}$$

As shown by Bacry et al. (2016), the estimation of this model does not pose particular difficulties. However the predictions of this model are at odds with empirical data.

This evidence motivates our work and the development of a model that takes into account the complex dependence between time and volume. In the univariate case, the most general Hawkes model specification is

$$\lambda_t = \mu + \int_{-\infty}^{t} \phi(t-s; v_s)\mathrm{d}N_s \tag{3.5}$$

where the kernel ϕ depends on the time distance to previous events as well as on their marks[1]. The interdependence structure of the time t and the mark process v_t of the kernel is very difficult to estimate directly. A simple workaround consists in considering the volume process v_t as the superposition of D unmarked point processes, each of which corresponds to one of the possible D values $\{V^i\}_{1\leq i\leq D}$ that v_t can take. One can then consider the D multivariate Hawkes process defined by (2.9), where λ^i represents the intensity process for the events associated to the value V^i. The so-obtained model is an equivalent representation of the original process.

This approach is convenient as it allows to treat the case of dependent mark with the already available tools developed for multivariate point process. For a mark variable that can take an infinite number of values, some sort of binning is necessary in order to map the marked process to a multivariate one[2].

In this application we follow this approach and we use the non-parametric estimation method described in Section 2.3.2 to estimate the kernels and the baseline intensities from empirical data. We start by applying this model to unsigned trades, then by increasing the total dimension of the model we extend the study to signed trades (i.e. differentiating between buyer initiated transactions and seller initiated ones), and finally to the whole first level of the order book.

In the next section we discuss some details of the estimation procedure. In Section 3.3 we describe our dataset and some of its empirical properties. Section 3.4 contains the empirical results on trades when volume is taken into account. Section 3.5 extends our Hawkes based analytical tool to the first level of the order book by including also limit

[1]Note that in the case of volume we furthermore require that

$$\phi(t, v_1 + v_2) = \lim_{\Delta t\to 0} \phi(t, v_1) + \phi(t + \Delta t, v_2). \tag{3.6}$$

That is in the limit where two events become indistinguishable in time, their effect must be the same of the combined event.

[2]Let us point out that the two approaches are strictly equivalent in the limit of infinitely many bins.

orders and cancellations, and accounting for order size. Finally, in Section 3.6 we present our conclusions.

3.2 Kernel estimation procedure and inhibition effects

The main features of the non-parametric estimation technique we use were already presented in Section 2.3.2. As remarked in Bacry et al. (2016) some care is needed in the choice of the grid for the empirical estimation of $g(t)$ and of the quadrature scheme, since, in typical application to financial data, its mass is found to span many orders of magnitude in t. Moreover, rapid variation is observed at short lags (a few microseconds on today exchanges), while a much slower one is present at higher lags (of the order of one minute). Thus, it is important to have a finer grid around 0 in order to capture the behaviour of g.

To accomplish these goals, we follow Bacry et al. (2016) and use a grid with dense, linearly spaced intervals at short lags followed by log-spaced intervals at larger lags. The same type of grid is used also in the numerical solution of the Wiener-Hopf system (2.19).

Specifically, we can summarize our procedure as follows

Algorithm 3.2.1. *Non-parametric estimation of Hawkes kernels*

1. *For each day of trading, the times and volumes of orders are extracted.*

2. *Each event is assigned to a component of the multivariate process according to its volume.*

3. *The conditional law*

$$g^{ij}(t) = \frac{\mathbb{E}\left[\mathrm{d}N_t^i | \mathrm{d}N_0^j = 1\right]}{\mathrm{d}t} - \Lambda^i, \ \forall\, t > 0$$

   is estimated using empirical means. For this estimation a linear-log binning is used, with edges:

$$[0, \delta_{lin}, 2\delta_{lin}, 3\delta_{lin}, h_{min}, e^{\delta_{log}}, e^{2\delta_{log}}, e^{3\delta_{log}}, \ldots, h_{max}]$$

   where δ_{lin}, h_{min} and δ_{log} are user-defined parameters. In this work we fixed $h_{min} = 10^{-3}s$ and $h_{max} = 2 \cdot 10^4 s$. δ_{lin} and δ_{log} are chosen so to have 50 bins in the linear part and 1500 bins in the log-spaced part.

4. *The kernels are then estimated by solving the integral equation* (2.19)*. Again a lin-log quadrature scheme similar to the one above is used. All kernels are estimated on the domain $[0, x_{max}]$ and the quadrature points are*

$$[0, \epsilon_{lin}, 2\epsilon_{lin}, 3\epsilon_{lin}, x_{min}, e^{\epsilon_{log}}, e^{2\epsilon_{log}}, e^{3\epsilon_{log}}, \ldots, x_{max}]$$

   again, ϵ_{lin}, x_{min}, ϵ_{log} and x_{max} are user-defined parameters. We choose $x_{min} = 0.5 \cdot 10^{-3}s$ and $x_{max} = 0.5s$. ϵ_{lin} and ϵ_{log} are chosen so to have 80 bins in the linear part and 80 bins in the log-spaced part.

5. *Finally the baseline intensities μ_i are recovered by solving for μ the stationarity condition*

$$\Lambda = (\mathbb{I} - \|\Phi\|)^{-1}\mu$$

The linear Hawkes process (2.9) allows for self- and cross-excitation but not for inhibition. In fact, the non-negativity of the kernels implies that the occurrence of an event never causes a decrease in the intensity[3]. However, in real systems such as the order book we can expect the presence of inhibitory effects alongside exciting ones.

A proper account for inhibition in a Hawkes processes framework requires to abandon the linear specification (2.9) in favor of nonlinear formulations that ensure the positiveness of the intensity, while at the same time allowing the presence of negative valued kernels and thus inhibitory effects (Brémaud and Massoulié (1996); Sornette and Ouillon (2005); Bowsher (2007); Zheng et al. (2014)). This comes however at the expenses of mathematical tractability.

An interesting case for practical application, discussed in Brémaud and Massoulié (1996); Reynaud-Bouret et al. (2010) and Hansen et al. (2015) is

$$\lambda_t = \left(\mu + \int_{-\infty}^{t} \phi(t-s)\mathrm{d}N_s\right)^+ \tag{3.7}$$

where $(x)^+ = x$ if $x \geq 0$ and $(x)^+ = 0$ otherwise. This case is relevant because if $\mathbb{P}\left[\mu + \int_{-\infty}^{t} \phi(t-s)\mathrm{d}N_s < 0\right]$ is negligible, then it is almost equivalent to the linear case. If this requirement is met, then the non-parametric estimation procedures developed in Bacry and Muzy (2016) and Hansen et al. (2015) have been shown to lead to reliable results even in presence of moderate inhibition. On the contrary, the non-parametric approach developed in Lewis and Mohler (2011) leverages a probabilistic interpretation of the kernel and therefore leads always to non-negative estimates that rule out inhibition.

When we consider the nonparametric estimation method of Equation 2.19, it is important to stress that the solution ϕ of the equation above is not guaranteed to be positive. Moreover it is possible to show that the negativity of the conditional law g implies the negativity of the kernel ϕ. More precisely,

Proposition 3.2.1. *Let $g(t)$ and $\phi(t)$ be two matrices of functions in L_1 that satisfies*

$$g(t) = \phi(t) + \phi * g(t) \quad \forall\, t > 0$$

elementwise. Moreover, let the following assumptions hold:

1. *$g(t)$ is bounded;*
2. *the spectral radius of the matrix $||\phi||_1 = \{\|\phi^{ij}(t)\|_1\}$ is less than one;*
3. *$\phi^{ij}(t)$ is a causal function, i.e. $\phi^{ij}(t) = 0\ \forall\, t < 0,\ \forall i, j$;*

Then, if at least one entry of every column or row of g take on negative values for some $t > 0$, then also some elements of ϕ take on negative values.

Proof. See 3.A. □

Assumptions 1-3 basically require that $\phi(t)$ possesses the characteristics for being the kernel of a stationary Hawkes process and $g(t)$ those for being a conditional law. Note in particular that condition 2 was shown by Brémaud and Massoulié (1996) to hold as a stability condition even when λ_t is a nonlinear positive Lipschitz function of $\phi * \mathrm{d}N$ and where the kernels can take negative values.

[3]This also implies that in a linear Hawkes process framework there cannot be a minimum time separation between events as there is always a positive probability of a new event arrival.

Whenever some inhibition effect is present, the conditional laws (2.17) will take negative values. In fact, we would have that the expected arrival rate of type i event given that event j has occurred is lower than the unconditional rate Λ_i. Therefore, the above proposition suggests that if some inhibition effect is present, then with our procedure we should see some (at least partially) negative kernel. We will actually encounter some of them, particularly in section 3.5.

3.3 Data and main statistics

In this Section we describe our dataset and we provide an empirical analysis of its main characteristics.

3.3.1 Data

In this paper we use level-I order book data provided by QuantHouse EUROPE/ASIA (http://www.quanthouse.com) for two future contracts, namely the German Bund Future and the DAX Future[4]. The data span the period from July 2013 to November 2014.

The dataset consists of snapshots of the first level of the order book, each with a timestamp indicating the record time. For each day, a line is added by the market every time a change in the first level of the order book is registered. Outstanding quantities at the best quotes as well as the corresponding prices and the number of proposals are specified for each line. In the cases where the change in the book is determined by the occurrence of a trade, two lines are added with the same timestamp, the first one with the state of the first level immediately before the trade together with the trade details (size, side, price), and the other with the state of the first order book level immediately after the trade. Timestamps are provided by the market itself with microsecond precision.

By comparing consecutive lines, we can then precisely infer the type of event that led to the change, i.e. market order, limit order or cancellation using the quantity and price information available in the dataset. For instance, if the available quantity at the best ask increases or a new best ask is generated inside the spread, we can say that a limit order has arrived that determined the change. Thanks to the additional line present when a trade occurs, we can also discriminate between a quantity decrease or price move due to a trade from that due to a cancellation. For example, if the available quantity at the best ask decreases and no trade line was recorded, we ascribe the effect to a cancel order.

We process the whole dataset to extract limit order, cancel order and trade events. We remark once again that we consider only orders at the best quotes (first level of the order book). It is important also to remark that we treat multiple orders that happen at the same time and on the same side of the order book as a single event (for instance a market order that hits two limit orders present in the book at the same price is regarded as a single trade). It is still possible to have two simultaneous events on opposite side of the book, i.e. one at the ask and one at the bid. This occurrence is however very rare thanks to the fine time resolution of the dataset, resulting in a ratio of less than 0.2% of simultaneous events to the total number of events.

3.3.2 Empirical properties of durations and volumes

Figure 3.1 shows the histograms of all inter-event times (right panel) and of inter-trades times alone (left panel). We never observe events closer than about 10μs, and this may

[4]This is the same dataset used in Bacry et al. (2016)

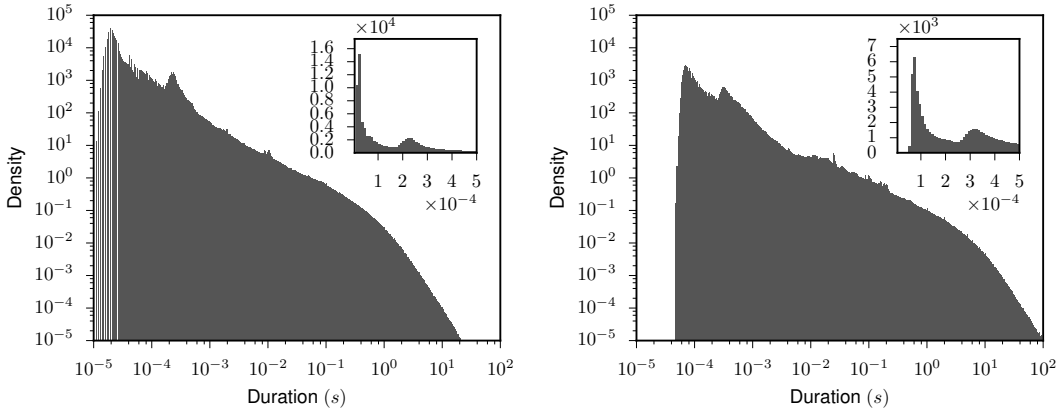

Figure 3.1: Histograms of the inter-event times. All events (left) and only trades (right). Main plots use log-spaced bins, while the inset show equally spaced histograms for short durations. Data refer to the Bund future, no significant difference is observed for DAX future data.

be a minimum technical delay due to data processing. For trades alone, the minimum observed time distance is larger, at about 50μs, possibly due to longer processing time needed in case of transactions.

In the duration distributions, we note a major peak around 30μs when all events are considered and around $70 - 100\mu$s when only trades are taken into account. There is a second noticeable peak between 200 and $300\mu s$ for all events and around $300 - 400\mu$s for trades. Finally, even quite far in the tail, small peaks are visible. Those might be the results of order splitting and automatic trading.

Figure 3.2 shows the empirical distribution of the trade sizes, measured in number of contracts, for the DAX and Bund futures. A positive sign indicates buyer initiated transaction while a negative sign stands for seller initiated ones. For both assets we note sharp peaks in correspondence of "round" order sizes such as 1, 50, 100. In the DAX future market, transactions of size above 100 contracts are very infrequent, while in the Bund Future market volumes of 1000 contracts are not uncommon. This difference is probably due to the fact that the Bund belongs to the class of the so called "large-tick" assets, while the DAX future is considered a "small tick" one. That is, in the former case, the spread is almost always equal to one tick, thus resulting in fewer price changes and larger available volumes at the best quotes (Eisler et al., 2012).

We also examined serial correlations in the trade sizes both in trade time and in real (continuous time). We find that autocorrelation in trade time is very small, and that in real time can be almost entirely explained by correlation in the number of transactions.

3.3.3 High frequency characteristics of Eurex market

The $300\mu s$ peak observed in the duration distribution is likely related to technical characteristics of the Eurex market. In particular, on the exchange website is reported that

"For futures orders Eurex Exchange currently offers customers (daily) average roundtrip times as low as 0.2-0.35 milliseconds in co-location"[5].

[5]From `http://www.eurexchange.com/exchange-en/technology/co-location-services` visited on 7 May 2015.

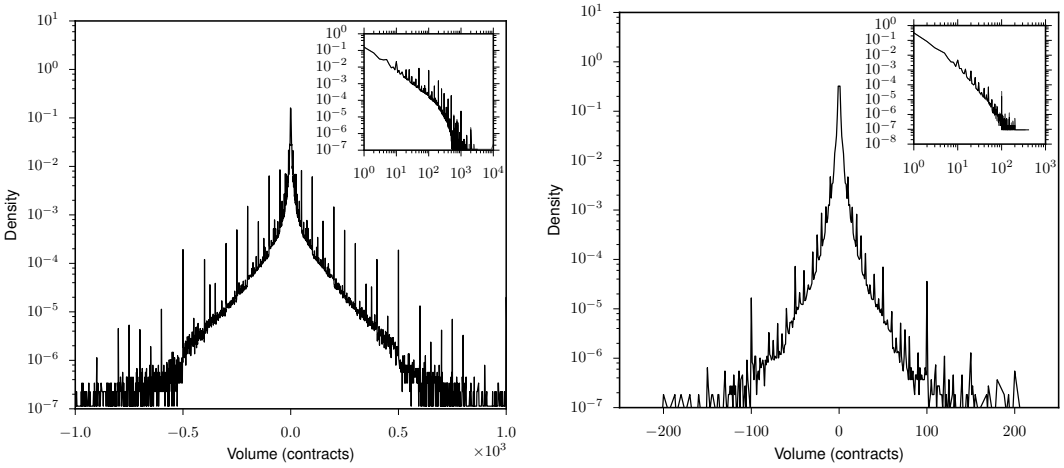

Figure 3.2: Empirical distribution of the trade volumes for the Bund (left) and DAX (right) futures. The inset shows the distribution in log-log scale where buy and sell are aggregated.

This means that the minimum time it takes for an agent to react to an event on the market is around $300\mu s$, and explains the bump we see in the duration distribution. As a consequence, no direct causal relationship can hold between events that are closer than this timescale, except if they are the result of the same agent's actions. We need to keep this fact in mind when analyzing the result of our study in the next sections.

3.4 Mapping Traded volume to multivariate Hawkes processes

We first apply the multivariate Hawkes framework to the series of trades, thus disregarding other types of order book events. We map trades of different sizes to separate the components of a multivariate Hawkes point process. To keep the total dimension of the resulting multivariate process manageable, we divide the range of order sizes into a small number of bins (six). Once each transaction has been assigned to a component of the multivariate Hawkes process, we use the non-parametric estimation method described before to determine the kernels and baseline intensities of the Hawkes process. The whole trading day was used (from 8am to 10pm) without taking into account intraday seasonality (the results do not change dramatically by restricting the analysis to the most active hours see 3.B).

3.4.1 Unsigned Trades

We begin by considering unsigned trades, i.e. we do not distinguish between buyer initiated trades and seller initiated ones. We fix as many bins for the unsigned volume as component of our multivariate Hawkes model. Then we assign each transaction to a component based on the bin in which its volume falls. In Table 3.1 we report the bin choice as well as the average number of events per day in each bin for the Bund and DAX futures respectively. We note the overwhelming prevalence of size one trades in the case of the DAX future, where they represent more than 60% of the total.

	Volume (contracts)	Avg. N	Fraction (%)	Volume (contracts)	Avg. N	Fraction (%)
B_1	1	11049	31.3	1	28140	64.5
B_2	2	4979	14.1	2	7308	16.8
B_3	3	2169	6.2	3	2753	6.3
B_4	$(3, 7]$	5571	15.8	$(3, 5]$	2959	6.8
B_5	$(7, 20]$	5619	15.9	$(5, 10]$	1700	3.9
B_6	$(20, \infty)$	5876	16.7	$(10, \infty)$	735	1.7

Table 3.1: Bund (left) and DAX (right) Future: Binning scheme and average number N of events per day in each bin in the six dimensional case

Conditional laws The first step of the procedure is the estimation of the conditional laws $g^{ij}(t)$. In the left panels of Figure 3.3 we plot the diagonal conditional laws estimates $g^{ii}(t)$ for the Bund and DAX futures. These functions attain their largest values at very short lags. In particular, we note two main peaks, the first around 100 μs and the second around 300 μs. We also note that this second peak tends to become more relevant for larger volumes. Another interesting feature that emerges is the presence of several sharp peaks at "round" time values (0.1 s, 1 s, ...). These peaks are more evident for the smallest volumes and they are likely the result of automatic trading and order splitting. For lags larger than about one second the diagonal conditional laws show a power law decay with exponents smaller than one. The right panels of the same figures show instead the terms $g^{6j}(t)$, i.e. those that measure the effect of the different volumes on large trades. At short lags we again note the two major peaks described before. Here we note how the second peak is more pronounced when large-large trades are involved. We will comment more on this when examining the kernels.

Kernels shape Once the conditional laws have been estimated, we solve the Wiener-Hopf system (2.19) for the kernels. A total of 36 kernels are estimated non parametrically and each function $\phi^{ij}(t)$ is estimated for values of t up to 0.5s. Beyond this value the estimation becomes very noisy. Given their number, in examining them, we focus on those contributing to the intensity of the smallest and largest bins. In Figure 3.4 we plot the functions ϕ^{1j} and ϕ^{6j} for the Bund and the DAX futures. To begin with, we remark that taking the volume of the trades into account does add new information. Indeed, the shapes of the kernels involving different sizes are markedly distinct. This has the important consequence that models where the time dependence is separated from size dependence appear to be inadequate.

We now examine the most relevant facts that emerge from the analysis of Figure 3.4. First, we note that the influence of large trades is more intense and more persistent at longer time scales. It thus takes more time for the market to forget about a large trade than it takes for a small one. This can be linked to the fact that the execution of a large trade is more informative for the rest of the market than that of a small one. This is even more relevant in the DAX case where large trades are particularly rare.

Second, we observe sharp peaks located around $300\mu s$ in the right figures, i.e. those depicting the influence over large trades. These peaks are visible also in the left plots, albeit they are not the main ones there. In light of Section 3.3.3, we interpret these peaks as corresponding to the market reaction to some trader's action. This is consistent with

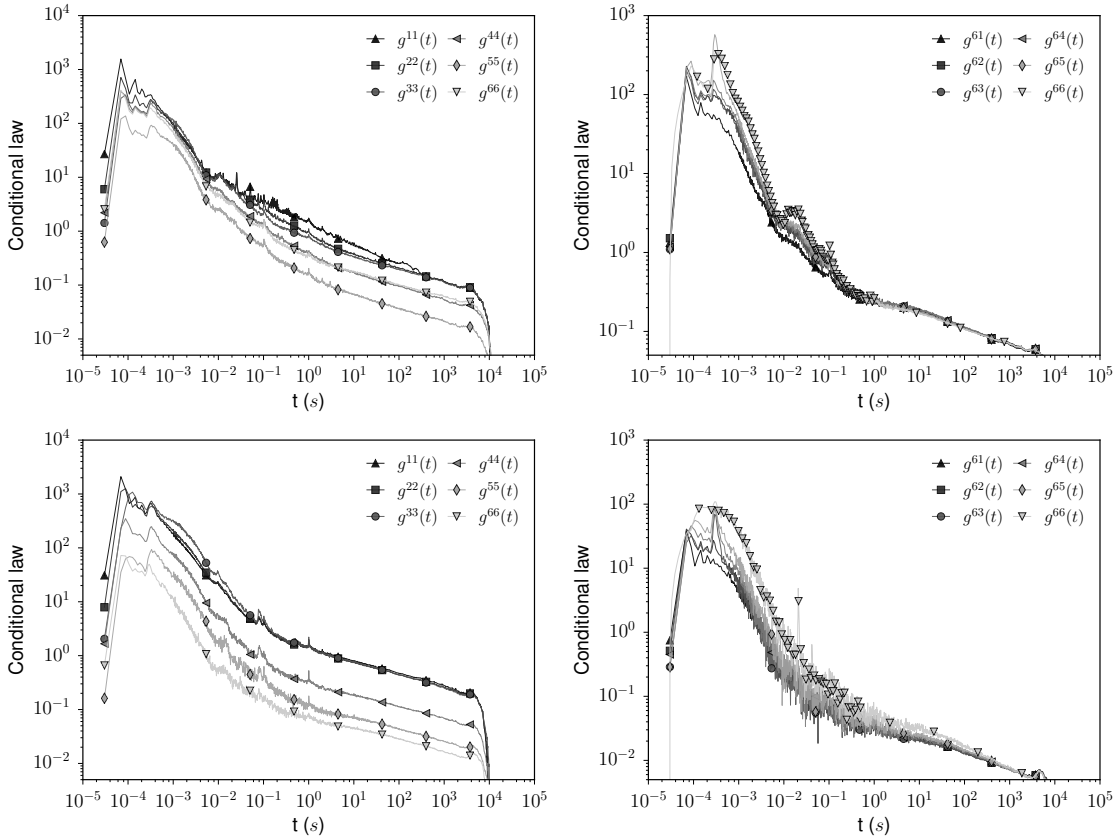

Figure 3.3: Conditional law estimates. Left column: diagonal terms. Right: the row corresponding to largest volumes. Figures on top refers to the Bund future, those at the bottom to the DAX future.

the fact that the $300\mu s$ peaks are more significant when large trades are involved. Indeed, a single agent does not typically execute several large order one after the other, and we know that at least the reaction time is needed for an interaction.

The large peaks found on the left figures at $100\mu s$ are instead likely the result of order splitting, i.e. the same trader executing several consecutive orders. Actually, we also found the peak at $100\mu s$ to be stronger in small size diagonal term ϕ^{11}, ϕ^{22}, which further corroborate the role played by order splitting as traders do not change the size too much when splitting.

Kernels norms To complete the analysis of the kernel matrix, we report in Table 3.2 the norms $n_{ij} = \int_0^\infty \phi^{ij}(t)\mathrm{d}t$ of the kernels together with the rescaled norms $\tilde{n}^{ij} = \frac{\Lambda_j}{\Lambda_i} n^{ij} = 1 - \frac{\mu_i}{\Lambda_i}$. It is important to underline the different meaning of these two quantities. The norm n^{ij} is the average number of events of type i triggered by *one* event of type j.Thus, the larger this number is, the bigger is the influence of events of type j on events of type i. The rescaled norms represent instead the fraction of the average intensity Λ_i attributable to excitation from component j. The former is useful to understand what component i is more affected by the arrival of an event j (reading by column) and also how many events of type i are expected after an event of type j (reading by row). The latter instead tells us how much of the activity of component i is explained by component j (reading by row). The comparison of the two matrices is particularly relevant when the components have very heterogeneous average intensities. In the same spirit of Biais et al. (1995), the largest

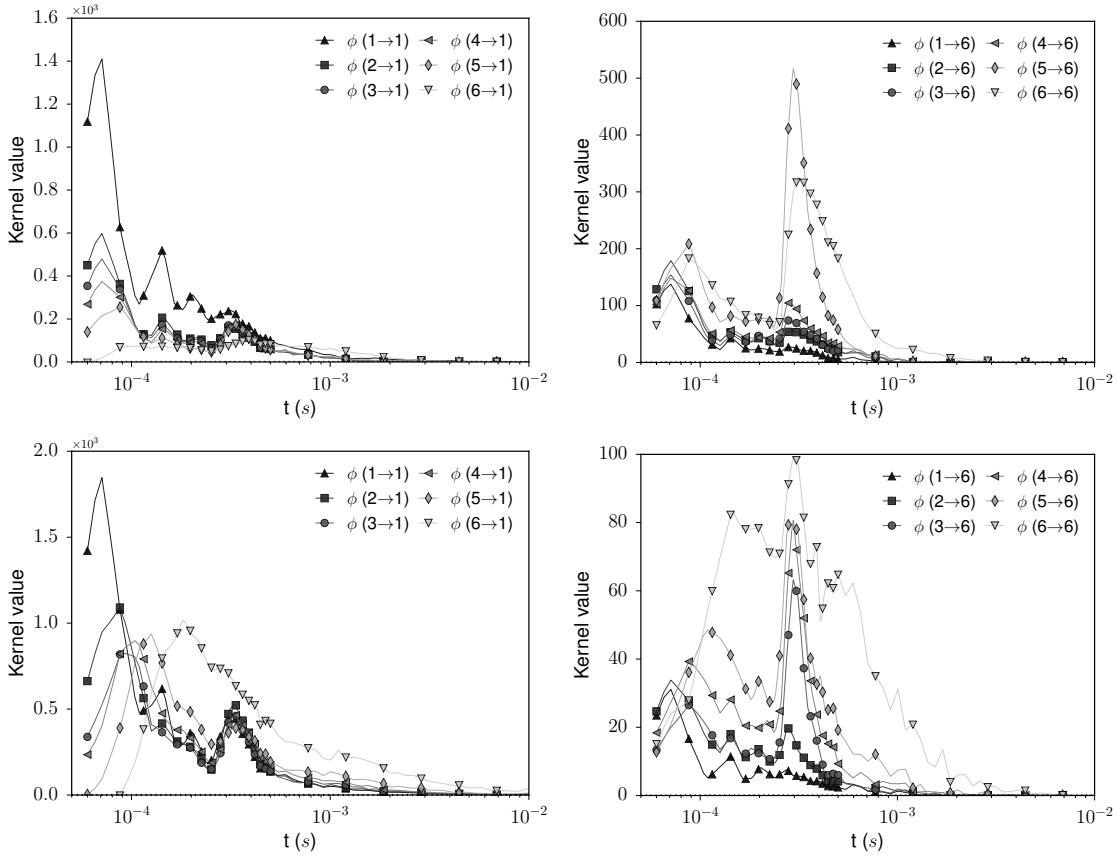

Figure 3.4: Kernel estimates. Figures at the top refers to the Bund, while those at the bottom to the DAX. Left hand side figures: influence on smallest volume. Right hand side: the row corresponding to largest volumes.

three values of each column of the kernel norm matrix are reported in bold face to help visualizing the structure of the matrix. For the rescaled norm matrix, we instead highlight in bold face the largest three values in each row.

In the case of the Bund, the two matrices appear quite similar since, by construction, the average intensities are almost equal across all the components. We observe that self-excitation is preponderant, followed in importance by the excitation from large volumes. The components of the DAX model have instead very different average intensities, with the first component largely dominating. It appears that for this small-tick asset the unusual occurrence of a large trade has a dramatic impact on the activity. However, the scarcity of such large orders is such that when we consider the fraction of each intensity attributable to excitation from other components, size one trades clearly make up the largest amount.

Baseline intensities Lastly, we examine the contribution of the baseline intensities μ_i. In Table 3.3 we report the ratio between the exogenous intensity μ_i and the average intensity Λ_i for each bin. Since the kernels are estimated only up to $t = 0.5s$, the norms can be underestimated, because the contribution from the tail is disregarded. With this in mind, for both assets it appears nevertheless that large trades are more exogenously driven than the others. The exogenous fractions μ_i/Λ_i for the DAX are higher than those obtained for the Bund, with all but size one trades having exogenous components accounting for 40-50% of the average intensity.

Bund						
	≤ 1	≤ 2	≤ 3	≤ 7	≤ 20	> 20
≤ 1	**0.46**	**0.10**	**0.09**	**0.08**	0.08	0.16
≤ 2	**0.04**	**0.36**	0.08	0.07	0.07	0.11
≤ 3	0.02	0.04	**0.20**	0.05	0.04	0.06
≤ 7	**0.03**	**0.07**	**0.11**	**0.29**	**0.12**	**0.17**
≤ 20	0.02	0.05	0.07	**0.09**	**0.27**	**0.20**
> 20	0.02	0.03	0.03	0.04	**0.09**	**0.22**
	≤ 1	≤ 2	≤ 3	≤ 7	≤ 20	> 20
≤ 1	**0.46**	**0.05**	0.02	0.04	0.04	**0.08**
≤ 2	**0.09**	**0.36**	0.04	0.08	0.08	**0.13**
≤ 3	0.08	0.08	**0.20**	**0.14**	0.11	**0.16**
≤ 7	0.06	0.06	0.04	**0.29**	**0.13**	**0.18**
≤ 20	0.04	0.05	0.03	**0.09**	**0.27**	**0.21**
> 20	0.03	0.02	0.01	**0.04**	**0.09**	**0.22**

DAX						
	≤ 1	≤ 2	≤ 3	≤ 5	≤ 10	> 10
≤ 1	**0.41**	**0.36**	**0.38**	**0.44**	**0.61**	**1.27**
≤ 2	**0.07**	**0.18**	**0.14**	**0.15**	**0.19**	**0.33**
≤ 3	**0.02**	**0.04**	**0.08**	0.06	0.07	0.11
≤ 5	0.02	0.04	0.05	**0.10**	**0.08**	**0.12**
≤ 10	0.01	0.02	0.03	0.04	0.07	0.09
> 10	0.01	0.01	0.01	0.02	0.03	0.08
	≤ 1	≤ 2	≤ 3	≤ 5	≤ 10	> 10
≤ 1	**0.41**	**0.09**	0.04	**0.05**	0.04	0.03
≤ 2	**0.26**	**0.18**	0.05	**0.06**	0.05	0.03
≤ 3	**0.22**	**0.11**	**0.08**	0.06	0.04	0.03
≤ 5	**0.20**	**0.09**	0.04	**0.10**	0.04	0.03
≤ 10	**0.20**	**0.08**	0.05	0.06	**0.07**	0.04
> 10	**0.23**	**0.08**	0.04	0.08	0.07	**0.08**

Table 3.2: Kernel norm matrices in the unsigned trades case. For each asset we report in the top panel the kernel norms $n_{ij} = \int_0^\infty \phi^{ij}(t)\mathrm{d}t$ and in the bottom panel the rescaled kernel norms $\tilde{n}^{ij} = \frac{\Lambda_j}{\Lambda_i} n^{ij}$. To highlight the structure, the largest three values in each column (row) of the upper (lower) panel are in bold face.

			Bund			
Bin	1	2	3	$(3,7]$	$(7,20]$	>20
$R_i = \frac{\mu_i}{\Lambda_i}$	31.62 %	22.58 %	23.76 %	23.63 %	31.06 %	58.93 %
			DAX			
Bin	1	2	3	$(3,5]$	$(5,10]$	>10
$R_i = \frac{\mu_i}{\Lambda_i}$	34.44 %	36.98 %	45.84 %	49.81 %	49.79 %	42.14 %

Table 3.3: Ratios μ_i/Λ_i for the two assets examined. All values in percent.

	Volumes (contracts)	Avg. N events	Fraction (%)	Avg. N events	Fraction (%)
S_1	1	5678	16.10	14066	32.24
S_2	$(1,3]$	3558	10.09	5047	11.57
S_3	$(3,10]$	4132	11.72	2340	5.37
S_4	$(10,\infty)$	4397	12.47	376	0.86
B_1	1	5371	15.23	14082	32.28
B_2	$(1,3]$	3591	10.18	5019	11.50
B_3	$(3,10]$	4152	11.77	2322	5.32
B_4	$(10,\infty)$	4384	12.43	370	0.85

Table 3.4: Bund (left) and DAX (right) Future: Binning scheme and average number of events per day in each bin in the signed case.

3.4.2 Signed Trades

We now take the "sign" of the transaction into account. If a market order hits the ask side of the book we label it as a buy trade, while if it hits the bid side we call it a sell trade. To analyze the interaction between different order sizes, we proceed as before by binning the trades according to their volume. To keep the overall dimension of the model manageable we use four bins of volume for each side (buy/sell). The binning schemes for the two assets examined are detailed in Table 3.4.

In Figure 3.5 we plot the kernels for the sell components obtained on Bund (top 4 panels) and DAX (bottom 4 panels) data, the results on the buy components are almost exactly symmetrical and thus not shown. We again observe two main peaks at $100\mu s$ and $300\mu s$ respectively, with the latter being more relevant when large trades are involved. Now that trade sign is taken into account we immediately notice that short term influence is almost entirely due to same-sign trades, in line with the hypothesis that is the result of order splitting. Excitation from opposite side trades is negligible, except for the largest ones that show some influence around the millisecond timescale. In the DAX figures, the excitation from large trades of opposite sign over larger trades is even more evident and the cross excitement starts to become relevant around time lags of 0.5 milliseconds. A possible explanation is that at this timescale, enough time has passed for the sign of the order to become less relevant thus the contribution is towards a general increase in activity rather than targeting specifically more buying or selling.

In Table 3.5 we report the kernel norms for the Bund and the DAX. We can divide the matrix into the four sectors sell-sell, sell-buy, buy-sell and buy-buy. The two same-sign blocks and the two opposite-sign blocks appear identical to each other as we would

expect, i.e the buy-sell symmetry is respected. Moreover, opposite-sign trades provide almost no contribution, with the notable exception of the column of largest size trades. In the Bund case, we find the same-sign blocks to be almost perfectly symmetric. Indeed, the asymmetry observed in the unsigned case, that is that large trades influence small ones more than the other way round, is now found to be due to excitation from opposite-sign large trades. In DAX future instead, each of the four main blocks is markedly asymmetric, with influence of large trades being more pronounced. As already noted this is related to the much lower frequency of large orders.

Finally, we mention that the baseline intensities confirm what emerged for the unsigned case, namely that a significant fraction of the trades is of exogenous origin and this fraction is larger for big trades.

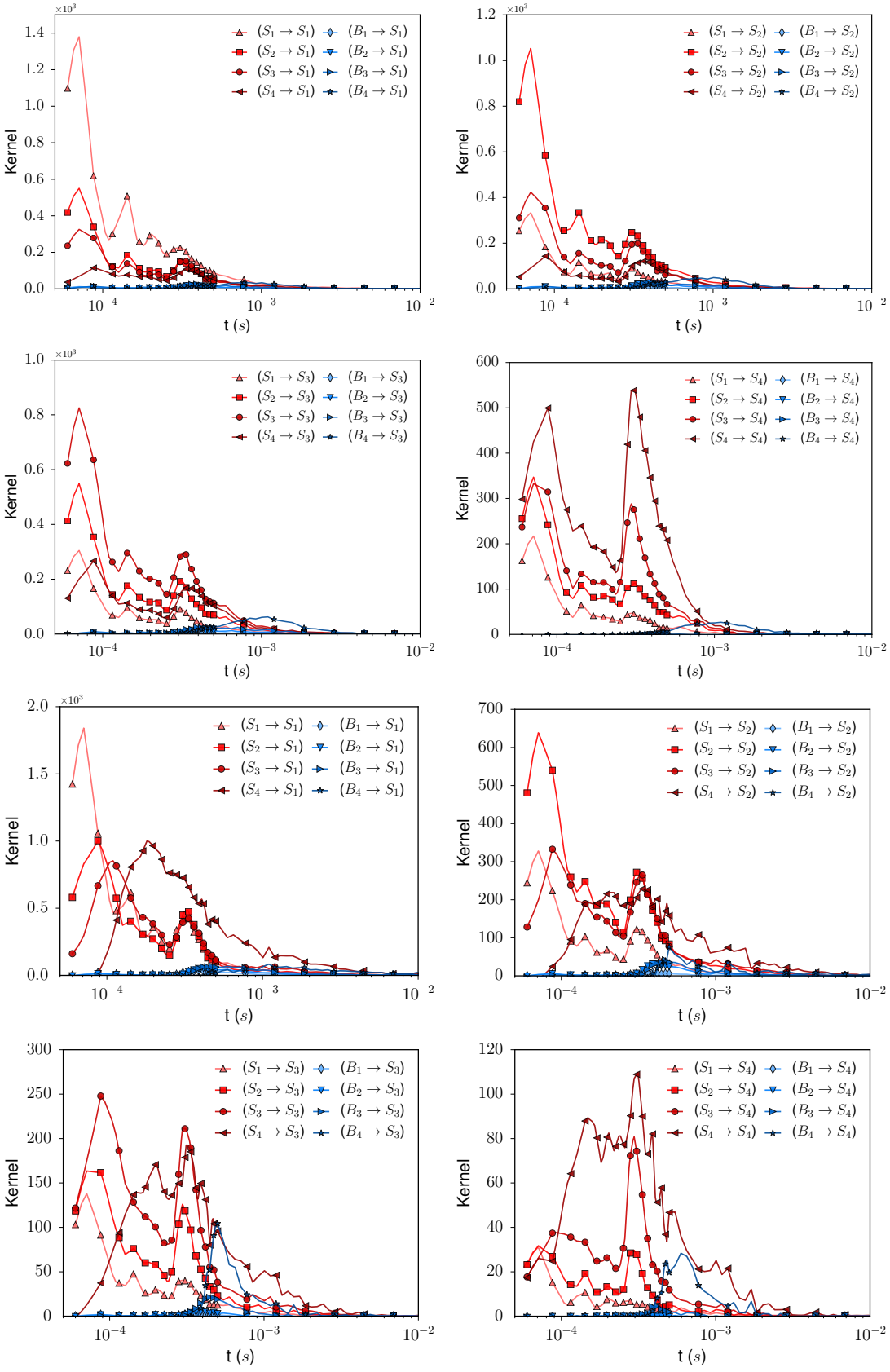

Figure 3.5: Estimated kernels relative to the four sell components in the signed trades model. Top four panels refer to the Bund, bottom four to the DAX.

Bund								
	S_1	S_2	S_3	S_4	B_1	B_2	B_3	B_4
S_1	**0.44**	**0.09**	0.05	0.06	0.02	0.01	0.02	0.08
S_2	**0.06**	**0.34**	**0.09**	0.06	0.00	0.02	0.02	0.08
S_3	**0.04**	**0.10**	**0.30**	**0.11**	0.00	0.01	0.04	**0.12**
S_4	0.03	0.06	**0.10**	**0.24**	-0.00	-0.01	0.00	0.08
B_1	0.02	0.01	0.02	0.07	**0.43**	**0.09**	0.06	0.06
B_2	0.00	0.02	0.03	0.08	**0.06**	**0.34**	**0.09**	0.07
B_3	0.00	0.01	0.04	**0.12**	**0.04**	**0.09**	**0.30**	**0.12**
B_4	0.00	-0.01	-0.00	0.08	0.03	0.06	**0.10**	**0.25**

	S_1	S_2	S_3	S_4	B_1	B_2	B_3	B_4
S_1	**0.44**	**0.06**	0.04	0.04	0.01	0.01	0.01	**0.06**
S_2	0.09	**0.34**	**0.10**	0.08	0.00	0.02	0.03	**0.10**
S_3	0.06	0.08	**0.30**	**0.12**	0.00	0.01	0.04	**0.13**
S_4	0.04	0.05	**0.10**	**0.24**	-0.00	-0.00	0.00	**0.08**
B_1	0.02	0.01	0.02	**0.06**	**0.43**	**0.06**	0.04	0.05
B_2	0.01	0.02	0.03	**0.10**	0.08	**0.34**	**0.11**	0.08
B_3	0.00	0.01	0.04	**0.13**	0.05	0.08	**0.30**	**0.12**
B_4	0.00	-0.00	-0.00	**0.08**	0.03	0.05	**0.10**	**0.25**

DAX								
	S_1	S_2	S_3	S_4	B_1	B_2	B_3	B_4
S_1	**0.34**	**0.28**	**0.36**	**0.86**	**0.07**	**0.08**	**0.14**	**0.43**
S_2	**0.07**	**0.19**	**0.17**	**0.29**	0.01	0.04	0.06	0.15
S_3	0.03	0.06	0.12	0.16	0.00	0.00	0.02	0.05
S_4	0.01	0.01	0.02	0.08	-0.00	-0.00	-0.00	0.01
B_1	**0.07**	**0.08**	**0.14**	**0.46**	**0.34**	**0.28**	**0.36**	**0.82**
B_2	0.01	0.03	0.06	0.15	**0.07**	**0.18**	**0.17**	**0.28**
B_3	0.00	0.00	0.01	0.05	0.03	0.06	0.12	0.17
B_4	-0.00	-0.00	-0.00	0.01	0.01	0.01	0.02	0.06

	S_1	S_2	S_3	S_4	B_1	B_2	B_3	B_4
S_1	**0.34**	**0.10**	0.06	0.02	**0.07**	0.03	0.02	0.01
S_2	**0.20**	**0.19**	**0.08**	0.02	0.04	0.03	0.03	0.01
S_3	**0.18**	**0.12**	**0.12**	0.02	0.01	0.01	0.02	0.01
S_4	**0.22**	**0.15**	**0.15**	0.08	-0.01	-0.02	-0.00	0.01
B_1	**0.07**	0.03	0.02	0.01	**0.34**	**0.10**	0.06	0.02
B_2	0.04	0.03	0.03	0.01	**0.21**	**0.18**	**0.08**	0.02
B_3	0.02	0.01	0.01	0.01	**0.19**	**0.12**	**0.12**	0.03
B_4	-0.02	-0.01	-0.00	0.01	**0.25**	**0.13**	**0.15**	0.06

Table 3.5: Kernel norm matrices in the signed trades case. For each asset we report in the top panel the kernel norms $n_{ij} = \int_0^\infty \phi^{ij}(t)\mathrm{d}t$ and in the bottom panel the rescaled kernel norms $\tilde{n}^{ij} = \frac{\Lambda_j}{\Lambda_i} n^{ij}$. To highlight the structure, the largest three values in each column (row) of the upper (lower) panel are in bold face.

					Bund							
	L_1^a	L_2^a	L_3^a	L_4^a	C_1^a	C_2^a	C_3^a	C_4^a	T_1^a	T_2^a	T_3^a	T_4^a
mil	11.46	15.24	12.22	4.41	10.61	13.33	10.05	3.66	1.80	1.24	1.44	1.53
%	6.60	8.78	7.04	2.54	6.11	7.68	5.79	2.11	1.04	0.71	0.83	0.88
	L_1^b	L_2^b	L_3^b	L_4^b	C_1^b	C_2^b	C_3^b	C_4^b	T_1^b	T_2^b	T_3^b	T_4^b
mil	11.49	15.24	12.19	4.28	10.64	13.30	9.92	3.50	1.88	1.22	1.43	1.53
%	6.62	8.78	7.02	2.47	6.13	7.66	5.71	2.02	1.08	0.70	0.82	0.88
					DAX							
	L_1^a	L_2^a	L_3^a	L_4^a	C_1^a	C_2^a	C_3^a	C_4^a	T_1^a	T_2^a	T_3^a	T_4^a
mil	29.07	3.84	0.71	0.05	30.29	2.45	2.53	0.76	5.14	1.82	0.82	0.13
%	18.80	2.48	0.46	0.03	19.59	1.58	1.64	0.49	3.32	1.18	0.53	0.08
	L_1^b	L_2^b	L_3^b	L_4^b	C_1^b	C_2^b	C_3^b	C_4^b	T_1^b	T_2^b	T_3^b	T_4^b
mil	28.88	3.81	0.71	0.06	29.95	2.41	2.51	0.76	5.12	1.82	0.83	0.13
%	18.68	2.46	0.46	0.04	19.37	1.56	1.62	0.49	3.31	1.18	0.54	0.08

Table 3.6: Total number of events in our database, divided by type and size.

3.5 Level one order book with volume

In this section we extend the model beyond trades including a total of three types of events, namely limit orders, cancellations, and trades. For each one, we differentiate between ask and bid side and we consider four bins of volume with edges at 1, 3, and 10 contracts. The resulting Hawkes model has thus 24 components, and 576 kernels. The total number of events recorded in each component is detailed in Table 3.6 for both assets. We use the same estimation procedure as before to recover the kernels and the baseline intensities. Finally, we indicate with L_i^x ($x \in \{a, b\}$, $i = 1, .., 4$) the limit order events at the ask ($x = a$) or at the bid ($x = b$) with volume in bin i. Similar notation is used for cancellations (C_i^x) and trades (T_i^x).

3.5.1 Kernel norms

We start our analysis by examining the matrix of the norms $\|\phi^{ij}\|_1$, that give us a summary of the main mutual influences detected in the order book with our procedure. Since we find the bid/ask symmetry to be fairly well respected, we plot only the quadrants Ask→Ask and Bid→Ask in order to improve readability.

We first examine the detected relationship between same side orders. In Table 3.7 we report the ask/ask quadrant for the Bund and Dax futures, whereas the rescaled kernel matrices are reported in Table 3.12. We stress once again that in the case of the DAX the four volume components have very different average intensities. As a result, the normalized norms matrix looks very different from the original norm matrix. We note that diagonal terms, corresponding to self excitation, play an important role in the dynamics. This was first observed in Biais et al. (1995) ("diagonal effect"). In particular, these authors find that the conditional probability than a certain type of order is observed after the same type of order just occurred is higher than its unconditional probability. Also in Bacry et al. (2016) diagonal terms (limit/limit, cancel/cancel, trade/trade) were found to be important. By adding order size to the analysis, we determine that this interaction is stronger for same size orders.

Bund												
	L_1^a	L_2^a	L_3^a	L_4^a	C_1^a	C_2^a	C_3^a	C_4^a	T_1^a	T_2^a	T_3^a	T_4^a
L_1^a	**0.25**	**0.10**	**0.11**	**0.17**	**0.13**	-0.02	-0.03	-0.02	**0.03**	0.02	0.02	-0.04
L_2^a	**0.14**	**0.24**	**0.18**	**0.26**	-0.04	**0.18**	-0.05	-0.03	-0.00	0.03	0.02	-0.02
L_3^a	**0.06**	**0.10**	**0.31**	**0.18**	-0.01	-0.02	**0.18**	0.03	0.02	0.05	**0.09**	**0.14**
L_4^a	0.02	0.03	0.04	**0.20**	-0.00	-0.01	0.00	0.11	0.00	0.01	0.02	0.09
C_1^a	**0.17**	-0.02	-0.01	-0.07	**0.25**	**0.14**	**0.13**	**0.15**	0.02	0.04	0.08	**0.29**
C_2^a	-0.05	**0.17**	-0.05	-0.07	**0.14**	**0.25**	**0.18**	**0.21**	0.00	-0.02	0.02	0.11
C_3^a	-0.03	-0.03	**0.22**	-0.02	**0.06**	**0.09**	**0.25**	**0.15**	-0.03	-0.00	0.00	**0.15**
C_4^a	-0.01	-0.01	-0.01	0.15	0.01	0.01	0.03	**0.24**	0.00	0.02	0.04	0.13
T_1^a	0.00	-0.00	0.00	0.01	0.00	0.01	0.00	-0.01	**0.41**	**0.10**	0.06	0.05
T_2^a	-0.00	0.00	0.00	0.00	0.00	0.01	0.00	-0.00	**0.06**	**0.36**	**0.08**	0.05
T_3^a	-0.00	-0.00	0.01	0.01	0.00	0.01	0.00	-0.01	**0.04**	**0.08**	**0.30**	0.10
T_4^a	0.00	-0.00	0.00	0.03	0.01	0.02	0.00	-0.01	0.03	**0.05**	**0.10**	**0.22**

DAX												
	L_1^a	L_2^a	L_3^a	L_4^a	C_1^a	C_2^a	C_3^a	C_4^a	T_1^a	T_2^a	T_3^a	T_4^a
L_1^a	**0.34**	**0.06**	-0.11	-0.16	**0.14**	**0.11**	**0.44**	**0.26**	**0.06**	**0.07**	**0.14**	**0.50**
L_2^a	0.01	**0.23**	**0.08**	0.04	-0.00	**0.13**	**0.04**	**0.04**	0.01	0.03	0.07	**0.22**
L_3^a	0.00	0.01	**0.24**	0.06	-0.00	-0.00	0.02	0.01	0.00	0.01	0.02	0.06
L_4^a	0.00	0.00	0.00	**0.13**	-0.00	-0.00	0.00	0.00	-0.00	0.00	-0.00	-0.00
C_1^a	**0.18**	-0.14	-0.28	-0.39	**0.36**	**0.25**	**0.52**	**0.69**	**0.21**	**0.51**	**0.37**	-1.63
C_2^a	0.00	**0.23**	0.05	-0.03	**0.01**	**0.16**	0.00	0.01	0.00	0.00	0.01	-0.02
C_3^a	**0.02**	0.02	**0.16**	-0.07	-0.00	0.02	**0.27**	-0.03	0.01	0.01	0.00	-0.04
C_4^a	0.00	0.01	0.03	**0.14**	**0.00**	0.01	-0.00	**0.26**	0.00	0.00	0.01	-0.02
T_1^a	**0.02**	0.02	**0.05**	**0.10**	-0.02	0.01	-0.01	-0.02	**0.31**	**0.23**	**0.29**	**0.68**
T_2^a	0.00	**0.03**	0.04	**0.10**	-0.01	-0.00	-0.01	-0.02	**0.06**	**0.15**	**0.12**	**0.22**
T_3^a	0.00	0.00	0.04	0.07	-0.00	-0.00	-0.00	0.00	0.03	0.05	0.11	0.13
T_4^a	-0.00	-0.00	-0.00	0.06	-0.00	-0.00	0.00	0.00	0.01	0.01	0.02	0.06

Table 3.7: Ask-Ask quadrant of the kernel norm matrices for the order book case. To highlight the structure, the largest four values in each column are in bold face.

Another feature, more visible for the Bund, concerns the blocks $L_i^x \to C_j^x$ and $C_i^x \to L_j^x$. They describe the influence of cancellations on limits on the same side of the book (and vice-versa), and present positive diagonal terms and slightly negative and mostly negligible off diagonal terms. This means that excitation is found only between orders of the same (or similar) size. The presence of strong excitation between orders of the same size appears very reasonable. Indeed, one can only cancel orders that he placed before. The fact that the same structure is found in both the blocks $L^x \to C^x$ and $C^x \to L^x$ suggests that these blocks reflect repositioning of limit orders, for example because the price has moved.

We already discussed in the previous section the structure of the trade-trade interaction. Here we note that trades appear to be mostly influenced by other trades with very little excitation from other types of orders. This is somewhat expected, since market orders have been found in different studies (Toke, 2015) to trigger limit orders. Trades, especially large ones appear to have also a significant effect on cancellations. We will discuss in more detail the effects of trades in Section 3.5.3.

Table 3.8 shows instead the quadrant ask/bid of the estimated kernel matrix. The rescaled version is instead reported in Table 3.13. First, we observe that trades, particularly large ones, influence significantly limit orders on the opposite side of the book. Values of the kernel norms $T_4^x \to L^y$ $(x \neq y)$ are in fact far higher than those for small trades $T_1^x \to L^y$, with $T_2^x \to L^y$ and $T_1^x \to L^y$ taking values in between. This is more evident in the case of the DAX. The DAX is a case of a small-tick asset, so large trades are more likely to move the midprice. As a consequence, new bid limit orders are submitted in order to follow the price move upwards and vice-versa.

Finally, we note the mutual influence between limit and cancel orders on opposite side of the book and the inhibitory effect $L_i^x \to L_j^y$ and $C_i^x \to C_j^y$ $(x \neq y)$, i.e. between the same order type on opposite side of the order book. This can be linked to a fair price view, in the sense that when the flow of limit orders is more intense on side x it means that the "fair price" is closer to side y, and thus the flow of limit orders on side y decreases.

The norm matrix provides a summary of the mutual influence structure, however, to complete the analysis it is important to look at the precise shape of the kernels which add noteworthy information. In the following, we will examine the shape of the kernels focusing in particular on the effect of the different order sizes.

3.5.2 Effects of limit orders and cancellations

To identify the main effect of the arrival of an event of type Z_j^x (e.g. limit, cancel, trade with size j), we plot all the kernels $\phi(Z_j^x \to \cdot)$ i.e. we plot along one column of the kernel matrix. To improve readability we separate the effect on the same side of the book from those on the opposite side.

In Figure 3.6 we plot the kernels that describe the influence of a small and respectively a large limit order on the DAX future. Let us look first at the left column of the figure, that shows the influence on the same side of the book. We notice immediately that both a small and a large limit order trigger mainly small limit orders. However, while the small/small kernel attains high values already at very short time scales (below 300μs), the large/small one is negligible until that scale. This is a feature we already noticed in Section 3.4 and suggests that when a large order arrives, what follows is mainly the reaction of the market to the new information, and it takes some time to react. While the reaction part is present also for small orders, here a significant part seems to be the result of the same trader executing a series of orders at very short distances.

Small limit orders also trigger cancellations of the same size on both sides of the book. This may be the result of traders repositioning their limit orders because e.g. the midprice

Bund												
	L_1^b	L_2^b	L_3^b	L_4^b	C_1^b	C_2^b	C_3^b	C_4^b	T_1^b	T_2^b	T_3^b	T_4^b
L_1^a	0.01	0.01	-0.02	-0.06	**0.04**	**0.00**	**0.02**	**0.08**	**0.07**	0.02	**0.05**	**0.35**
L_2^a	-0.02	-0.02	-0.04	-0.09	**0.02**	**0.09**	**0.04**	**0.15**	**0.01**	**0.07**	**0.06**	**0.59**
L_3^a	-0.02	-0.02	-0.03	-0.07	**0.03**	**0.02**	**0.05**	**0.06**	**0.02**	**0.08**	**0.22**	**0.68**
L_4^a	-0.00	-0.00	-0.01	-0.01	**0.01**	**0.02**	**0.01**	**0.03**	**0.02**	**0.03**	**0.07**	**0.45**
C_1^a	**0.04**	**0.03**	**0.02**	**0.10**	0.00	-0.02	-0.01	-0.03	-0.01	-0.02	-0.03	-0.07
C_2^a	**0.04**	**0.11**	**0.04**	**0.11**	-0.01	-0.02	-0.02	-0.03	-0.05	-0.04	-0.05	-0.10
C_3^a	**0.03**	**0.06**	**0.07**	**0.14**	-0.01	-0.02	-0.03	-0.02	-0.04	-0.03	-0.01	-0.10
C_4^a	**0.02**	**0.03**	**0.03**	**0.08**	-0.00	-0.01	-0.01	0.01	-0.00	0.01	0.03	0.02
T_1^a	0.00	0.00	-0.00	-0.01	0.00	-0.00	-0.00	0.01	0.01	0.00	0.01	0.06
T_2^a	-0.00	0.00	-0.00	-0.01	0.00	-0.00	-0.00	0.00	0.00	**0.02**	0.02	0.07
T_3^a	-0.00	0.00	-0.00	-0.01	0.00	-0.01	0.00	0.00	0.00	0.00	0.02	0.09
T_4^a	-0.00	-0.00	-0.00	-0.00	0.00	-0.00	0.00	0.00	-0.00	-0.01	-0.01	0.05

DAX												
	L_1^b	L_2^b	L_3^b	L_4^b	C_1^b	C_2^b	C_3^b	C_4^b	T_1^b	T_2^b	T_3^b	T_4^b
L_1^a	-0.07	-0.20	-0.35	-0.80	**0.13**	**0.23**	**0.28**	**0.26**	**0.34**	**0.54**	**1.12**	**2.15**
L_2^a	-0.01	0.00	-0.03	-0.02	**0.00**	**0.05**	**0.03**	**0.09**	**0.06**	**0.18**	**0.32**	**0.61**
L_3^a	0.00	-0.00	0.00	0.02	-0.00	0.00	**0.01**	**0.01**	0.01	0.03	**0.08**	**0.10**
L_4^a	0.00	-0.00	-0.00	0.00	-0.00	0.00	0.00	**0.00**	0.00	0.00	0.00	0.01
C_1^a	**0.22**	**0.40**	**0.67**	**1.31**	-0.02	-0.11	-0.11	-0.14	-0.10	-0.09	-0.25	-0.96
C_2^a	**0.01**	**0.03**	**0.04**	0.08	-0.01	**0.01**	-0.00	-0.00	0.01	0.03	0.01	-0.02
C_3^a	0.01	0.01	0.02	-0.00	-0.00	**0.01**	**0.01**	-0.01	**0.02**	**0.04**	0.02	-0.07
C_4^a	0.00	0.00	0.00	0.00	**0.00**	0.00	-0.00	0.00	0.00	0.01	0.01	-0.01
T_1^a	**0.03**	**0.06**	**0.10**	**0.26**	-0.01	0.00	-0.02	-0.02	**0.04**	**0.05**	**0.09**	**0.29**
T_2^a	**0.01**	**0.05**	**0.08**	**0.23**	-0.01	-0.01	-0.01	-0.01	0.01	0.02	0.03	0.09
T_3^a	0.00	0.01	0.04	**0.11**	-0.00	-0.00	-0.00	-0.00	0.00	0.00	0.01	0.03
T_4^a	0.00	0.00	0.00	0.01	**0.00**	-0.00	-0.00	-0.00	0.00	-0.00	0.00	0.01

Table 3.8: Ask-Bid quadrant of the kernel norm matrices for the order book case. To highlight the structure, the largest four values in each column are in bold face.

has shifted. Note that the $L_1^a \to C_1^b$ kernel features high values also before the 300μs reaction time, unlike the $L_1^a \to C_1^a$ kernel. This is probably due to the fact that if a trader decides to change side on the book then he will almost simultaneously place a new limit order on one side and cancel his previous one on the other.

A large limit order also triggers many cancellations and, to a lesser extent, trades on the opposite side. A possible reading is the following. A large limit order on ask side conveys the information that the "fair price" is actually closer to the bid side. Then agents rush to cancel their outstanding limit order at the bid, but some get caught by fast traders that place market orders at the bid to take advantage of the situation.

Effect of cancellations, reported in Figure 3.7 follow the same lines, albeit the effect on trades is negligible. Similar considerations hold also in the Bund case, the main difference being that in the DAX case the occurrence of a large order (in our binning scheme) is a much rare event and thus has more dramatic consequences.

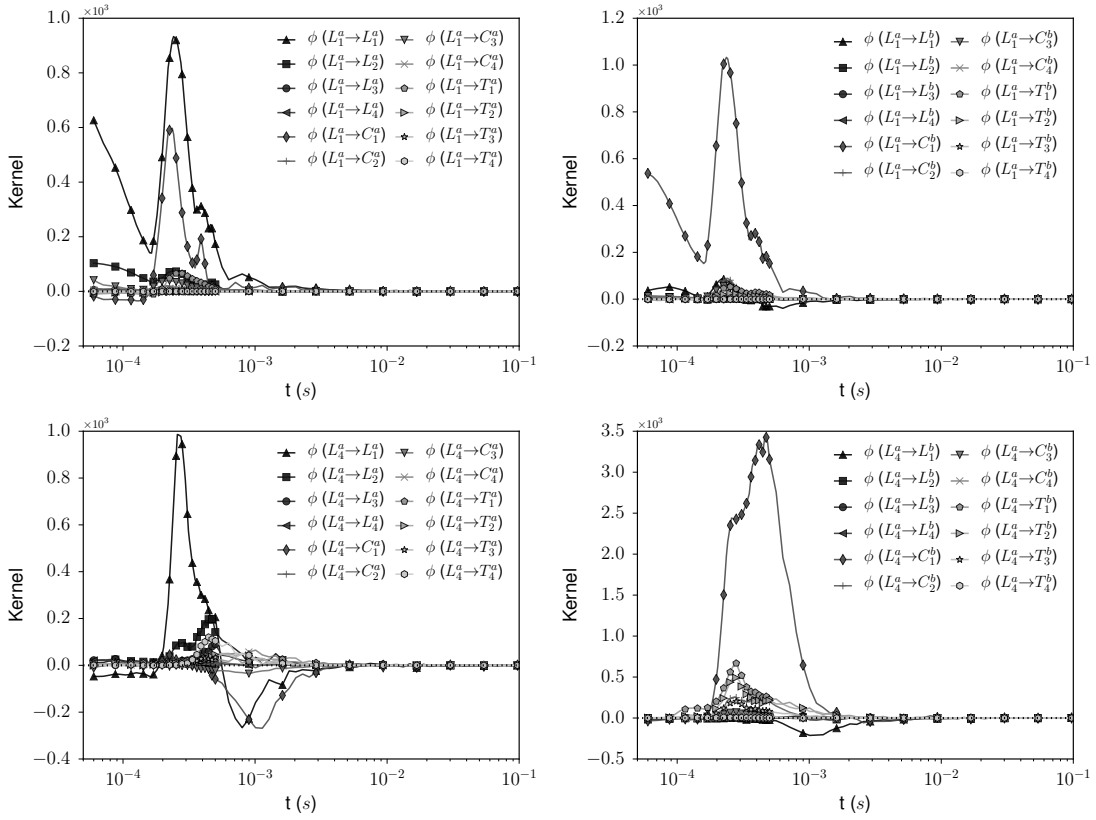

Figure 3.6: DAX future: kernels describing the influence of a small (top) and large (bottom) limit order on other events. The left column refers to same side of the book, the right side to the opposite one.

3.5.3 Effects of Trades

We now look at the effect of trades on other events. In Figures 3.8 and 3.9 the relative kernels are plotted for the DAX and the Bund futures, respectively. As we already pointed out, below the market reaction time we note the effect of order splitting, that manifests itself in the large positive values of the small-trade/small-trade kernel. In the case of the DAX this effect is absent for large trades, since there is no point in executing two or more large orders one immediately after the other. In the Bund case, where a 10 contracts order

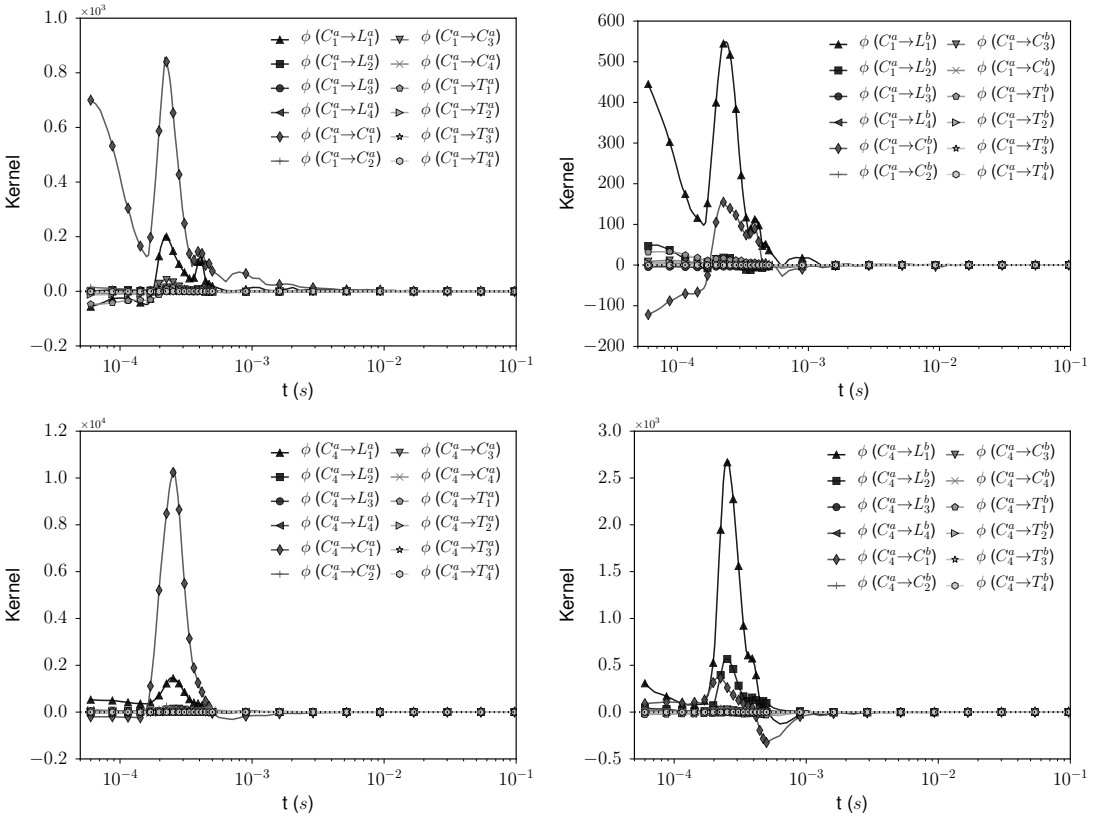

Figure 3.7: DAX future: kernels describing the influence of a small (top) and a large (bottom) cancel order on other events. The left column refers to same side of the book, the right side to the opposite one.

is still not so large we note some excitation from large trades to small ones at the shortest time scale.

Another feature that emerges concerns the kernels $T^a \to C_1^a$. For both assets, these kernels are negative before the market reaction time, then they assume very large positive values around 300μs and then around $1m$s they revert to zero or to negative values in the case of large trades. We can give the following reading of this effect. When a large trade arrives, it will likely eat up a significant part of the outstanding liquidity. Of course, those limit orders that have been matched cannot be canceled anymore. Thus, the negative values of this kernel before the market reaction time are explained by this mechanical effect. The reaction of market participants with a market making strategy to a large trade is then to cancel their outstanding limit orders on the side hit by the trade, and to place limit orders on the opposite side. This is a consequence of the fact that a trade at the ask signals that the "true price" is closer to that side, and thus liquidity must adapt to the new situation. We remark again that the effect of large trades on liquidity is both more intense (compare the different y-scales of the figures) and much more persistent in time.

3.5.4 Baseline intensities

We complete the analysis of the order book section by examining the exogeneity of the different types and sizes of orders, measured by the baseline intensities μ_i, we present the ratios μ_i/Λ_i in Table 3.9.

We can make three kind of considerations. The first concerns the degree of exogeneity

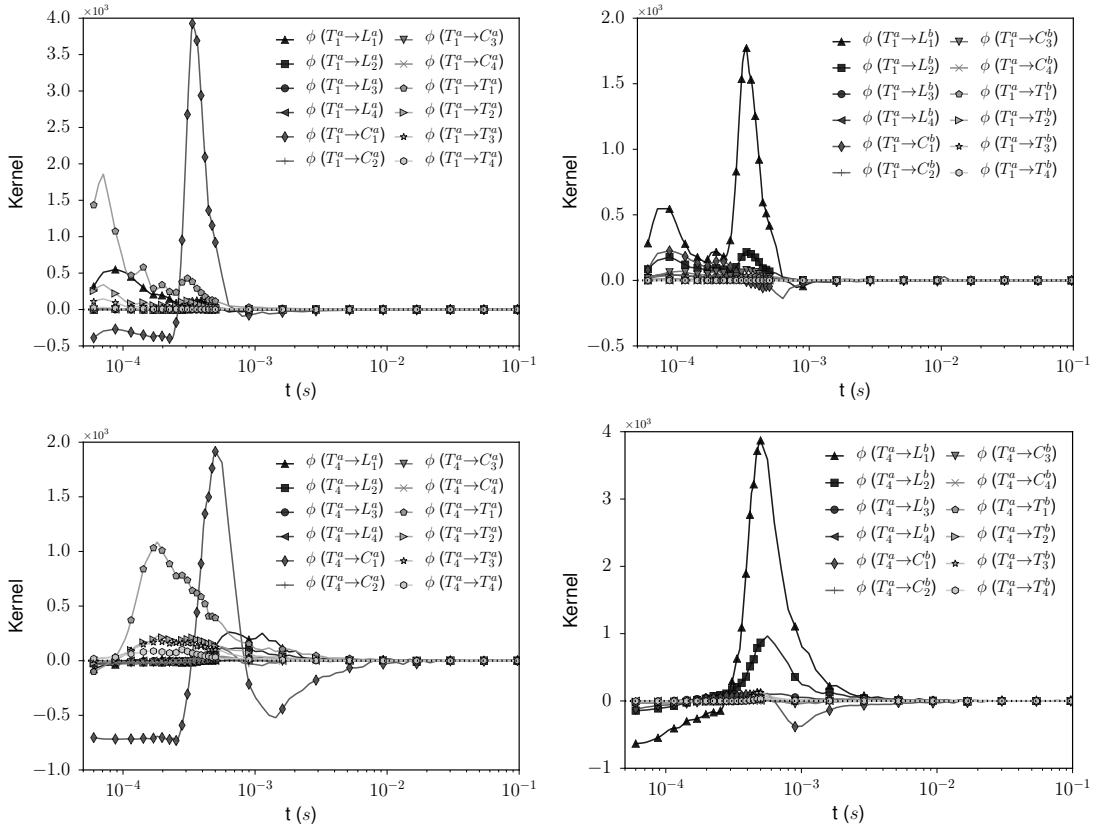

Figure 3.8: DAX future: kernels describing the influence of a small (top) and a large (bottom) trade on other events. The left column refers to same side of the book, the right side to the opposite one.

of the different types of orders. Trades are the most exogenous, followed by limit orders. It is no surprise that the cancellations activity can be in large part explained by cross and self excitation, since cancel orders are mostly the reaction to some event. It is interesting to compare the ratios μ_i/Λ_i we obtained when examining trades alone with those of the full order book model. From Table 3.10 we note that, when including all the other orders into the picture the fraction of intensity explained by the baseline intensity decreases, albeit not dramatically. In fact, market orders have been shown to lead limit and cancels rather than be driven. Moreover, we have seen that most of the cross excitation comes from other trades, thus no big difference is expected between the full order book case and the trade-only case. Nevertheless, some trade activity is explained by the other events notably by large limit orders as we pointed out in the previous section.

The second consideration regards the effect of the order size. In line with the analysis conduced so far, we find large orders to be more exogenous than small ones. It is reasonable to assume that when a trader places a large order is because she has an "independent" reason to do so, either a clear view on the underling asset or a necessity to execute a large quantity immediately, rather than because she is following the market or engaging in market making.

Finally, we observe that the ratios μ_i/Λ_i for large orders on the DAX are much higher than those on the Bund (see Tables 3.9 and 3.10). This is mostly a consequence of the different volume distributions in the two assets. On the DAX future, orders larger than 10 are much more unusual, and perceived as big, than they are on the Bund.

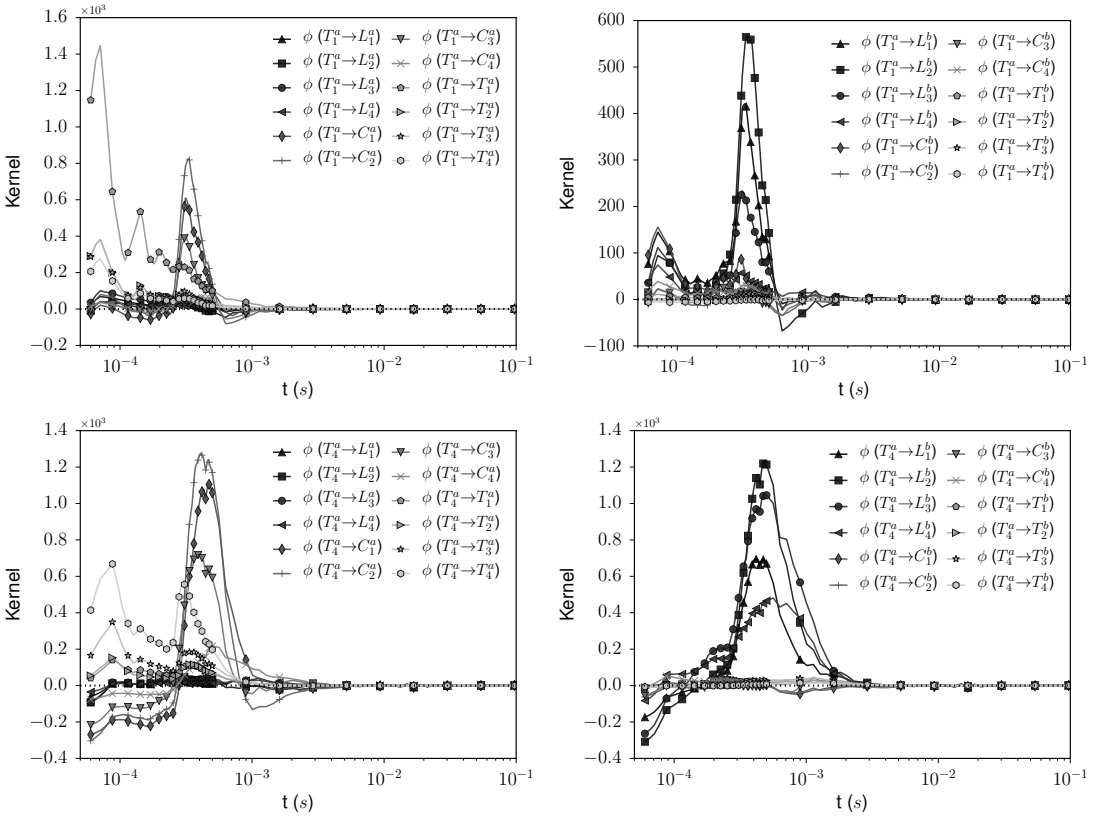

Figure 3.9: Bund future: kernels describing the influence of a small (top) and a large (bottom) trade on other events. The left column refers to same side of the book, the right side to the opposite one.

Bund											
L_1^a	L_2^a	L_3^a	L_4^a	C_1^a	C_2^a	C_3^a	C_4^a	T_1^a	T_2^a	T_3^a	T_4^a
20.6	16.8	13.0	16.5	9.0	16.5	13.3	7.8	27.3	16.8	17.4	36.8
L_1^b	L_2^b	L_3^b	L_4^b	C_1^b	C_2^b	C_3^b	C_4^b	T_1^b	T_2^b	T_3^b	T_4^b
19.9	16.5	13.1	17.1	8.0	16.5	13.3	7.6	29.2	17.0	18.0	36.4
DAX											
L_1^a	L_2^a	L_3^a	L_4^a	C_1^a	C_2^a	C_3^a	C_4^a	T_1^a	T_2^a	T_3^a	T_4^a
19.9	19.1	33.7	72.3	10.7	10.8	16.4	18.5	27.2	30.5	44.8	50.6
L_1^b	L_2^b	L_3^b	L_4^b	C_1^b	C_2^b	C_3^b	C_4^b	T_1^b	T_2^b	T_3^b	T_4^b
20.1	20.0	35.0	71.2	11.0	11.5	16.4	18.4	26.7	30.5	44.8	49.6

Table 3.9: Ratios $\frac{\mu_i}{\Lambda_i}$ expressed in percent.

Bund							
T_1^a	T_2^a	T_3^a	T_4^a	T_1^b	T_2^b	T_3^b	T_4^b
4.3	4.3	6.6	10.7	2.4	3.6	5.6	10.4
DAX							
T_1^a	T_2^a	T_3^a	T_4^a	T_1^b	T_2^b	T_3^b	T_4^b
5.5	7.3	4.0	7.5	5.7	6.8	2.8	6.6

Table 3.10: Difference in the ratios $\frac{\mu_i}{\Lambda_i}$ (expressed in percent) between the signed trade only case examined in Section 3.4 and the full order book case.

3.6 Conclusions

In this paper we showed the effectiveness of multivariate Hawkes processes coupled with non-parametric kernel estimation as a toolkit for the analysis of high frequency order book dynamics when the effect of some variable besides the time and type of orders is of interest. We examined in particular the role played by the order size and we have shown that a simple multiplicative model, where time and size are factored, is not able to capture the complex interplay between these variables. On the contrary, our approach has proved to be capable of highlighting several features of the dynamics, from typical reaction time to complex interaction between different order types. Of course, while the methodology is completely general, the features we highlighted pertain to the particular market and instrument that we examined and may in principle be different in other markets.

There are two main strengths of our approach as compared to the existing ones. First, Hawkes processes are point processes easily applicable in multiple dimensions and in this paper we have exploited this property by considering different types of orders (limit orders, cancellations, and trades) and different volume size. Second, we used a non parametric estimation method that limits the constraint imposed by traditional parametric models.

When volume is not considered we recovered many of the results of Bacry et al. (2016) for limit order book dynamics. Moreover by using the multivariate approach we were able to separate contribution from different order sizes. The different impact of large orders clearly emerge from our analysis, as well as the longer persistence of their effect. Our work thus shows that the role of order size needs to be taken into account for a complete understanding of the order book dynamics. Our study still leave out some relevant information such as the size of the queue and the dynamics of the price and this will be the objective of future works.

We conclude by noting that our methodology is well suited to the application to different systems where the interaction of different event types as well as their marks are relevant.

Appendix

3.A Proofs

Here we report the proof of Proposition 3.2.1.

Proof. Let us suppose that $\phi(t) \geq 0$ for all $t \geq 0$, meaning that all the elements are positive functions. Define, for every $t \geq 0$, $S^-_{kj} = \{s \in \mathbb{R}^+ : g^{kj}(t-s) < 0\} \neq \emptyset$ and $S^+_{kj} = \{s \in \mathbb{R}^+ : g^{kj}(t-s) \geq 0\}$.

Then we rewrite the integral equation (2.19) as:

$$\begin{aligned} g^{ij}(t) &= \phi^{ij}(t) + \sum_k \int_0^\infty g^{kj}(t-s)\phi^{ik}(s)\mathrm{d}s \\ &= \phi^{ij}(t) + \sum_k \int_{S^+_{kj}} g^{kj}(t-s)\phi^{ik}(s)\mathrm{d}s + \sum_k \int_{S^-_{kj}} g^{kj}(t-s)\phi^{ik}(s)\mathrm{d}s \end{aligned} \tag{3.8}$$

Rearranging the terms we get:

$$g^{ij}(t) - \sum_k \int_{S^-_{kj}} g^{kj}(t-s)\phi^{ik}(s)\mathrm{d}s = \phi^{ij}(t) + \sum_k \int_{S^+_{kj}} g^{kj}(t-s)\phi^{ik}(s)\mathrm{d}s \tag{3.9}$$

The right hand side is always positive, so for the equation to hold we need, for all t such that $g^{ij}(t) < 0$:

$$g^{ij}(t) \geq \sum_k \int_{S^-_{kj}} g^{kj}(t-s)\phi^{ik}(s)\mathrm{d}s \tag{3.10}$$

Let us define

$$m_{kj} = \inf_{\mathbb{R}^+} g^{kj}(t) \tag{3.11}$$

and

$$m_j = \min_k m_{kj} \tag{3.12}$$

Now we can write

$$g^{ij}(t) \geq \sum_k \int_{S^-_{kj}} g^{kj}(t-s)\phi^{ik}(s)\mathrm{d}s \geq \sum_k m_{kj} \int_{S^-_{kj}} \phi^{ik}(s)\mathrm{d}s \geq m_j \sum_k \int_{S^-_{kj}} \phi^{ik}(s)\mathrm{d}s \tag{3.13}$$

if we take this equation for the component (i^*, j) such that $m_j = m_{i^*j}$, that is the g that reaches the lowest value on the column j, we find that the above conditions requires

$$\sum_k \int_{S^-_{kj}} \phi^{i^*k}(s) \geq 1 \Rightarrow \sum_k \int_0^\infty \phi^{i^*k}(s) = \sum_k \|\phi^{i^*k}\|_1 \geq 1 \tag{3.14}$$

since the ϕ are positive and causal functions. If the matrix $\mathbf{g}(t)$ has at least one entry on every column (or row) wich takes negative values, then the minimum row (column) sum of $\hat{\phi}_0$, r (c) must be greater than one. This implies that also the spectral radius, ρ, of $\hat{\phi}_0$ is greater than one (see Minc (1988) pp. 24-25).

□

3.B Robustness checks

In this appendix we briefly discuss two possible issues that can arise when dealing with real data, namely the effects of finite time resolution and those of the intraday activity pattern. We find that none of these issues appears to significantly influence the results of our analysis.

Timestamps randomization As noted in the main text, in our data we never observe two trades closer in time than 50 μs. This could be the result of some technical time required to process the order and write the data in the database. For changes in the best quote the minimum time difference observed is 10 μs, despite the timestamps being recorded with microsecond precision. So there is the possibility that actual events are sometimes closer in time to each other than they result from the database.

We check the outcome of the non-parametric estimation also with randomized data. To randomize the data, we first round all the trades timestamps to the nearest $10\mu s$. Then, we subtract to them a random number uniformly distributed in $[0, 50)\,\mu s$. The effect of randomization on short durations are shown in Figure 3.10.

The estimation parameters are the same as those used in Section 3.4. In Table 3.11 (top) the rescaled kernels norms are presented, we note that the outcome is almost identical to the non randomized case (see Table. 3.2). We find also that the shape of the obtained kernels is essentially the same, except that the peaks at 100 and $300\mu s$ are broader as a result of the broader distribution of the inter event times due to randomization. We conclude that the time resolution of our data is fine enough not to pose noticeable problem to the estimation procedure and we therefore prefer to use the original timestamps.

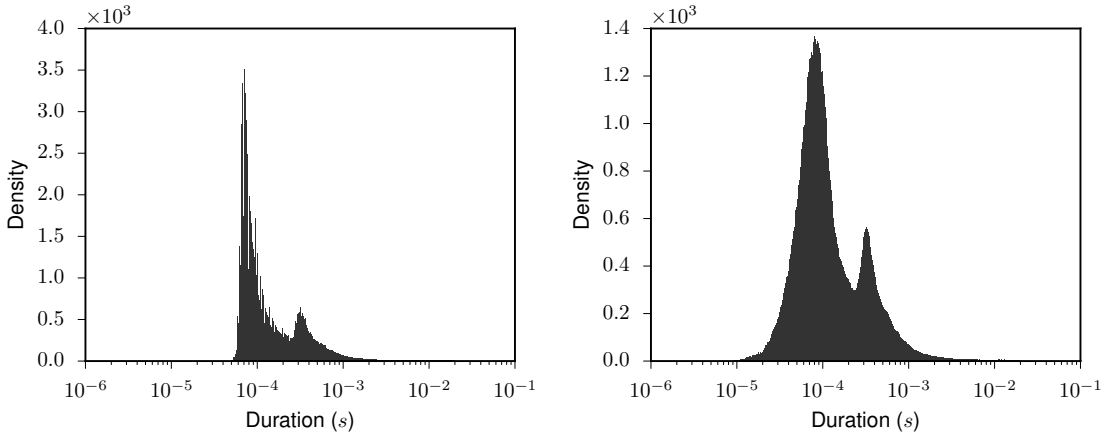

Figure 3.10: Bund future: Inter events distribution before (left) and after (right) randomization.

Intraday Seasonality In principle intraday non stationarities could pose a problem for estimation. To check for daily-pattern effects on the estimation procedure we repeat the estimation on trade events comprised between 11:00 and 17:00 Frankfurt time. We consider the same six bins of volume as specified in Table 3.1. The matrix of the norms is reproduced in Table 3.11 (bottom). We also computed the relative differences

$$\frac{n_{ij} - n'_{ij}}{n_{ij}}$$

Where, n_{ij} is the rescaled norm of ϕ_{ij} estimated on the whole period (08:00-22:00), while n' is the analogous quantity estimated on the restricted period. We find that the differences are of the order of a few percent, with the highest differences (about 10%) appearing on the smallest norms where we also have the highest estimation error. We therefore conclude that our main results are robust also with respect to seasonality effects. This can be linked to the fact that our study focuses only on the very short time scale dynamics (we estimate the kernels only up to $0.5s$).

3.C Rescaled kernel matrices for the first level of the order book

Bund Randomized						
	≤ 1	≤ 2	≤ 3	≤ 7	≤ 20	> 20
≤ 1	**0.52**	**0.05**	0.02	0.04	0.04	**0.09**
≤ 2	**0.10**	**0.40**	0.04	0.08	0.08	**0.13**
≤ 3	0.08	0.08	**0.24**	**0.14**	0.11	**0.15**
≤ 7	0.06	0.06	0.05	**0.32**	**0.13**	**0.18**
≤ 20	0.04	0.04	0.02	**0.09**	**0.29**	**0.22**
> 20	0.03	0.02	0.01	**0.04**	**0.09**	**0.24**

Bund 11-17						
	≤ 1	≤ 2	≤ 3	≤ 7	≤ 20	> 20
≤ 1	**0.52**	0.04	0.01	0.03	**0.04**	**0.09**
≤ 2	**0.08**	**0.39**	0.04	0.07	0.08	**0.14**
≤ 3	0.07	0.08	**0.24**	**0.14**	0.11	**0.15**
≤ 7	0.06	0.06	0.05	**0.31**	**0.14**	**0.19**
≤ 20	0.04	0.04	0.03	**0.09**	**0.28**	**0.23**
> 20	0.03	0.02	0.01	**0.04**	**0.09**	**0.24**

Table 3.11: Bund future: rescaled kernel norms for the six dimensional case. Results from randomized data (top), and results form trades occurred only between 11:00 and 17:00 local time (bottom). The three largest values in each row are highlighted in bold face.

Bund

	L_1^a	L_2^a	L_3^a	L_4^a	C_1^a	C_2^a	C_3^a	C_4^a	T_1^a	T_2^a	T_3^a	T_4^a
L_1^a	**0.25**	**0.14**	**0.12**	0.07	**0.12**	-0.02	-0.03	-0.01	0.00	0.00	0.00	-0.01
L_2^a	**0.10**	**0.24**	**0.14**	0.08	-0.03	**0.16**	-0.03	-0.01	-0.00	0.00	0.00	-0.00
L_3^a	0.06	**0.12**	**0.31**	**0.07**	-0.01	-0.02	**0.15**	0.01	0.00	0.00	0.01	0.02
L_4^a	0.05	**0.10**	**0.11**	**0.20**	-0.01	-0.02	0.00	**0.09**	0.00	0.00	0.01	0.03
C_1^a	**0.18**	-0.03	-0.01	-0.03	**0.25**	**0.17**	**0.13**	0.05	0.00	0.00	0.01	0.04
C_2^a	-0.04	**0.19**	-0.05	-0.02	**0.11**	**0.25**	**0.14**	0.06	0.00	-0.00	0.00	0.01
C_3^a	-0.04	-0.05	**0.27**	-0.01	**0.06**	**0.12**	**0.25**	0.05	-0.01	-0.00	0.00	0.02
C_4^a	-0.04	-0.03	-0.02	**0.18**	0.03	0.05	**0.08**	**0.24**	0.00	0.01	0.01	**0.05**
T_1^a	0.02	-0.00	0.01	0.02	0.01	**0.04**	0.01	-0.01	**0.41**	**0.07**	**0.04**	0.04
T_2^a	-0.01	0.03	0.01	0.02	0.01	**0.07**	0.02	-0.01	**0.09**	**0.36**	**0.09**	0.06
T_3^a	-0.01	-0.02	**0.09**	0.04	0.03	**0.10**	0.00	-0.03	0.05	0.07	**0.30**	**0.10**
T_4^a	0.00	-0.03	0.02	**0.09**	0.04	**0.16**	0.02	-0.03	0.03	0.04	**0.09**	**0.22**

DAX

	L_1^a	L_2^a	L_3^a	L_4^a	C_1^a	C_2^a	C_3^a	C_4^a	T_1^a	T_2^a	T_3^a	T_4^a
L_1^a	**0.34**	0.01	-0.00	-0.00	**0.14**	0.01	**0.04**	0.01	**0.01**	0.00	0.00	0.00
L_2^a	**0.10**	**0.23**	0.02	0.00	-0.01	**0.08**	**0.02**	0.01	0.01	0.01	0.01	0.01
L_3^a	**0.04**	**0.06**	**0.24**	0.01	-0.03	-0.01	**0.07**	0.01	0.03	0.02	0.02	0.01
L_4^a	**0.03**	0.00	**0.01**	**0.13**	-0.05	-0.01	0.00	**0.02**	-0.00	0.01	-0.00	-0.00
C_1^a	**0.17**	-0.02	-0.01	-0.00	**0.36**	0.02	**0.04**	0.02	**0.04**	0.03	0.01	-0.01
C_2^a	0.00	**0.37**	**0.02**	-0.00	**0.09**	**0.16**	0.00	0.00	0.00	0.00	0.00	-0.00
C_3^a	**0.26**	**0.04**	**0.05**	-0.00	-0.04	0.02	**0.27**	-0.01	0.02	0.01	0.00	-0.00
C_4^a	**0.16**	**0.06**	0.03	0.02	**0.07**	0.03	-0.01	**0.26**	0.02	0.01	0.01	-0.00
T_1^a	**0.12**	0.01	0.01	0.00	-0.13	0.01	-0.00	-0.00	**0.31**	**0.08**	**0.05**	0.02
T_2^a	0.04	**0.07**	0.02	0.00	-0.12	-0.00	-0.01	-0.01	**0.16**	**0.15**	**0.06**	0.02
T_3^a	0.01	0.01	**0.04**	0.01	-0.12	-0.01	-0.01	0.00	**0.17**	**0.10**	**0.11**	0.02
T_4^a	-0.00	-0.03	-0.01	0.04	-0.21	-0.00	0.01	0.03	**0.27**	**0.15**	**0.15**	**0.06**

Table 3.12: Ask-Ask quadrant of the rescaled kernel norm matrices for the order book case. To highlight the structure, the largest four values in each row are in bold face.

Bund												
	L_1^b	L_2^b	L_3^b	L_4^b	C_1^b	C_2^b	C_3^b	C_4^b	T_1^b	T_2^b	T_3^b	T_4^b
L_1^a	0.01	0.01	-0.02	-0.02	**0.04**	0.00	**0.01**	**0.02**	0.01	0.00	0.01	**0.05**
L_2^a	-0.02	-0.02	-0.03	-0.03	0.01	**0.08**	**0.03**	**0.03**	0.00	0.01	0.01	**0.06**
L_3^a	-0.02	-0.02	-0.03	-0.03	**0.03**	0.02	**0.04**	0.02	0.00	0.01	**0.03**	**0.08**
L_4^a	-0.01	-0.02	-0.02	-0.01	0.02	**0.07**	0.02	**0.02**	0.01	0.01	**0.02**	**0.15**
C_1^a	**0.05**	**0.04**	**0.02**	**0.04**	0.00	-0.02	-0.01	-0.01	-0.00	-0.00	-0.00	-0.01
C_2^a	**0.03**	**0.13**	**0.04**	**0.04**	-0.01	-0.02	-0.01	-0.01	-0.01	-0.00	-0.01	-0.01
C_3^a	**0.03**	**0.09**	**0.09**	**0.06**	-0.01	-0.03	-0.03	-0.01	-0.01	-0.00	-0.00	-0.02
C_4^a	**0.06**	**0.10**	**0.10**	**0.10**	-0.01	-0.03	-0.03	0.01	-0.00	0.00	0.01	0.01
T_1^a	0.01	0.00	-0.02	-0.02	**0.02**	-0.03	-0.01	**0.01**	**0.01**	0.00	0.01	**0.05**
T_2^a	-0.01	**0.02**	-0.03	-0.03	0.01	-0.02	-0.00	0.01	0.00	**0.02**	**0.02**	**0.08**
T_3^a	-0.02	0.01	-0.01	-0.02	0.01	-0.05	**0.01**	**0.01**	0.00	0.00	**0.02**	**0.10**
T_4^a	-0.03	-0.01	-0.04	-0.01	**0.01**	-0.02	**0.00**	**0.01**	-0.00	-0.01	-0.01	**0.05**

DAX												
	L_1^b	L_2^b	L_3^b	L_4^b	C_1^b	C_2^b	C_3^b	C_4^b	T_1^b	T_2^b	T_3^b	T_4^b
L_1^a	-0.07	-0.03	-0.01	-0.00	**0.14**	0.02	0.02	0.01	**0.06**	**0.03**	**0.03**	0.01
L_2^a	-0.04	0.00	-0.01	-0.00	0.00	**0.03**	0.02	0.02	**0.07**	**0.08**	**0.07**	0.02
L_3^a	**0.04**	-0.01	0.00	0.00	-0.14	0.01	0.02	0.01	**0.07**	**0.08**	**0.08**	0.02
L_4^a	**0.07**	-0.01	-0.00	0.00	-0.10	0.00	**0.02**	0.01	**0.05**	0.02	**0.03**	0.02
C_1^a	**0.21**	**0.05**	**0.02**	**0.00**	-0.02	-0.01	-0.01	-0.00	-0.02	-0.01	-0.01	-0.00
C_2^a	**0.17**	**0.05**	0.01	0.00	-0.08	0.01	-0.00	-0.00	**0.02**	**0.02**	0.00	-0.00
C_3^a	**0.12**	**0.01**	0.01	-0.00	-0.02	0.01	0.01	-0.00	**0.04**	**0.03**	0.01	-0.00
C_4^a	**0.09**	**0.01**	0.01	0.00	**0.04**	0.01	-0.01	0.00	0.00	**0.02**	0.01	-0.00
T_1^a	**0.17**	**0.04**	0.02	0.00	-0.06	0.00	-0.01	-0.00	**0.04**	**0.02**	0.01	0.01
T_2^a	**0.22**	**0.10**	**0.04**	0.01	-0.09	-0.01	-0.02	-0.01	**0.03**	0.02	0.01	0.01
T_3^a	**0.13**	**0.06**	**0.04**	0.01	-0.05	-0.01	-0.01	-0.00	**0.01**	0.00	0.01	0.01
T_4^a	**0.01**	0.00	0.01	**0.01**	**0.03**	-0.01	-0.01	-0.00	0.00	-0.01	0.00	**0.01**

Table 3.13: Ask-Bid quadrant of the rescaled kernel norm matrices for the order book case. To highlight the structure, the largest four values in each row are in bold face.

Chapter 4

Hawkes processes with local non-stationarities (I)

4.1 Intoduction

Complex system dynamics is characterized by a subtle interplay between endogenous and exogenous effects. Financial markets are, in this respect, paradigmatic. The dynamics of asset prices, in fact, is affected by the arrival of exogenous news, which modifies the valuation of the fair price of the asset, and by the arrival of endogenously generated events (e.g. trades, orders, and price changes) in reaction to past events (Engle and Ng, 1993; Birz and Lott, 2011; Groß-Klußmann and Hautsch, 2011). Disentangling and quantifying the relative importance of these two drivers have been debated widely in the last half century, since the formalization of the Efficient Market Hypothesis. There is a growing empirical evidence that a significant fraction of market activity[1] and volatility is explained by the endogenously generated trading activity. For example, large price movements (jumps) are only partly explained by public news (Cutler et al., 1989; Joulin et al., 2008).

The debate about the endogenous and exogenous component of price dynamics has recently received a new impulse thanks to the application of Hawkes processes to the modeling of financial data (Hewlett, 2006; Bowsher, 2007; Bauwens and Hautsch, 2009; Embrechts et al., 2011; Toke, 2011; Bacry et al., 2013, 2012; Bormetti et al., 2015). Hawkes processes Hawkes (1971), originally introduced in seismology, describe a point (counting) process where the intensity is not constant but depends on the past history of the counting process, weighted by a suitably chosen kernel. Due to the natural interpretation of a Hawkes process as a branching process, in terms of immigrant baseline events and offspring events, it is direct to interpret the integral of the kernel as the fraction of the activity due to the endogenous self excitation. The empirical analysis on equity and future data has indicated (Filimonov and Sornette (2012); Hardiman et al. (2013); Filimonov and Sornette (2015); Hardiman and Bouchaud (2014)) that, according to Hawkes modeling, a very large fraction of the market activity is explained by the endogenous component. Markets seem to be very close to a "critical" state where all the dynamics is endogenously generated (see Filimonov and Sornette (2012); Hardiman et al. (2013) for more details).

However, in some situations the arrival of exogenous news has typically a very large impact on the trading activity. An example, which is the one investigated in the present paper, is the impact of macroeconomic news on foreign exchange (FX) markets. This

[1] In this paper we will use the term *market activity* to describe the rate of price changes per unit time, irrespectively on the size, similarly to what investigated in Filimonov and Sornette (2012); Hardiman et al. (2013)

large impact is due to the fact that macroeconomic news (unemployment rate, GDP data, central banks intervention, etc.) affects significantly the valuation of the economy of a country and therefore of its currency. Moreover these announcements are scheduled in advance and are monitored with great attention by FX market participants.

The impact of exogenous news on asset prices has been investigated by a number of authors starting with the pioneering work of Cutler et al. (1989). In Mitchell and Mulherin (1994) authors analysed the connection between the daily number of news announcements reported by Dow Jones & Company and aggregate measures of market activity such as trading volume and market returns. Avellaneda and Lipkin (2003) studied the phenomenon of stock pinning observed at the expiration date of the corresponding options. Joulin et al. (2008) and Groß-Klußmann and Hautsch (2011) explored the role of news in high-frequency volatility and order book dynamics. A more extensive survey of the literature on news impact can be found in Lillo et al. (2015).

The effects of scheduled macroeconomic announcements on the FX market have been the object of several works. Ederington and Lee (1993) found that this sort of announcements is responsible for a large part of the volatility pattern observed in these markets. Andersen and Bollerslev (1998) provided a characterization of the volatility in the DM/USD rate, separating contributions of intraday activity patterns, macroeconomic announcements, and volatility persistence. More recently, Bauwens et al. (Bauwens et al., 2005) have a investigated volatility dynamics around news announcements on high-frequency EUR/USD data in the framework of ARCH-type models. These works focus either on large time horizons or choose a discrete time setting.

In this paper we focus on modeling the dynamic evolution in physical (continuous) time of market activity before and after a macroeconomic news announcement. We consider the rate of change in the best quotes as a proxy of market activity, and we take a point process perspective. Our main original contribution is a Hawkes model of the FX market activity where one of the kernels describes the (causal and possibly non-causal) effect of the macro news on the market activity. The model has thus two Hawkes kernels, one for the endogenous and one for the exogenous component. We find that the description of the market activity improves significantly when one introduces the exogenous kernel. It is important to stress that most of the work done so far concerning Hawkes processes in financial data has considered stationary processes. In fact, when non stationarities are present (e.g. the intraday pattern), one typically tries to remove them, for example by deseasonalizing the data. However in some cases, non stationarities are important and bring relevant economic information that one wants to model. The example considered in this paper is the market activity around macro announcements. Our approach demonstrates how it is possible to reconcile non stationarities like news announcements in the Hawkes-processes framework, that has proven to be a versatile tool for modeling of high-frequency market dynamics in physical time.

The paper is organized as follows. In Section 4.2 we present our dataset of prices and news and the main variables introduced to characterize the news. In Section 4.3 we show how Hawkes models are able to describe the dynamics of FX rates when only the endogenous component is considered. Section 4.4 presents our modeling approach, where a new component of the Hawkes process is added, taking into account the effect of macro news. In Section 4.5 we present an extension of the model where a non-causal kernel is added, taking into account the fact that the existence of the announcement (but not its content) is known in advance. Finally, in Section 4.6 we present our conclusions.

	# of days	Events per day (Avg.)	Avg. quote duration (s)
		Small tick size	
EUR/USD	190	58,902	0.55
EUR/JPY	190	34,857	0.91
USD/JPY	190	24,477	1.32
		Large tick size	
EUR/USD	60	20,229	1.60
EUR/JPY	60	19,012	1.70
USD/JPY	60	7244	4.47

Table 4.1: Summary statistics for the FX dataset during London core trading hours (07:30-16:30), before and after the tick size increase. Events refer to cases when either the bid or the ask change (with a time resolution of 100ms).

4.2 Data

We investigated EBS live data from January 1, 2012 to December 18, 2012, a total of 353 days, for three pairs of currencies, namely EUR/USD, EUR/JPY and USD/JPY. The time resolution of the database is 100ms and we used the *ebsMarketReferenceTime* as the reference time. Since the time discretization might affect the estimation of the Hawkes process, as customary done, we randomize the times by subtracting from each time stamp expressed in second a random number uniformly distributed in (0,0.1] s.

We filter the data by considering the events where either the best ask or the the best bid change. Although FX trading is active 24h a day 7 days a week, the activity on Saturdays and Sundays recorded in our dataset is almost negligible and therefore we excluded them. This left us with 250 days. FX trading activity displays an high intraday seasonality and therefore we decided to restrict our work to London core trading hours (07:30-16:30), where the highest activity is observed. Table 4.1 shows the number of bid or ask changes observed during the London core trading hours for the three rates. Notice that we have divided the sample in two, depending on the tick size[2]. In fact EBS increased the tick size of a factor five on September 23, 2012 and, as it can be seen from the Table, this fact has reduced the number of best quote changes by a factor two or three, depending on the rate.

The FX market is strongly affected by macro announcements. In this paper we are interested in modeling the price dynamics around these announcements. In particular here we focus on announcements whose publishing is scheduled well in advance, such as Central Banks rate decisions or announcements of unemployment or GDP data. We obtained the list of all economic announcements of 2012 from the website *www.dailyfx.com*. The news data report date and time of the news release, the currency which is most affected by the news, a brief description of the news, an estimate of the importance (high, medium, low), and, where available, the expected (Forecast) and actual value. The Forecast value is an average of analysts' expectations. Not all the news in the dataset concern the publication of an indicator which can be quantified by a number. An example are press conferences of central banks' governors. Since there is no numeric value associated to them, there is not any forecast either. Since we treat only the EUR/USD, EUR/JPY and USD/JPY rates, we retained only the news that are linked to one of these currencies. In addition, in this work we consider only news that are rated Medium or High importance, so we discarded

[2]The tick size is the discretization step of price.

those rated Low. Quite often, related indicators are announced at the same time. Since we are concerned with the impact of a news before and after its announcement, we kept only the most relevant news for each distinct time. The total of the news retained after this selection is 723. The totals divided per currency and importance are reported in Table 4.2. The much smaller number of JPY news is due to the fact that we focus on London core trading hours, and most JPY news are released outside this time interval.

	HIGH	MEDIUM	TOTAL
EUR	130	293	423
USD	107	186	293
JPY	1	6	7
TOTAL	238	485	723

Table 4.2: Number of news affecting each currency disaggregated by importance.

4.2.1 Intensity bursts and news surprise

In order to characterize the news, we introduce two additional metrics. The first one measures how much the news has impacted the trading activity, while the second one measures the degree of surprise of the news, i.e. the difference between the forecast and the actual value.

To be more specific, for each one minute interval i we compute the number N_i of changes in any of the best quotes, and we compare it with the Simple Moving Average (A_n) of the same quantity over the previous n one minute intervals. Since the events in minute i do not enter in the calculation of $\mathrm{A}_n(i)$, we consider it as a measure of the expected activity for the interval i. Similarly to what is done in price jumps detection Joulin et al. (2008); Bormetti et al. (2015), we define the *impact* θ_i as the ratio between the observed number of events at i and the expected number of events according to the Simple Moving Average A_n, namely

$$\theta_i = \frac{N_i}{\mathrm{A}_n(i)} \tag{4.1}$$

We shall use the quantity θ_i for two purposes. First we use it to identify "jumps" in the market activity and to answer the question of how frequently a high activity is associated with a recent news. Second we shall use is to answer the reverse question, i.e. how large is θ_i after a macro news and to classify news according to their effect on the trading activity.

The second metric measures the difference between expectations and actual values announced by the news. For this reason we call it "surprise" of the news. To compute it, we can only consider news for which the Forecast value is given. In the end, starting with 723 news, we are left with a total of 559 news or about 77% of the original sample.

For each news we have the Actual and the Forecast value, which we denote I_A and I_F respectively. The measures of surprise that we use are the absolute surprise

$$S_{abs} = |I_A - I_F|, \tag{4.2}$$

and the relative surprise

$$S_{rel} = \frac{|I_A - I_F|}{I_F} \times 100. \tag{4.3}$$

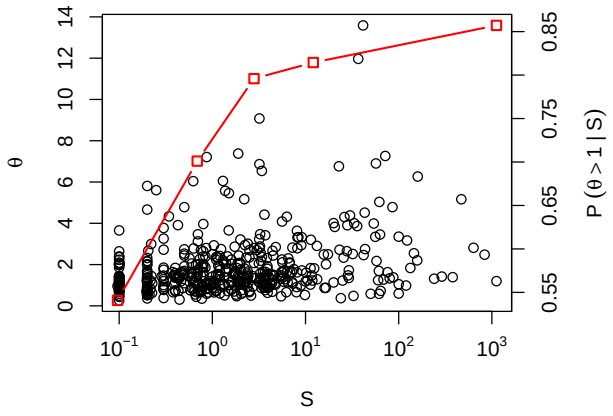

Figure 4.1: Scatterplot of θ against our measure of news surprise S (black circles, left y axis). The figure refers to EUR/USD data for θ. The red square line (right y axis) is an estimation of $P(\theta > 1|S)$. For reference, when $S = 0$ this probability is 0.73

In principle one could distinguish between "good news" and "bad news", i.e. if the indicator was better or worse than expected. However this is difficult to do in an unsupervised way, since a positive increment is good or bad depending on the type of news. Therefore we chose to use the absolute value of the difference. The choice between the absolute and the relative surprise is not immediate. On one hand, in fact, indicators span a wide range of values, some are of the order of unity other are given in thousands or in percentage and hence the most reasonable choice seems the relative surprise. On the other hand, for some indicators, such as unemployment rate, that are given in percentage, even a small absolute change might be significant. For these reasons we use the absolute surprise for indicators given in percentage and the relative surprise in all other cases. In the following, the notation S denotes this combined measure.

The correlation between surprise and θ is surprisingly low and the scatterplot is very noisy[3]. By restricting to Medium and High importance news, the Spearman correlation coefficient between surprise and θ is 0.34 and different from zero in a statistically significant way. This effect comes mainly from the High importance news, for which the Spearman correlation raises to 0.62. Figure 4.1 shows the scatterplot of θ vs. the surprise for EUR/USD for Medium and High importance news. This relation is confirmed by $P(\theta > 1|S)$, which is the probability that θ is larger than one conditional to a given value of the surprise (see Figure 4.1). We observe that this probability is systematically larger than 1/2, indicating that news trigger an increase in activity and this increase depends on the surprise of the news. For EUR/USD and EUR/JPY we obtain an almost monotone trend, while USD/JPY presents a noisier pattern. It seems that even this very simple measure of surprise, that does not take into account the specificity of each indicator, is nevertheless able to capture some relevant information.

4.3 Hawkes models of FX market activity

In this Section we present the baseline model for market activity. It consists of a Hawkes process with a constant baseline intensity and a self-excitation mechanism describing the endogenous effect Filimonov and Sornette (2012); Hardiman et al. (2013). In the next

[3]One could try to reduce this noise by normalizing the surprise indicator by the historical standard deviation, thus taking into account the typical uncertainty of the surprise of each news. This however would require much longer time series in order to have enough observations for each type of news.

Section we will introduce also an exogenous news term. Here we consider a univariate Hawkes process

$$\lambda_t = \mu + \int_{-\infty}^{t} \phi(t-s) dN_s \tag{4.4}$$

As a consequence of Proposition 2.2.1, the process is stationary if $n \equiv \int_0^\infty \phi(\tau) d\tau < 1$. As described in Brémaud and Massoulié (2001); Hardiman et al. (2013), the process can however be stationary in the critical case $n = 1$ if $\mu = 0$. If $n < 1$, from (2.14), it follows that n is related to μ through the relation

$$\Lambda \equiv \mathbb{E}\left[\lambda_t\right] = \frac{\mu}{1-n}. \tag{4.5}$$

The quantity n is the fraction of the average rate explained by the self-exciting mechanism.

The Hawkes process is characterized by the functional form of the kernel $\phi(t)$. In empirical applications, one can choose a priori a functional form for the kernel, dependent on some parameters, and then use maximum likelihood method to determine the best fitting parameters' values [4].

In our application, for the endogenous kernel we use either a double exponential[5] or a power-law decaying kernel. The double exponential kernel $\phi_{\mathbf{DE}}(t)$ is defined as

$$\phi_{\mathbf{DE}}(t) = \alpha_A e^{-\beta_A t} + \alpha_B e^{-\beta_B t}, \tag{4.6}$$

where α_q and β_q $(q = A, B)$ are the constant parameters of the model. Here two different time scales are present and their relative weight is controlled by the amplitudes α. For definiteness we choose $\beta_A \geq \beta_B$, i.e. the A subscript refers to the shorter time scale and B to the longer one.

The power-law kernel we use is in fact not a truly power-law decaying function, but instead, for computational reasons, we use an exponential sum that carefully approximates a function with a power law tail. In particular we adopt the same specification of Hardiman et al. (2013), which reads

$$\phi_{\mathbf{PL}}(t) = \frac{n}{Z}\{\sum_{k=0}^{M-1} (a_k)^{-p} e^{-\frac{t}{a_k}} - S e^{-\frac{t}{a_{-1}}}\} \tag{4.7}$$

where

$$a_k = \tau_0 m^k. \tag{4.8}$$

The parameters are n, p and τ_0. M is the number of exponential terms and m is a scale parameter. We fix $M = 15$ and $m = 5$ as in Hardiman et al. (2013). The parameter m controls how well a single exponential approximates the power law. If one wishes the approximation to be valid over many orders of magnitude, high values of m and M are necessary. Z and S are constants such that $\phi_{\mathbf{PL}}(0) = 0$ and $\int_0^\infty \phi_{\mathbf{PL}}(t) \mathrm{d}t = n$. This kernel has a smooth cut-off at short lags, provided by the negative exponential, whose time scale is regulated by τ_0 once m is fixed. This approximated power law kernel presents an exponential cut-off at large t. This difference with respect to a true power law is however irrelevant, in fact, for our choice of M and m, it manifests itself only for t larger than $\approx 10^9$s, a value much larger than the duration of a trading day. This approximation of the power law in terms of an exponential sum allows the log-likelihood to be computed recursively, reducing the computational cost form $O(N^2)$ to $O(N)$, where N is the sample size. This is particularly relevant given the large number of data available.

[4] Another parametric method was recently proposed in Da Fonseca and Zaatour (2014).

[5] We have also considered the single exponential case. In agreement with the results of Hardiman et al. (2013) for other financial assets, we find that FX data are very poorly described by Hawkes processes with a single exponential kernel.

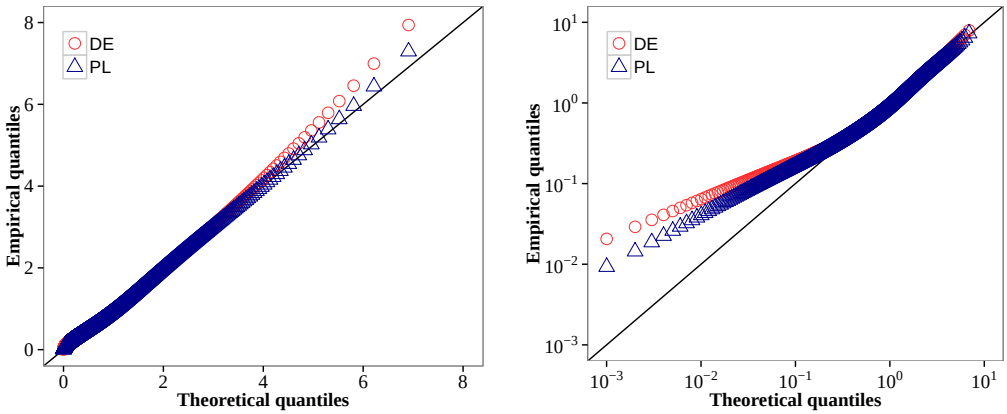

Figure 4.2: Comparison of the distribution of the residuals from the double exponential and power low model against the theoretical exponential distribution. Top: lin-lin scale. Bottom: log-log scale. The residuals from all the separate days and for all the three pairs are pooled together.

4.3.1 Results

We estimated the best parameters for each day (i.e. the 9-hour window from 07:30 to 16:30 UTC) separately by maximizing the log-likelihood. In this way we obtain a set of optimal parameters for each day.

In order to compare the performance of the two kernels we studied the distribution of the residuals. In fact the time change theorem Bauwens and Hautsch (2009) states that if the true data generating process is a point process with intensity λ, then the residuals

$$\Lambda_i = \int_{t_i}^{t_{i+1}} \lambda(s)\mathrm{d}s$$

are independent and identically distributed with a standard exponential distribution. In Figure 4.2 we compare by means of a quantile-quantile plot the empirical distribution of the residuals of both the double-exponential kernel and the power-law kernel against the standard exponential distribution. We find that the power-law kernel performs better than the double exponential. In fact, both the right tail and the left tail of the residual distribution for the power law kernel are closer to the theoretical one than the one for the exponential kernel. An analysis of the log-log plot in Figure 4.2 indicates that the major discrepancies occur for the left tail of the distribution, which corresponds to the residuals computed on short durations. The duration distribution of the data is peaked around 100 ms even after the randomization procedure. The double exponential kernel that presents its maximum at $t = 0$ is not able to reproduce this feature, while the power law kernel, which is equipped with an exponential cutoff at short lags performs better. To support the conclusions drawn from the QQ-plot we mention that we performed a test of exponentiality on the residuals, namely the Excess Dispersion (ED) test of Engle and Russell (1998). The test was performed on each day separately for the three currency pairs. The null hypothesis of the residuals being exponentially distributed was rejected at the 5% level in 92.8% of the days for the DE kernel and only in 42.2 % of the days for the PL kernel (aggregating across the pairs). To conclude the comparison between the two kernels, we stress that also likelihood values are consistently higher for the power law kernel, thus favouring this choice.

We now turn to the analysis of the values obtained for the parameters. We report the results from the power law kernel. Figure 4.3 shows the time series of parameters estimated

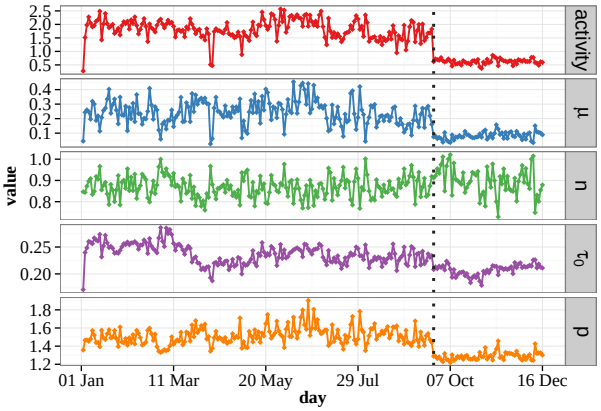

Figure 4.3: Time series of the Hawkes model parameters for EUR/USD. Market activity in events per second is also shown. The dotted line signals the tick size increase.

on EUR/USD data. Trading activity measured in events per second is also shown. In Table 4.3 we report the average value of the parameters obtained with the power law kernel before and after the tick increase for all the three pairs. The uncertainties on the values, as estimated from the Hessian of the log-likelihood, are about 10% on μ, 2% on n and τ_0, and about 1% on p.

	Before tick increase		
	EUR/USD	EUR/JPY	USD/JPY
activity (events/s)	1.82 (20)	1.08 (28)	0.76 (31)
μ (s^{-1})	0.24 (35)	0.17 (44)	0.12 (36)
n	0.87 (6)	0.86 (9)	0.84 (6)
τ_0 (s^{-1})	0.24 (8)	0.20 (11)	0.21 (7)
p	1.51 (6)	1.46 (10)	1.52 (7)
	After tick increase		
	EUR/USD	EUR/JPY	USD/JPY
activity (events/s)	0.62 (15)	0.59 (18)	0.22 (26)
μ (s^{-1})	0.08 (31)	0.08 (32)	0.03 (36)
n	0.90 (7)	0.91 (7)	0.95 (8)
τ_0 (s^{-1})	0.21 (4)	0.22 (16)	0.18 (4)
p	1.29 (3)	1.26 (3)	1.22 (3)

Table 4.3: Average values of the parameters before and after the tick size increase. Standard deviations across the sample expressed in percent is also reported in parentheses.

We note that the periods before and after the tick increase are very different. This is true also for the other pairs. After the tick increase the number of quote changes per second drops sharply as expected. In fact, after the increase a change in a quote implies a larger movement in the posted price. This drop is mirrored by a drop in the baseline intensity μ. The parameter p controlling the power law decay decreases sharply after the tick increase, indicating an increase of the number of long durations, in accordance with what is observed in the data. The criticality parameter n is quite close to 1 and relatively insensitive to the tick size change.

Overall, the Hawkes model equipped with the power law kernel provides a pretty good description of the empirical data. In line with Hardiman et al. (2013), we find that a large part of the activity is explained by the self-exciting mechanism (n close to 1) and hence appears to be endogenously generated.

4.4 Modeling market activity around macro news

We now present the main original contribution of this paper, namely an extension of the Hawkes model including a news-related (exogenous) term. The model aims at reproducing in a Hawkes framework the impact of news arrival on market activity. It is important to note that the news process is considered as given, that is, the model is not meant to describe the news process itself. The news process is completely deterministic in this context. As we said, this is actually also the real case in FX market, where the news and the forecast value, when available, are announced at predetermined times.

Let N_t^{news} be the counting process describing the arrival of a news. In order to avoid confusion with the times $\{t_i\}_{i\in\mathbb{N}}$ of arrival of the market activity process, we indicate with $\{z_j\}_{j\in\mathbb{N}}$ the arrival time of the j-th news. The intensity of the market activity counting process is then

$$\begin{aligned}\lambda_t &= \mu + \int_{-\infty}^{t} \phi(t-s)\mathrm{d}N_s + \int_{-\infty}^{t} \phi_N(t-s)\mathrm{d}N_s^{\text{news}} \\ &= \mu + \sum_{t_i<t} \phi(t-t_i) + \sum_{z_j<t} \phi_N(t-z_j)\end{aligned} \tag{4.9}$$

where $\phi(t)$ is the endogenous self-exciting kernel and $\phi_N(t)$ is the exogenous kernel that accounts for the news-induced excitement.

The condition for stationarity is obtained by taking the unconditional expectation on both side of (4.9),

$$\begin{aligned}\Lambda &= \mu + \int_{-\infty}^{t} \phi(t-s)\mathbb{E}\left[\mathrm{d}N(s)\right] + \int_{-\infty}^{t} \phi_N(t-s)\mathbb{E}\left[\mathrm{d}N^{\text{news}}(s)\right] = \\ &= \mu + \int_{-\infty}^{t} \phi(t-s)\mathbb{E}\left[\lambda(s)\right]\mathrm{d}s + \int_{-\infty}^{t} \phi_N(t-s)\mathbb{E}\left[\lambda_N(s)\right]\mathrm{d}s = \\ &= \mu + \Lambda\int_{-\infty}^{t} \phi(t-s)\mathrm{d}s + \Lambda_N\int_{-\infty}^{t} \phi_N(t-s)\mathrm{d}s = \\ &= \mu + \Lambda\int_{0}^{\infty} \phi(\tau)\mathrm{d}\tau + \Lambda_N\int_{0}^{\infty} \phi_N(\tau)\mathrm{d}\tau\end{aligned} \tag{4.10}$$

where $\Lambda_N \equiv \mathbb{E}\left[\lambda_N(t)\right]$ is the average intensity of the news process. Solving for Λ, we finally get

$$\Lambda = \frac{\mu + \Lambda_N \int_0^\infty \phi_N(\tau)\mathrm{d}\tau}{1 - \int_0^\infty \phi(\tau)\mathrm{d}\tau}. \tag{4.11}$$

Provided that Λ_N exists and is finite, the condition of stationarity is thus the same as in the pure self-exciting case, namely $\int_0^\infty \phi(\tau)\mathrm{d}\tau < 1$. In particular, it is not necessary that also $\int_0^\infty \phi_N(\tau)\mathrm{d}\tau < 1$ in order to have stationarity of market activity. As we will see below, in real data $\int_0^\infty \phi_N(\tau)\mathrm{d}\tau \gg 1$.

In order to test the model of Eq. (4.9) it is necessary to specify the functional form of $\phi(t)$ and $\phi_N(t)$. As we have seen above, the double exponential kernel and the quasi-power law kernel give good results for the endogenous component, hence we adopted these forms

also here. For the exogenous kernel ϕ_N we chose a single exponential specification of the type

$$\phi_N(t) = \alpha_N e^{-\beta_N t}. \tag{4.12}$$

Here, α_N gives the magnitude and β_N fixes the time scale. Moreover, it seems reasonable, at least as a first approximation, that the news impact could be described by a single timescale.

Both by using the double exponential $\phi_{\mathbf{DE}}(t)$ and the quasi-power law kernel $\phi_{\mathbf{PL}}(t)$ the expression for the log-likelihood can be analytically derived. In the former case we estimate 7 parameters, while in the latter case this number is equal to 6.

We estimated and tested our model on the series of best quote changes for the three currency pairs in our database. In each day there are on average three news announcements, and we calibrated the model on a window where only one news is present. Therefore we selected those news events that are "isolated", i.e. there is no other news in a three hour window centered at the time of the news. In this way the news event happens always at $t = 5400$ s, i.e. at the median point of the window. We then estimated the parameters of the model in each window as described in Appendix A. Notice that it is immediate to extend our modeling to the case where more than one news is present in the investigated time interval. In the following we focus on windows with only one news mostly for the sake of presentation clarity.

Since we are considering only High and Medium importance news, this filtering procedure returns a total of 266 isolated news. The news are the same for all the pairs. The average number of quote changes in each window is $17,579$ for EUR/USD, $11,199$ for EUR/JPY, and $7,322$ for USD/JPY.

4.4.1 Comparison of models with and without news

In order to visualize the difference between the model with and without the news component, we present here calibrated numerical simulations of the two models and we compare them with the real data. In Figure 4.4 (left) we consider an important news, namely, the release of US change in non-farm payrolls which turned out to be much lower than expected. The model without the news term fails to reproduce the trend of activity after the news. On the contrary, the model of Eq. (4.9) with the endogenous power law kernel seems to reproduce pretty well both the magnitude and the temporal decay of the news impact. The model with the double exponential kernel overestimates the impact immediately after the news arrival. Moreover, it is possible to note that the models without the news term significantly overestimates the activity before the news. This is because they attribute the extra activity due to the news to the self-exciting mechanism, thus increasing the overall activity instead of concentrating it in correspondence of the news.

Figure 4.4 (right) refers instead to the effects of the announcement of the Euro-Zone industrial new orders figures on USD/JPY. The figures were in line with expectations and the news had a scarce impact. Again, the model behaves nicely, in that it correctly captures the absence of impact in this case.

We compared the performance of the model with and without the news term across the whole set of news we examined. For each of the 266 news examined and for each currency pair, we simulated the two models 25 times. We then count the number of events produced by each model's simulation in the 5 minutes that follow the news announcement. In the top panel of Figure 4.5 we plot the ratio between the mean of the simulated counts and the observed (real) counts $\frac{<N_{\rm sim}>}{N_{\rm real}}$ in the 5 minute window that follows the news. The abscissa is the rank of the news in terms of its impact θ in ascending order. We

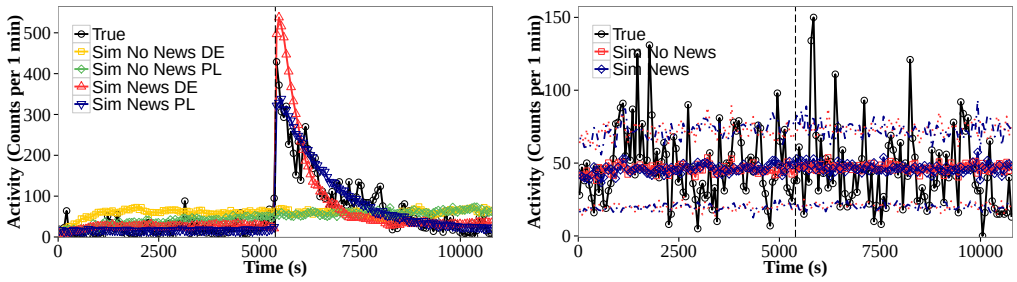

Figure 4.4: Actual and simulated activity around two macronews, namely EUR/USD rate around US change in non-farm payrolls on April 6 (left) and USD/JPY around Euro-Zone Industrial New Orders data on January 5 (right). The dashed line corresponds to the time of the announcement. Results from simulations of the models without news are also shown for comparison. For each model we run 100 simulations and we plot the average with one standard deviation confidence bands. Confidence bands are omitted in the top panel for improving readability. Parameter estimates were $\{\ \alpha_N^{\mathbf{DE}} = 3.2,\ \beta_N^{\mathbf{DE}} = 1.6 \cdot 10^{-2},\ \alpha_N^{\mathbf{PL}} = 1.45,\ \beta_N^{\mathbf{PL}} = 8.2 \cdot 10^{-4}\ \}$ for the first news and $\{\ \alpha_N^{\mathbf{DE}} = 3.0,\ \beta_N^{\mathbf{DE}} = 7.0,\ \alpha_N^{\mathbf{PL}} = 2.72,\ \beta_N^{\mathbf{PL}} = 7.0\ \}$ for the second one (all values in s^{-1}).

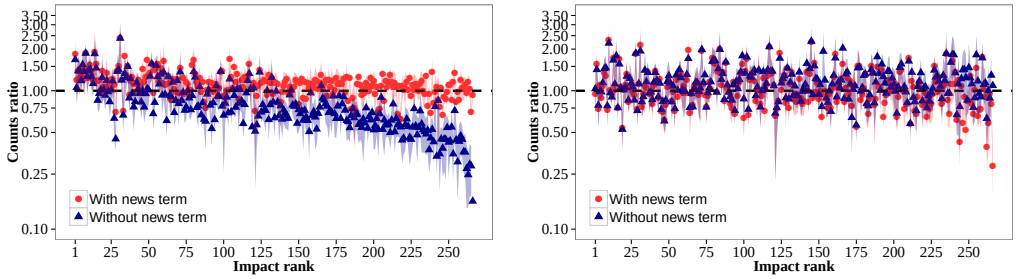

Figure 4.5: Ratio between mean of simulated counts and real counts in the 5 minutes window that follows (left panel) or precedes (right panel) the news plotted as a function of news' impact rank. Simulations were repeated 25 times for each news. All 266 news in our sample are considered. The figure shown refers to EUR/USD pair. A one standard deviation confidence interval is also plotted.

observe that the two models are essentially equivalent for low impact news, whereas the model with the news term performs increasingly better as the impact grows. In fact, the model without the news term systematically underestimate the number of events after high impact news, since the self-exciting endogenous mechanism is unable to reproduce such a sudden increase in activity. It is worth noting that the model with the news term does not artificially introduce a news effect when the news had little impact. For reference we also show in the bottom panel of Figure 4.5 the above ratio computed in the 5 minutes that precede the news. We observe that the two models produces almost identical outputs in this case.

These considerations are reinforced by observing Figure 4.6. The plot is built as follows: for each of the two models and for every news in the sample we divide the 3 hours window in bins of 5 minutes, and compute $\frac{<N_{\text{sim}}>}{N_{\text{real}}}$ in each bin. We also divide the news in four groups based on quartiles of the impact θ, labeled as "Very Low", "Low", "High",and "Very High". We then take the average of the ratio across all news in the same group for each 5 minutes bin. In this way we measure the averave behavior predicted by the models as a function of time distance from the news announcement. Also in this plot we observe that the models behave in the same way for very low impact news. However, as soon as the news has a significant impact, the model without the news term strongly

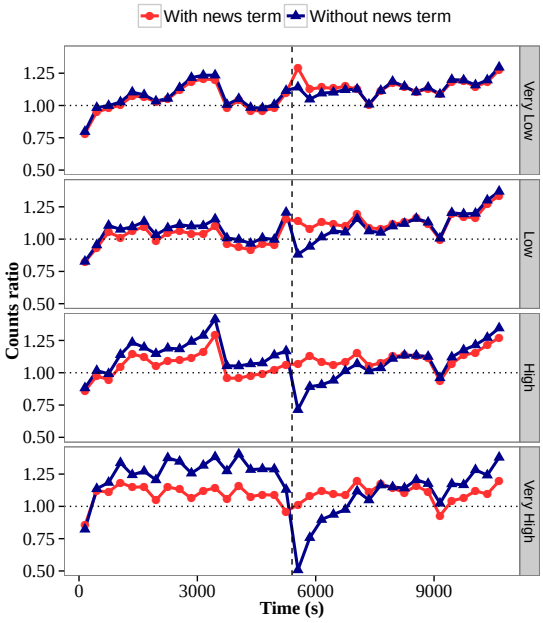

Figure 4.6: Ratio between mean of simulated counts and real counts computed in 5 minutes bin. Results form each bin are averaged across all the news in the same group of impact θ. The plot refers to EUR/USD data.

underestimates the activity that follows the news. Moreover, for very high impact news the model without the news term also systematically overestimate the activity before the news. This is because, without the exogenous kernel, the extra activity due to the news is attributed to the endogenous mechanism and thus it is spread over the full time interval. The result of this comparison indicates that the model with the news term performs much better than the one without it when the news had a sizable impact. When the news produced little or no impact, the two models are essentially equivalent.

In our model the total intensity λ_t is the sum of three contributions: the baseline intensity μ, the endogenous component $\int \phi(t-s)\mathrm{d}N_s$, and the exogenous component $\int \phi_N(t-s)\mathrm{d}N_s^{\text{news}}$. We can therefore analyse how the contribution of each term varies over time around the news announcement. The top panel of Figure 4.7 shows the fraction of the intensity coming from each component as a function of time for an important news. After the news release, the contribution of the exogenous term rises quickly, while the weight of the endogenous one decreases. Also the contribution of the baseline intensity becomes negligible immediately after the announcement. The exogenous contribution then slowly decays towards zero and the endogenous and the baseline components regain their pre-news level. It is interesting to contrast these figures with the middle panel of Figure 4.7, where the contributions to λ_t for the same news are shown for the model without the news term. The increment of activity after the news is now attributed to the endogenous component. The contribution of the baseline intensity has a similar trend in the two cases, though we note that its contribution before the news is higher in the model with the exogenous term. Finally, the bottom panel of figure 4.7 shows the same decomposition for a low impact news. It is possible to appreciate how, in this case, the exogenous component decays much faster and has a marginal role.

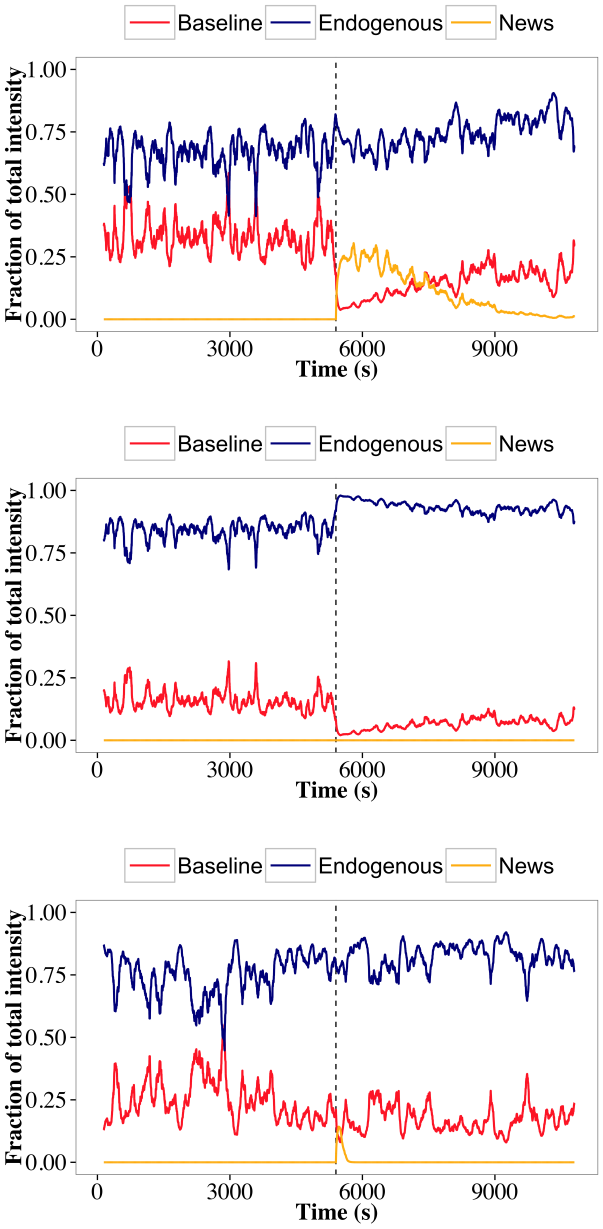

Figure 4.7: Components of the intensity λ_t as a function of time. The components sum to one and the curves are smoothed by using a simple moving average over 100 s. The top and middle panel refer to EUR/JPY, in correspondence of the announcement of the high importance and surprise news of US new jobs data on November 2 at 12:30. In the top (middle) panel we consider Hawkes model with (without) the exogenous news term. The bottom panel refers to EUR/USD during a medium importance news, namely the announcement of US new home sales data on September 26 at 14:00.

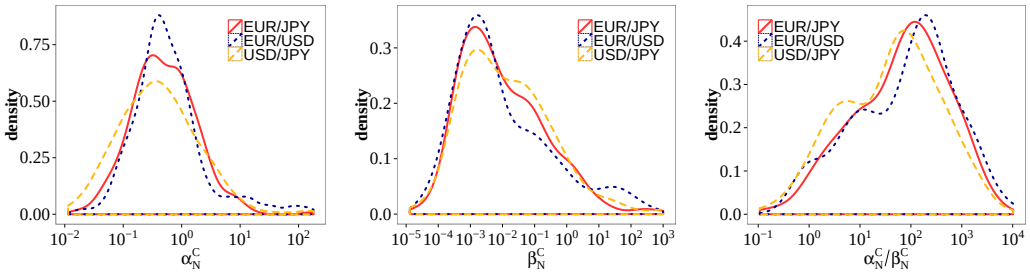

Figure 4.8: Estimated probability density function of α_N, β_N, and α_N/β_N for the three rates. The scale of the x axis is logarithmic, hence values of α_N and of the ratio α_N/β_N exactly equal to zero are not plotted. Their number is reported in Table 4.4.

	Number of news with $\alpha_N = 0$	$\overline{\theta}$
EUR/USD	10	0.70
EUR/JPY	13	0.69
USD/JPY	13	0.59

Table 4.4: Number of α_N found exactly equal to zero for each pair. The average value of the corresponding θ, $\overline{\theta}$, is also reported. We require α_N to be equal or greater then zero, and enforce this in the optimization. Hence, values equal to zero are found on the boundary of the searchable region.

4.4.2 Regression results

The regression of the model on the real data gives a set of parameters of the endogenous and exogenous kernel for each investigated news and exchange rate.

Figure 4.8 shows the distribution of the parameters α_N, β_N, and α_N/β_N for the three rates. Remarkably, the distributions are very similar across rates. It is worth mentioning that the figure does not show a delta peak in the distribution of α_N at the value zero. Their number along with the corresponding average impact is detailed in Table 4.4. We note that the news-related parameters have a very broad distribution. This is due both to the very different effects that news produce and to the uncertainty in the estimation. As a matter of fact, uncertainty on the news-related parameters, calculated from the inverse of the Hessian matrix, is very high, with averages of about 50% for both parameters. Finally, it is worth noticing that the distribution of α_N/β_N has most of its mass for values much larger than 1.

We now consider the problem of quantifying how much we improve the description of the data when we use the model with the exogenous term as compared to the performance of the simple model with only the endogenous factor. This is a typical problem of model selection (Claeskens and Hjort, 2008). Since we can compute the maximum likelihood and the number of parameters is different in the two models, we made use of the Akaike information Criterion (AIC) (Akaike, 1974) defined as

$$\mathrm{AIC} = 2k - 2\ln\mathcal{L}. \tag{4.13}$$

To corroborate our results we computed also the Bayesian Information Criterion (BIC) (Schwarz, 1978)

$$\mathrm{BIC} = k\ln n - 2\ln\mathcal{L} \tag{4.14}$$

where n represents the sample size. BIC thus generally penalizes more the introduction into the model of new free parameters than AIC does. The model with the lowest value of

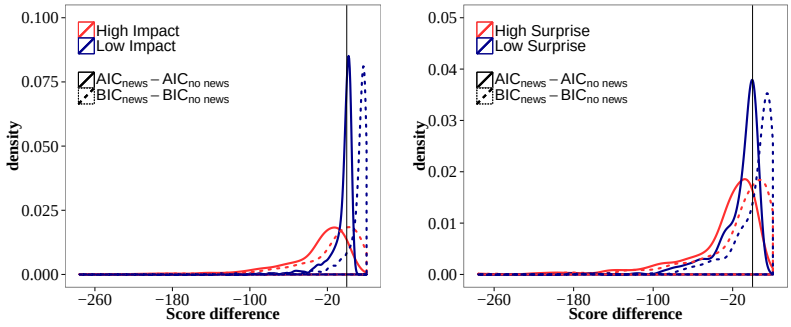

Figure 4.9: Distribution of the difference in AIC and BIC scores, $\text{AIC}_{\text{News}} - \text{AIC}_{\text{no News}}$ and $\text{BIC}_{\text{News}} - \text{BIC}_{\text{no News}}$, between the model with the news term and the one without it. Top: distributions for high impact news, $\theta > \theta_{\text{median}}$, and low impact news, $\theta \leq \theta_{\text{median}}$. Bottom: distributions for high surprise news $S > S_{\text{median}}$, and low surprise news, $S \leq S_{\text{median}}$. All the currency pairs are aggregated. Median values of θ for each pair are calculated before the aggregation.

AIC or BIC is the one to be selected. Given a set of models $\{i = 1, 2, .., M\}$ the relative likelihood RL of model i with respect to the best model in the set, for which $\text{AIC} = \text{AIC}_{\text{min}}$, is

$$\text{RL} = \exp\left(\frac{\text{AIC}_{\text{min}} - \text{AIC}_i}{2}\right). \tag{4.15}$$

AIC scores for the models with the power law endogenous kernel are always better than those for the models with the double exponential kernel. This is true as well for the BIC scores. Hence, in the following we compare the results of the models with the power law endogenous kernel that differ for the presence of the news term. Figure 4.9 shows the distribution of the difference in AIC and BIC scores between the model with the news term and the one without it, namely $\text{AIC}_{\text{News}} - \text{AIC}_{\text{no News}}$, and $\text{BIC}_{\text{News}} - \text{BIC}_{\text{no News}}$. The distribution is calculated separately for low impact and high impact news, and for high surprise and low surprise news. The figure is obtained pooling the values from all currency pairs. For AIC, only a small fraction of the differences are positive, meaning that the model with the news term is almost always better. However, for low-impact and low surprise news, there is a sharp peak around zero. A difference in AIC scores exactly equal to 4 corresponds to the case when the two models have the same likelihood, and hence the model with more parameters is penalized by a factor two times the number of extra parameters (which is two). Hence values around 4 mean that the models are essentially equivalent. As noted before, BIC penalizes more the extra parameters, so there is a larger portion of positive differences. However, for high-impact and high-surprise news the bulk of the distribution is on negative values. Indeed, negative differences attain even very large absolute values (as much as 260).

Looking at the average relative likelihoods RL for the two groups we obtain

Indicator	High	Low
θ	$5.6 \cdot 10^{-8}$	$1.0 \cdot 10^{-1}$
S	$1.5 \cdot 10^{-7}$	$1.3 \cdot 10^{-3}$

This means that for high impact news the model without the news term is about 10^{-8} times as probable as the model with the news term to minimize the information loss, i.e. the model with the news component is much better. Unsurprisingly, for low impact news

the difference between the models reduces, albeit AIC still favours the model with the news term.

As a final investigation we consider how the endogenous term parameters are influenced by the presence of the news term. For both models with the news term and the one without it, we consider separately the distribution of the parameters for news with $\theta \leq \theta_{\text{median}}$ ("Low") and those with $\theta > \theta_{\text{median}}$ ("High").

We find that the parameter that exhibits the most relevant difference when the news term is included is n. For high impact news n takes lower values in the model that includes the news term. In fact, an important part of the activity is now explained by the news term, whereas, in the previous model, it was attributed to the endogenous term. The baseline intensity μ has instead higher values in the model with the news term. This is at first sight somewhat surprising, since one could expect that without the news term a higher part of the observed intensity is attributed to the baseline intensity μ. The introduction of the news term impacts on n, reducing its value when the contribution of the news is significant (i.e. when the news term explains a large part of the activity which was previously attributed to the endogenous mechanism). Thus, also the contribution of the endogenous component to the average intensity Λ is lowered. On the other hand, the exogenous component contributes to Λ with $\frac{\Lambda_N \cdot \frac{\alpha_N}{\beta_N}}{1-n}$. In our case, $\Lambda_N = 1/5400$. What we observe is that this contribution is not enough to compensate for the lower values of n, and hence a higher μ is needed to maintain the same value of Λ. The tail of the endogenous kernel is estimated to decay faster to zero in presence of the news term and for high impact news. Instead, for low impact news, both models give about the same values. Finally, the values of τ_0 are not very sensitive to the presence of the news term.

In conclusion, the endogenous parameters are influenced by the presence of the news term. In particular, after the news an important fraction of the intensity is now attributed to the news term. We note that the value of n estimated with the news term may be underestimated when no news are present (e.g. before the announcement) and this in turn leads to higher values of μ to compensate. This result is due to the fact that we employ constant parameters, while it might be that the "right" values of n before and after a news are different. As we have seen, on average a large fraction of activity (≈ 0.9) is explained by the endogenous term.

4.4.3 Relation between the kernel parameters and the properties of the news.

In this Section we investigate how the parameters estimated by the Hawkes process with the exogenous news term are correlated with other measures of the effect of the news on the market activity, namely the jump parameter θ and the news surprise S (see Section 4.2.1).

We study which parameter of the news kernel is mostly affected when a news with high impact arrives. To this end, for each rate we split the sample of news in two subsamples, the first corresponding with news with $\theta > \theta_{median}$ and the second with $\theta < \theta_{median}$. For each of the three parameters α_N, β_N, and α_N/β_N we perform a t-test of the hypothesis that the means for each parameter are the same for high and low impact news. The results are summarized in Table 4.5. We observe that, while we cannot reject the hypothesis that α_N and β_N individually have the same mean in the two samples, the ratio α_N/β_N has different sample mean for large and small θ. The behaviour is similar for all the rates. This means that, even when a news has a large impact, the type of response of the kernel can affect either the amplitude or the time decay of the kernel (or both), but not in a

	parameter	t	df	$p-$value
EUR/USD	α_N	-0.11	238.28	$9.1{\cdot}10^{-1}$
	β_N	-1.27	144.30	$2.1{\cdot}10^{-1}$
	$\frac{\alpha_N}{\beta_N}$	3.95	181.06	$1.1{\cdot}10^{-4}$
EUR/JPY	α_N	-1.05	133.28	$3.0{\cdot}10^{-1}$
	β_N	-1.50	132.00	$1.35{\cdot}10^{-1}$
	$\frac{\alpha_N}{\beta_N}$	4.75	157.13	$4.5{\cdot}10^{-6}$
USD/JPY	α_N	-1.32	133.81	$1.9{\cdot}10^{-1}$
	β_N	-1.65	132.00	$1.0{\cdot}10^{-1}$
	$\frac{\alpha_N}{\beta_N}$	3.24	166.21	$1.4{\cdot}10^{-3}$

Table 4.5: Results of the t-test of the hypothesis that the parameters for high impact and low impact news have the same mean. The test is two sided and the Welch approximation is used for the degrees of freedom (df) to account for unequal variances in the two samples. We report the value of the t-statistic along with the computed $p-$value.

systematic way. The ratio between these quantities instead shows very different results for high and low impact news.

Finally, we examine the relation between the values of the parameters and the surprise indicator S defined in Section 4.2.1. To this end, we restricted the analysis to the news for which it was possible to calculate a surprise value, removing thus those for which the Forecast and Actual fields are not available. This left us with 213 news. Since our measure of surprise is pretty rough, we divided the sample in two subsamples, one of High Importance and one of Medium Importance news. As for θ, we further divided each of subsamples in two groups, one with a surprise value above and one below the sample median. For the rate EUR/USD a clear picture emerges. The rates α_N and β_N are not statistically different in the two groups, both for High and for Medium importance news. On the contrary, for High Importance news the ratio α_N/β_N is significantly larger for large surprise news as compared with small surprise news, while this effect is less evident for Medium importance news (p-values around of 10%). To illustrate the effect on High importance news, we divided the news and the corresponding estimated parameters into four groups based on the surprise value. The intervals are delimited by the quartiles of S. Figure 4.10 shows the probability distribution of the ratio α_N/β_N in the four quartiles, showing that higher surprise news correspond to higher value of the ratio. Similar results are observed for the other rates. In conclusion, even with our simple measure S, different levels of surprise give rise to different kernel ratio distributions.

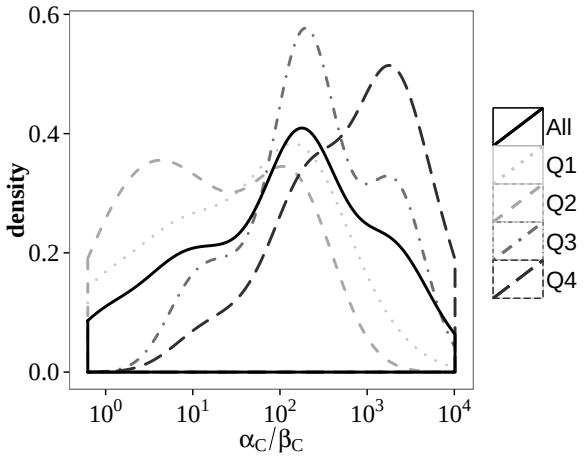

Figure 4.10: Estimated probability density function of the news parameter ratio α_N/β_N for four different groups formed on news surprise. Q1 indicates the fist quartiles of surprise, Q2 the second and so on. Also the total distribution is plotted for comparison. The rate is EUR/USD and we considered only High importance news.

4.5 A Hawkes process with a non causal kernel

The announcement of the macroeconomic news we are considering in our analysis is known in advance to market participants. Therefore it is natural to ask whether the point process describing the market activity is affected by the news arrival *before* its announcement. In terms of stochastic processes this implies the presence of a non-causal kernel, in the sense that the (pre-announced) news arrival affects the intensity at earlier times.

To consider this effect, we modify the model by introducing a non-causal term in the exogenous component of the intensity. Namely, the news kernel $\phi_N(t)$ is now given by the sum of a causal and a non-causal term

$$\phi_N(t) = \Theta(t)\phi_N^{\mathrm{C}}(t) + \Theta(-t)\phi_N^{\mathrm{NC}}(t) \tag{4.16}$$

where $\Theta(t)$ denotes the Heaviside step function and we assume $\Theta(0) = 0$. The intensity function now reads

$$\begin{aligned}\lambda_t &= \mu + \int_{-\infty}^{t} \phi(t-s)\mathrm{d}N_s + \int_{-\infty}^{+\infty} \phi_N(t-s)\mathrm{d}N_s^{\mathrm{news}} = \\ &= \mu + \sum_{t_i<t} \phi(t-t_i) + \sum_{z_j} \phi_N(t-z_j) = \\ &= \mu + \sum_{t_i<t} \phi(t-t_i) + \sum_{z_j<t} \phi_N^{\mathrm{C}}(t-z_j) + \sum_{z_j>t} \phi_N^{\mathrm{NC}}(t-z_j)\end{aligned} \tag{4.17}$$

We chose an exponential specification also for the non-causal term, so that $\phi_N(t)$ is now written as:

$$\phi_N(t) = \Theta(t)\alpha_N^{\mathrm{C}} e^{-\beta_N^{\mathrm{C}} t} + \Theta(-t)\alpha_N^{\mathrm{NC}} e^{\beta_N^{\mathrm{NC}} t} \tag{4.18}$$

with α_N^{C}, $\alpha_N^{\mathrm{NC}} \geq 0$, and β_N^{C}, $\beta_N^{\mathrm{NC}} > 0$.

The stationarity condition now reads

$$\Lambda = \frac{\mu + \Lambda_N \left[\int_0^\infty \phi_N^{\mathrm{C}}(\tau)\mathrm{d}\tau + \int_{-\infty}^0 \phi_N^{\mathrm{NC}}(\tau)\mathrm{d}\tau\right]}{1 - \int_0^\infty \phi(\tau)\mathrm{d}\tau}. \tag{4.19}$$

As before, provided that Λ_N exists and is finite, the condition of stationarity is thus the same of the self-exciting only case, namely $\int_0^\infty \phi(\tau)\mathrm{d}\tau < 1$.

We re-estimated the parameters of the model with the non-causal term on the same dataset used for the causal-only model. In Figure 4.11 we show simulations of the causal-only and non-causal models compared against the real data for an important news. We also add one standard deviation confidence intervals obtained by performing a large number of simulations of the model. We note that, especially by looking at the zoomed time series (bottom panel), before the announcement the confidence bands of the two models almost do not overlap, and the non-causal model reproduces quite well the news-anticipation effect when it is present, whereas the causal-only model misses completely this feature of the data.

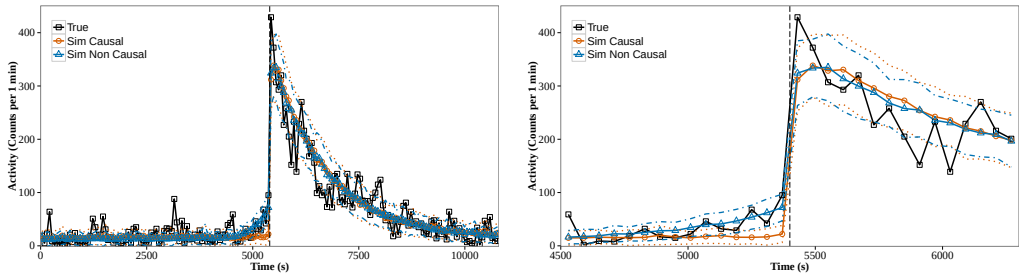

Figure 4.11: Actual and simulated activity measured as events per 1 min for EUR/USD on Friday April 06, when an important figure on US change in non-farm payrolls was released at 13:30. The figure produced an impact $\theta \approx 13$ and was much worse than expected (Surprise =44). The dashed vertical line corresponds to the time of the announcement. The bottom panel is a zoom of the top panel in a 30 minute window around the news. Results from simulation of the models with and without the non causal term are compared. The dotted and dashed-dotted lines are one standard deviation confidence intervals obtained from simulations of the calibrated model of the causal-only and non-causal model, respectively. Parameters estimates for the non causal term are $\alpha_N^{\mathrm{NC}} = 0.3$, $\beta_N^{\mathrm{NC}} = 3.0 \cdot 10^{-3}$ (all values in s^{-1}).

We then compare the extended model with the causal-only one via AIC and BIC. Figure 4.12 shows the distribution of the differences $\mathrm{AIC}_{\mathrm{NC}} - \mathrm{AIC}_{\mathrm{C}}$ and $\mathrm{BIC}_{\mathrm{NC}} - \mathrm{BIC}_{\mathrm{C}}$. Results from all currency pairs are pooled together. As the relatively large number of news for which $\alpha_N^{\mathrm{NC}} = 0$ suggests, the news-anticipation effect is often very small or absent. In these cases, the extended model with the non-causal term reduces in practice to the causal-only model. Hence, the extra parameters do not carry any benefit in these situation, and this is reflected in the information criteria scores. On the other hand, when the news anticipation effect is relevant, the improvement of the extended model is significant and AIC and BIC scores strongly favour the introduction of the non-causal term. Moreover the parameters of the causal news kernel are essentially unchanged by the introduction of the non-causal kernel.

Overall it seems that adding the non-causal term results in an improved description of the real data, albeit in some cases we simply recover the performance of the already discussed causal-only model.

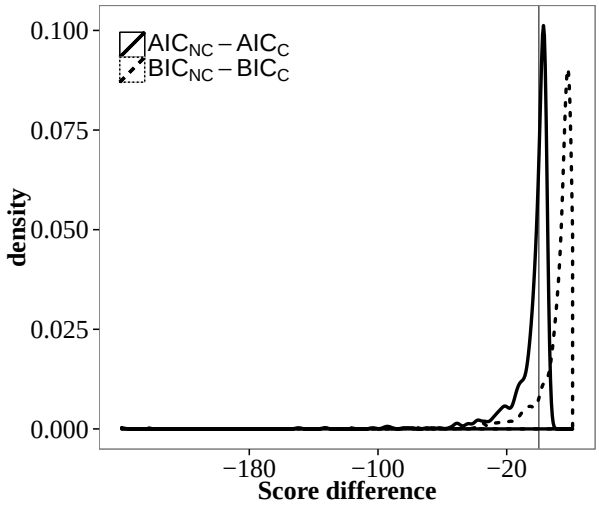

Figure 4.12: Estimated probability density function of the differences in AIC and BIC scores between the extended model with the non-causal term and the causal-only model. Differences from all three currency pairs are aggregated.

4.6 Conclusions

In this paper we have introduced a Hawkes process approach to the description of foreign exchange rate dynamics around the announcement of macroeconomic news. These news are very important and might affect in a significant way the trading activity in a time window around the announcement. We have first considered the unconditional Hawkes modeling, i.e. we have neglected the role of news. We found that, as seen in other asset classes, foreign exchange markets are very close to criticality, as measured by the integral of the kernel n. Moreover, while the parameters of the kernel show an abrupt change around the change of the tick size, the criticality parameter n seems to be pretty insensitive to it. We then considered our main original contribution, namely the introduction of a kernel describing the effect of macro news on the trading activity. Since announcements are known in advance, we have also considered the possible presence of a non causal kernel, describing how the market prepares itself to the news. In both cases we have shown that the model with the news kernel(s) outperforms the simple Hawkes model with only endogenous kernel. We also noted that, once the news term is introduced, the estimate of the criticality parameter n is smaller. This suggests that in the presence of localized exogenous excitations a Hawkes model which does not consider news-triggered non stationarity could overestimate n. Moreover, even if the non-causal part only helps in a fraction of cases, from the practical point of view of a market maker (or anyone trading around news) being able to gauge the market activity and liquidity at those particular points is very important and there is a premium to be able to get that estimate right (or big cost to getting it wrong). Finally, we have also explored the role of the surprise of the news with respect to forecasted value on the parameters of the kernel.

Despite being focused on financial market data, the approach presented in this paper could be useful in the analysis and modeling of other complex systems. In fact the presence of exogenous and endogenous drivers of the activity is ubiquitous in other systems monitored in continuous time. Describing the different source of excitation as correlated point processes is in general quite complicated, and the Hawkes approach proposed here could be a useful method to model, fit, and evaluate the relative importance the two drivers of

a generic stochastic dynamics.

In the next Chapter we improve on this work by introducing a procedure that allows us to estimate also the intensity burst location, thus widening the possible application domain beyond scheduled announcements. There we deal also with the presence of multiple bursts in the same window.

Chapter 5

Hawkes processes with local non-stationarities (II)

5.1 Introduction

The detection of anomalous dynamics and regime changes is an important problem which has a wide range of applications in many systems from risk management to preventing of fraud. Financial markets are emblematic in this respect. Market participants, and particularly those who act as intermediaries, need to be resilient to sudden changes in market conditions that can arise endogenously or when a new piece of information becomes available. However, the identification of anomalous dynamics might be challenging, especially when the "normal" dynamics is complex, possibly with non-linear and/or long range correlations. In finance, for example, a large amount of research has been devoted to the detection and characterization of "anomalous" price changes. Given an underlying price dynamics described by a continuous semimartingale, the problem is to test for the presence of price discontinuities (jumps - see for example Andersen et al. (2007); Lee and Mykland (2008); Bollerslev et al. (2009); Aït-Sahalia et al. (2009) to cite only a few).

In many real systems the observed time series is described by a point process. Examples include the arrival of phone calls, email, tweets, customers, queries, etc., and of course also financial time series, since at the smallest time scale the change in price is generically described by a (marked) point process. To the best of our knowledge the identification of anomalous changes in the intensity of a point process has remained relatively unexplored, especially when the "normal" dynamics presents correlations and bursts of activity. In this chapter we propose a method for tackling this problem. Specifically, we aim at identifying abrupt increases of the intensity of a otherwise stationary, yet correlated and bursty, point process, and where these increases relaxes back to the normal state after a certain period of time. We will call such events *intensity bursts (IBs)*. Detecting IBs might be important to identify anomalous activities in the system, detect the arrival of external perturbations, analyze the contagion effects and lead-lag relations in different co-evolving systems.

We propose a parametric model-based approach for the detection of intensity bursts, which relies on the Hawkes process (Hawkes, 1971). Being closely related to branching processes (Harris, 2002), the Hawkes model combines in a very natural way external (exogenous) influence on the system with internal (endogenous) self-excited dynamics. Here the probability of arrival of a new event is given by the combination of a baseline probability and a contribution from all the former generated events. In natural and socio-economic systems, Hawkes processes have become a very popular tool due to their flexibility and, at the same time, the simplicity of the calibration procedure. They have been applied in a

variety of research domains such as seismology (Ogata, 1988), genomics (Reynaud-Bouret et al., 2010), neurophysiology (Pierre Bremaud (1996); Chornoboy et al. (1988)) as well as in works on the spread of crime and violence (Lewis et al., 2012; Mohler et al., 2011) and on social network dynamics (Crane and Sornette, 2008; Blundell et al., 2012; Zhou et al., 2013). Within the domain of financial applications, the Hawkes process has become a widespread model for the dynamics of high-frequency price changes and order book evolution (Bacry et al., 2013; Bormetti et al., 2015; Hardiman et al., 2013; Filimonov and Sornette, 2012; Bowsher, 2007; Embrechts et al., 2011; Blanc et al., 2015) and also was extended to the modeling of the price shocks on a daily scale (see for instance Bauwens and Hautsch (2009) and Bacry et al. (2015) for a review).

We choose to model the "normal" dynamics of the point process with a Hawkes process, so that correlations and bursts are present. On top of this we assume that in the process there are few IBs that locally perturb the counting dynamics and, via the excitation mechanism of the Hawkes process, affect also the "normal" dynamics. In this setting, the parameters of the model and the location of the IBs are not known and must be inferred from the data.

In Rambaldi et al. (2015) two of us have considered the related but much simpler problem of inferring the parameters of the model when the location of the IBs is known. The considered example was the arrival of scheduled macro-economic announcements in financial markets. These events have a major and dramatic impact on both price and trading activity and cannot be described within the classical Hawkes model, but the their timing is known in advance. In the general case, the occurrence and timing of IBs are not known. The major limitation of Rambaldi et al. (2015) is precisely the assumption that the occurrence time of the IBs is known, which prevents the study of unexpected (surprise) IBs. Interestingly, even when the expected time is known in advance, a delay or anticipation of the actual arrival of the external event can bias the procedure.

In the present work, instead, we consider the general case when both IBs arrival times and the total number of IBs are unknown. We propose an efficient and robust procedure of estimation from the empirical data, and we also suggest a hypothesis test that distinguishes the genuine intensity burst from a statistical fluctuation that could be equally well explained within a standard self-excited dynamics. We validate our procedure and show that it is capable of reliably identifying sudden increases in the event rate with a relatively low false positive rate. Furthermore, being a model-based, our approach allows to estimate parameters of the intensity burst together with the properties of the underlying self-excited process, and use it for classification of the anomalous activity.

We apply our approach to the analysis of high-frequency financial data from a major FX market. We identify a large number of IBs and we compare them with price jumps and macro-economic announcements. Interestingly, we find a large number of IBs which are not explained (in terms of time proximity) by neither of these two possible causes. This is noteworthy, since it signals the presence of different market anomalies and we propose possible explanations for them.

We foresee several major directions along which our method could be useful. First, the detection of an anomalous an unexpected market activity is essential for monitoring liquidity and for intra-day risk management applications. Moreover, the analysis of the historic data is essential for forensic investigations and detection of fraud and market manipulations. Besides, our method is relevant even when the analysis of the intensity burst itself is not the primary interest, and the focus is on describing the underlying process. Indeed, the procedure we propose mitigates Hawkes process tendency to overestimate the degree of self-excitation when burst-like non-stationarities are present. Finally, we stress that

the potential range of applications spans well beyond the financial markets. For example our burst detection method could be of great importance in the analysis of social media dynamics such as YouTube views (Crane and Sornette, 2008) or twitter posts (MacKinlay, 2015), or even in the analysis of data of social conflicts and unrest (Donnay and Filimonov, 2014).

The remaining of the chapter is structured as follows. We start with the description of the Hawkes model that contains exogenous intensity bursts in Section 5.2. Section 5.3 presents the methodology for the statistical identification of intensity bursts. We test and validate our procedure on synthetic data in Section 5.4. Section 5.5 presents an application of our methodology on real financial data. We conclude in Section 5.6.

5.2 Hawkes process with an exogenous intensity burst

A univariate point process is a sequence of events that occurred at random times t_i and that is described by the corresponding counting process $N_t = \sum \mathbb{1}_{t_i<t}$, where $\mathbb{1}_A$ is the indicator function of the set A. The most known example of point process is the homogeneous Poisson process, where the intensity of the events is constant $\lambda_t = \lim_{\Delta\to 0} \Delta^{-1}\mathbb{E}\left[N_{t+\Delta} - N_t\right] = \mu$.

Hawkes (1971) extended the Poisson process to account for self-excitation in the system by modeling the intensity conditional on the history of the process $\mathcal{F}_t = \{t_i : t < t\}$: $\lambda_{t|\mathcal{F}_t} = \lim_{\Delta\to 0} \Delta^{-1}\mathbb{E}\left[N_{t+\Delta} - N_t|\mathcal{F}_t\right]$. In the standard univariate Hawkes model, this intensity reads

$$\lambda_{t|\mathcal{F}_t} = \mu + \int_{-\infty}^{t} \phi(t-s)\mathrm{d}N_s = \mu + \sum_{t_i<t} \phi(t-t_i), \tag{5.1}$$

where the first part (μ) gives the so-called *baseline intensity* of the classical exogenous Poisson process and the second one describes the explicit endogenous impact of the past events on the future. Here $\phi(t)$, called the *memory kernel*, is a non-negative function that specifies how past events contribute to the generation probability of future events and thus $\phi(t)$ controls the amplitude of the feedback mechanism.

The linear structure of the intensity (5.1) allows one to map the Hawkes process exactly onto a cluster process (Hawkes and Oakes, 1974), where the process consists of a superposition of random clusters, each of which starts with a single *immigrant*, generated from a homogeneous Poisson process with intensity μ. In turn the immigrants generate their offspring according to an inhomogeneous Poisson process with intensity $\phi(t)$, and these next-order events generate their own offspring with the same mechanism. Theoretically, such construction could be well-described using the theory of branching processes (Daley and Vere-Jones, 2008), and from a practical perspective such representation allows to perform efficient numerical simulation of the Hawkes process (Møller and Rasmussen, 2005). The branching context also provides an intuitive constraint on the kernel $\phi(t)$ that ensures the stability and stationarity of the system. The average number of offsprings generated by a single event is $n = \int_0^\infty \phi(t)$, called a *branching ratio*. For the process to be stable, this quantity has to be smaller than 1 ($n < 1$). The condition $n = 1$ corresponds to a tipping point and the system with $n > 1$ will exhibit exploding dynamics with total population increasing to infinity with probability one.

In Rambaldi et al. (2015) authors have further extended the Hawkes model.Instead of insisting that all immigrants arrive at random times following a homogeneous Poisson process, authors have proposed that some external events could have significantly affect the evolution of the system. Such external events can locally give rise to much more

immigrants than under the sole baseline intensity. This was modeled via a separate term in the intensity expression:

$$\lambda_t = \mu + \sum_{j=1}^{M} \phi_S^j(t - z_j) + \sum_{t_i<t} \phi(t - t_i) \tag{5.2}$$

Here the second term describes the impact of M exogenous events arriving at times z_j, that are assumed to be deterministic and known. These special exogenous events increase the rate of immigrants arrival via the memory kernels $\phi_S^j(t)$. The system further reacts in a regular way via the kernel $\phi(t)$ thus amplifying the effect of the external events.

Effectively, these special exogenous events and their clusters introduce Intensity Bursts (IBs) into the system. For example, within the modeling of socio-economic systems, such intensity bursts could correspond to the reaction of the system to major events, such as elections, referendums and other political events, regulatory changes or even announcements of the news related to a particular sector or company.

The expected number of new immigrants generated by a single exogenous IB is

$$f_j = \int_0^{+\infty} \phi_S^j(s)\mathrm{d}s, \tag{5.3}$$

and we will refer to this quantity as the *fertility* of the j-th IB. In contrast to the branching ratio n of the whole system, there are no restrictions on this parameter, except that the integral in (5.3) should converge. Moreover, we will be interested in cases when the exogenous event represent a "burst", and thus its fertility is much larger than of a regular event: $f_j \gg 1$. The immigrants directly triggered by the IB will produce their own offsprings according to the memory kernel $\phi(t)$ and thus the expected total number of events in each exogenous cluster is given by

$$S_j = f_j + f_j \frac{n}{1-n} = \frac{f_j}{1-n} \tag{5.4}$$

Finally, in order to fully specify the model, we need to define the functional form of memory kernels. There is no unique specification and the shape of the memory kernel will vary from one application to another. Within the domain of modeling the high-frequency financial data, which is the application discussed in this work, the question whether the kernel $\phi(t)$ should be short- or long-memoried is rather controversial. Some works show empirical support for the power law decaying functions (see e.g. Hardiman et al. (2013); Bacry and Muzy (2016)), while other works suggest use of sum of a few exponential functions (Lallouache and Challet, 2016; Martins and Hendricks, 2016) and point out dangers of using long-memory kernels on non-stationary data (Filimonov and Sornette, 2015).

For our purposes we will follow the original work of Rambaldi et al. (2015) and adopt an approximation of the power-law kernel via a sum of exponentials that was originally suggested in (Hardiman et al., 2013):

$$\phi(t) = \frac{n}{Z} \left\{ \sum_{k=0}^{K-1} (\tau_0 m^k)^{-p} e^{-\frac{t}{\tau_0 m^k}} - S e^{-\frac{tm}{\tau_0}} \right\}. \tag{5.5}$$

It approximates a power law decay with exponent p and it also features an exponential cutoff at short time scales. The parameter τ_0 controls the position of the maximum and n is the branching ratio. Further, $m = 5$, $K = 15$ and values of S, Z are fixed such

that $\phi(0) = 0$ and $\int_0^\infty \phi(s)\mathrm{d}s = n$. This kernel specification has been shown (Rambaldi et al. (2015), Hardiman et al. (2013)) to be well suited to model financial data. Moreover, the slow decay of the kernel creates relatively strong point clustering and long range dependencies between counts of the process. This allows us to test our procedure in a somewhat harsher environment, as it makes more difficult to separate genuine external IBs from endogenously generated bursts.

Finally, for the memory kernel $\phi_S(t)$ of the IB we will follow Rambaldi et al. (2015) and adopt the exponential specification:

$$\phi_S(t) = \alpha e^{-\frac{(t-z)}{\tau}} \mathbb{1}_{t>z} \tag{5.6}$$

where $\alpha > 0$ and $\tau > 0$ are parameters. In this case we have $f = \alpha\tau$. This specification reflects the idea that an external event has a very strong immediate impact and also some time persistence, and finally its effects fades completely. Nevertheless, in other domains different specification of ϕ_S could be more appropriate, and our procedure could be easily adapted to them.

5.3 Identification of the intensity bursts

In the original work of Rambaldi et al. (2015), the times of the IBs $\{z_j\}$ as well as their number M were assumed to be known. Here we extend the framework to define a rigorous procedure of the detection of both the number and the timestamps of the exogenous shocks.

5.3.1 Identification of a single IB

For clarity, let us start with the case where we assume that at most one IB is present, but we do not know its occurrence time. In this case our model reads

$$\lambda_t = \mu + \sum_{t_i<t} \phi(t - t_i) + \phi_S(t - z) \tag{5.7}$$

and now z is a parameter of the model.

In general, once the Hawkes model is specified, we can apply a range of methods in order to calibrate the model on the observed data and estimate its parameters. Such methods include maximum likelihood estimation (Rubin, 1972; Ogata, 1978; Ozaki, 1979) and method of moments (Da Fonseca and Zaatour, 2014). A range of non-parametric tools (Bacry et al., 2012; Bacry and Muzy, 2016; Lewis and Mohler, 2011; Kirchner, 2015) are also available, these are however mostly limited to the classical Hawkes specification (5.1). Within the parametric estimators the Maximum Likelihood Estimator is considered to be the de-facto standard for the Hawkes process family. Given the functional form of the intensity (5.7) and the observations $\mathcal{F}_T$ in the interval $[0, T]$ one can maximize the log-likelihood

$$\log \mathcal{L}(\theta|\mathcal{F}_t) = -\int_0^T \lambda_s \mathrm{d}s + \int_0^T \log \lambda_s \mathrm{d}N_s \tag{5.8}$$

in order to obtain the vector of parameters θ, which includes the time z of the IB and parameters of the kernel ϕ_S that defines the fertility f. However, a-priori we do not know if the realization contains any IBs. Thus, we need to decide whether the identified burst in activity is indeed genuine or could be attributed to the endogenous mechanism alone. For this we will test if the extension of the model from a simple Hawkes process (5.1) to

a model with a shock (5.7) improves the description of the data. A natural approach here is to compare the likelihood of model (5.7) evaluated at the optimal parameter with the corresponding best fit from a Hawkes model without IBs.

Ogata (1978) proved under certain regularity assumptions that the maximum likelihood estimate for a simple, stationary, univariate point process is consistent and asymptotically normal as the sample size tends to infinity. Moreover, he also established that the likelihood ratio test of a simple null hypothesis possesses the standard χ^2 distribution. Therefore, as suggested also in Gresnigt and Franses (2015), likelihood tests such as the Likelihood Ratio (LR) test and the Lagrange Multipliers (LM) test can in general be used to discriminate between different Hawkes specification.

However, our case presents some difficulties that make such tests not readily applicable. Our situation is akin to that found in regime shift problems when the change point in unknown. Moreover, some of the parameters are not identified under the null hypothesis that no IB is present. Indeed, the model with the IB term described by (5.6) reduces to the null model when $\alpha = 0$. Hence, τ and z are not identified under the null. One can also argue that the reduction to the null is not unique, since the effect of the IB is removed also for $z \geq T$, or $\tau \to 0$, or $z < 0$ and $\tau \to \infty$. As discussed in (Andrews, 1993; Davies, 1977, 1987; Hansen, 1996) the standard asymptotic theory of LR does not apply in these cases and other approaches have to be taken.

One stream of literature deals with this problem through the so called "sup" class of tests (see Lange and Rahbek (2009) for a survey), which rely on simulations of the LR distribution under the null to compute the appropriate p-values.

Another approach (Wong and Li, 2001) uses information criteria for model selection. The most widely employed are Akaike information criterion

$$\mathrm{AIC} = 2k - 2\log\mathcal{L} \tag{5.9}$$

where k denotes the number of estimated parameters, and the Schwartz (or Bayesian) information criterion (BIC)

$$\mathrm{BIC} = k\log N - 2\log\mathcal{L} \tag{5.10}$$

which penalizes more heavily extra parameters if, as in our case, the sample size $N > 7$ (i.e. $\log N > 2$). In this work we follow this second approach, and will document that the BIC performs well in our test (see Section 5.4).

We can summarize our procedure for the identification of a single IB as follows:

1. the null model (Hawkes model without any IB term (5.1)) is estimated using the Maximum Likelihood Estimator.

2. the alternative (extended) model (5.7) is estimated with the constraint of $z \in [0, T]$

3. the score $\Delta\mathrm{BIC} = \mathrm{BIC}_1 - \mathrm{BIC}_0$ is evaluated. If $\Delta\mathrm{BIC} < 0$ then the null model is rejected and we accept the extended model. Otherwise we retain the null model of no exogenous bursts.

Here we have denoted with BIC_M the score of the model with M IBs. In case of exponential memory kernel of IB (5.6) the difference in numbers of parameters of null and alternative model is 3 and thus $\Delta\mathrm{BIC} = 3\log N - 2(\log\mathcal{L}_1 - \log\mathcal{L}_0)$. We refer to Appendix 5.A for more details on likelihood optimization.

5.3.2 Identification of multiple IBs: pre-identification

We now discuss the extension of our procedure to the case of more than one IB in the window $[0, T]$, i.e. the model (5.2), where z_j are parameters of the model and also the total number of IBs M has to be determined. Because of the numerical complexity of the problem, often noisy data, and the curse of dimensionality, straightforward estimation of the full model (5.2) leads to an optimization of a multi-variate cost-function with multiple local extrema. In order to efficiently estimate the parameters z_j in (5.7) it is useful to restrict the search space for each z_j to some interval $[z_1^j, z_2^j] \subset [0, T]$. Doing so significantly improves the convergence and reduces the number of local minima. Moreover, it allows us to focus on one IB at a time and thus to increase sequentially the size of the model.

The problem we address in this section is thus how to efficiently and reliably reduce the search space for z via pre-identification. The following section discusses the complete procedure.

We adapt the method, originally proposed by Almgren (2012), for detection of price jumps in high-frequency data to a point process setting. Following his work we define the exponential averaging functions $u_L(t;\kappa)$ and $u_R(t;\kappa)$ as

$$u_L(t;\kappa) = \frac{1}{\kappa}\int_{-\infty}^{t} e^{-\frac{(t-s)}{\kappa}}\,\mathrm{d}N_s = \frac{1}{\kappa}\sum_{t_j<t} e^{-\frac{(t-t_j)}{\kappa}} \tag{5.11}$$

$$u_R(t;\kappa) = \frac{1}{\kappa}\int_{t}^{+\infty} e^{-\frac{(s-t)}{\kappa}}\,\mathrm{d}N_s = \frac{1}{\kappa}\sum_{t_j>t} e^{-\frac{(t_j-t)}{\kappa}}, \tag{5.12}$$

where κ is a parameter of the method.

The functions $u_L(t;\kappa)$ and $u_R(t;\kappa)$ provide a local estimate of the rate of the process. The first one considers only the past, while the second one considers only the future. Indeed in real time analysis it is impossible to calculate $u_R(t;\kappa)$. However, here we are interested in ex-post analysis of the historic data, so this limitation does not pose any problems. The strong sudden jump in activity (IB) at time t will dramatically increase $u_R(t;\kappa)$ and will not contribute to $u_L(t;\kappa)$, thus we can consider the difference

$$\Delta(t;\kappa) = u_R(t;\kappa) - u_L(t;\kappa) \tag{5.13}$$

in order to identify the times of jumps. An example is provided in Figure 5.1, where the function $\Delta(t,\kappa)$ is shown for different values of κ. The choice of κ will be discussed in Section 5.4.3. High values of Δ correspond to abrupt changes in the intensity of the process and thus provide a way to identify probable IB locations[1].

As candidates for the IB times $\bar{z}_j$ we select the local maxima of $\Delta(t;\kappa)$ over the set of the event times $\{t_i\}$. Moreover, we rank them from the most to the least relevant one, according to the value of Δ at the maximum. We start with the global maximum

$$\bar{z}_1 = \underset{t\in\{t_i\}}{\arg\max}\,\Delta(t;\kappa), \tag{5.14}$$

[1]For this purpose it is particularly efficient to evaluate $\Delta(t;\kappa)$ only at the event times. In fact, the recursive relations

$$\kappa\, u_L(t_i) = e^{-\frac{t_i - t_{i-1}}{\kappa}}\,(1 + \kappa\, u_L(t_{i-1}))$$

$$\kappa\, u_R(t_i) = e^{-\frac{t_{i+1} - t_i}{\kappa}}\,(1 + \kappa\, u_R(t_{i+1}))$$

hold.

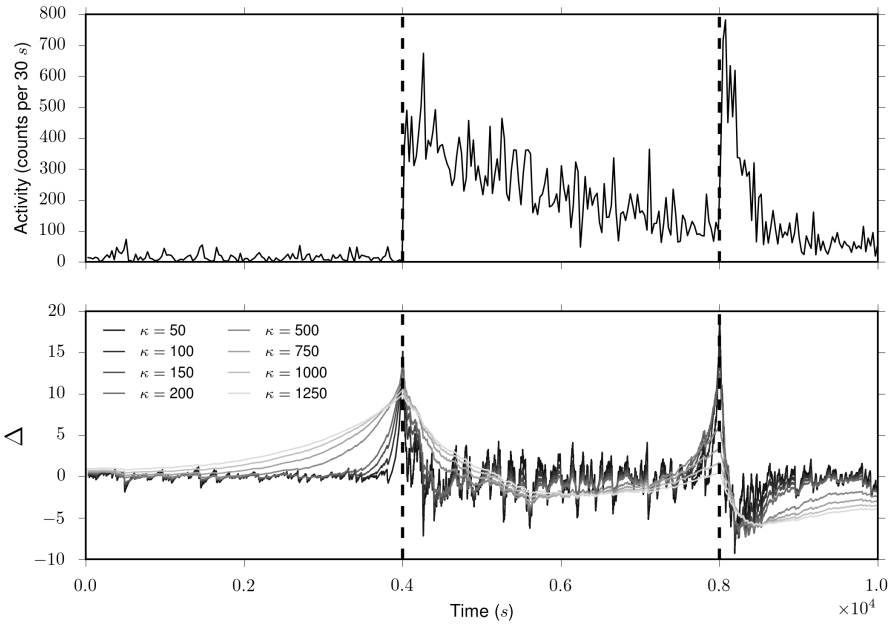

Figure 5.1: Simulation of a Hawkes process with two IBs (top panel) and the corresponding functions $\Delta(t;\kappa)$ for different choices of κ (bottom panel).

which defines a window $W_1 = [\bar{z}_1 - \frac{w}{2}, \bar{z}_1 + \frac{w}{2}]$ of size w around $\bar{z}_1$ where the optimal value of z_1 will be searched using the maximum likelihood estimator. Further we exclude all events which are closer than w to $\bar{z}_1$ and look for next local extremum:

$$\bar{z}_2 = \underset{t\in\{t_i:|t_i-\bar{z}_1|>w\}}{\arg\max} \Delta(t;\kappa), \tag{5.15}$$

which defines another window $W_2 = [\bar{z}_2 - \frac{w}{2}, \bar{z}_2 + \frac{w}{2}]$. The exclusion is important in order to select distinct maxima of the function Δ: as it is seen from Figure 5.1, the function Δ is rather persistent and if this precaution is not taken, then the second-to-best maximum will be most likely selected next to the global maximum. The procedure is repeated as many times as necessary to find the candidate times

$$\bar{z}_k = \underset{t\in\{t_i:|t_i-\bar{z}_1|>w,|t_i-\bar{z}_2|>w,\ldots,|t_i-\bar{z}_{k-1}|>w\}}{\arg\max} \Delta(t;\kappa) \tag{5.16}$$

and corresponding windows $W_k = [\bar{z}_k - \frac{w}{2}, \bar{z}_k + \frac{w}{2}]$

5.3.3 Identification of multiple IBs: estimation

Once a ranking of the windows W_j has been determined, we apply an iterative procedure to determine the optimal parameters of the model (5.2) as well as the total number M of IBs. First, we estimate the model with $M = 0$, that is a standard Hawkes model. Then, we estimate the model with $M = 1$ and $z_1 \in W_1$. At this point we use the BIC (or another model selection criterion) to decide whether to accept or reject the enlarged model. If the IB is not accepted, the procedure stops and the null model with $M = 0$ is selected. Otherwise, the model is extended to the case $M = 2$ and an additional IB term is added. The optimal value of z_2 is estimated within the window W_2, while keeping the value of z_1

fixed, and all the other parameters are optimized over again. The information criterion is again used to compare the penalized likelihood of the $M = 2$ model to that of the case $M = 1$: then the new extended version is either accepted or rejected. The iterative procedure stops when the addition of the M_{k+1} IB does not improve the likelihood of the model significantly over the model M_k.

We can summarize our complete procedure as follows:

1. Given a realization $\{t_1, t_2, \ldots, t_N\}$, we select values for κ and w.
2. We determine a ranking of the candidate IB locations $W_1, W_2, \ldots$.
3. We estimate a standard Hawkes model on the data ($M = 0$)
4. We add to the model one IB at a time
   - If $\mathrm{BIC}_{M=k+1} < \mathrm{BIC}_{M=k}$ the candidate IB z_{k+1} is added to the model and the next is examined,
   - Else if $\mathrm{BIC}_{M=k+1} > \mathrm{BIC}_{M=k}$ the procedure stops and we consider $M = k$ as the final number of IBs.

We note that this procedure relies on the ranking provided by the pre-identification procedure. Stopping criteria that rely less on the pre-identification procedure can also be considered. For example, instead of stopping at the first failure, additional IBs can be considered before exiting the loop. Again, we refer to Appendix 5.A for details on the optimization procedure.

5.4 Numerical simulations and model validation

In order to validate our procedure, we performed extensive numerical simulations. First, we examined the rate of false positives produced by our procedure. Then we tested its statistical power. Finally, we studied how discrimination works in the case where two IBs are present in the window. Since even in the case of the standard Hawkes process (5.1), bias and variance of the estimations of the parameters depend on the values of parameters themselves, we test various combinations of the parameters on various sample sizes in order to get a broad view on the efficiency of the detection procedure.

5.4.1 Absence of exogenous shocks

In order to study the rate of false positive detection we performed a Monte-Carlo simulations of a Hawkes process (5.1) with the endogenous kernel (5.5) and without any IB. The parameters are $p = 2.0$, $\tau_0 = 0.1$, and $n = \{0.3, 0.5, 0.7, 0.9\}$. We simulate 1000 realizations for each parameter set. Because the power of the statistical test strongly depends on the sample size, in our simulations we keep the expected sample size $\mathbb{E}[N] = \mu T/(1-n)$ constant by fixing simulation horizon $T = 3,600$ and varying the baseline intensity μ. We perform tests with sample sizes of approximately 1,000, 2,000, 5,000, and 10,000 events.

In each simulation we apply our procedure and check whether or not the extended model (5.7) is preferred over the standard Hawkes model. In general we found that AIC performs much worse than BIC, especially for high values of n. The rates of false positives, expressed in percent, resulting from the BIC are summarized in Table 5.1. As it is seen from the table, even for moderate sample sizes (1000 events) and very strong self-excitation of the underlying process ($n \sim 0.9$) the rate of false positives is around 1%, and in general the

	n			
size	0.3	0.5	0.7	0.9
1000	0.7	0.5	1.3	1.0
2000	0.4	0.4	0.7	0.6
5000	0.0	0.3	0.2	0.2
10000	0.1	0.0	0.3	0.2

Table 5.1: Percentage of false positives using Bayesian Information Criterion when the simulated model has no IBs. All values expressed in percent (%).

specificity of the method is close to 100%. This is further illustrated by the Figure 5.2 that presents the distribution of the observed scores ΔBIC.

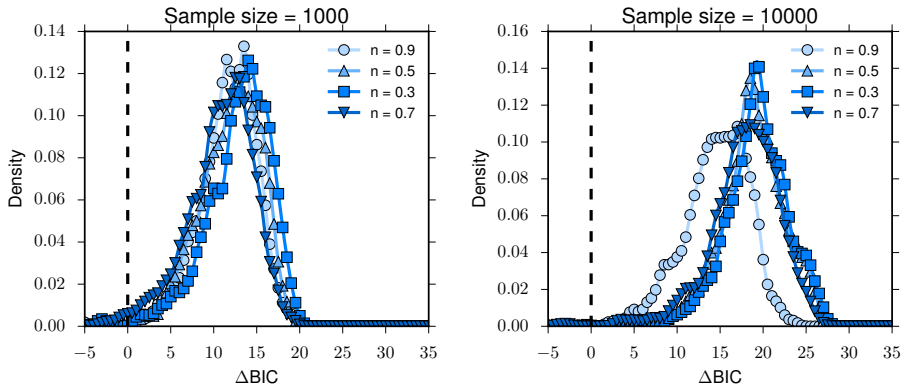

Figure 5.2: Distribution of the differences ΔBIC between the null model where no IB term is present and the complete model with a single IB estimated on simulations where no IB is present. For differences smaller than zero the null model is rejected.

5.4.2 Single IB with approximately known location

We now consider numerical simulations when one IB is present. Here we present results of simulations with parameters $p = 2.0$, $\tau_0 = 0.1$, $n = \{0.3, 0.5, 0.7, 0.9\}$. The simulation window is $[0, T]$ where $T = 3,600$. An IB is simulated at $z = T/2$, and 35 different combinations of IB parameters (f, τ) are considered (see below for details). As before, we vary the value of μ in order to keep the expected sample size

$$\frac{\mu + f(1 - e^{\frac{z-T}{\tau}})}{1-n} T \tag{5.17}$$

approximately constant. We performed two sets of simulations with average sample sizes 5,000 and 10,000 respectively. Since the primary concern of this experiment is the rate of false negatives resulting from our procedure and our choice of the BIC selection criterion, here we ignore the pre-identification routine and we will address the complete procedure in the next sections. Thus we limit the search space for z to a window of size 100 centered around the true position of the IB: $[\frac{T}{2} - 50, \frac{T}{2} + 50]$.

We report the results of the test in Appendix 5.B. Table 5.14 presents the percentage of correctly detected IBs for different values of the branching ratio n for the sample size of 5,000. Our results show that there is a region corresponding to low values of the fertility f and high values of the relaxation time τ of the IB (i.e. small values of $\alpha = f/\tau$), where

the addition of the IB term does not contribute significantly to the improvement of the likelihood and therefore the extended model is rejected.

Hawkes processes naturally produce clusters of events through the self-exciting mechanism. The average size of a cluster is given by $\frac{1}{1-n}$ (see e.g. Møller and Rasmussen (2005)) and its variance is $\frac{n}{(1-n)^3}$. Hence, if an exogenous IB produces an events cascade of similar size to the ones generated endogenously, then our procedure will not identify it as an IB. This is actually a desired feature since we want to identify IBs that are not compatible with the endogenous dynamics. Moreover, this also suggests that for high values of the branching ratio n it becomes more difficult to distinguish between an endogenously generated burst of events and an exogenous one. At the same time, the total number of events triggered by the IB (5.4) is an increasing function of n, and this explains why the detectable region is bigger for higher values of n.

5.4.3 Pre-identification

The efficiency of the pre-identification algorithm depends on the choice of the "smoothing parameter" κ. When κ is small, the approximation of the local intensity will be noisy and will result in detection of many local maxima. On the other hand, large values of κ result in averaging out most of the intensity fluctuations and leaving only big IBs to stand out, but potentially missing medium scale ones.

In order to test the effect of the choice of κ, we simulate the model (5.7) with different combinations of the parameters (α, τ) and then apply our pre-identification algorithm with various values of κ. Results are reported in Appendix 5.B, where Figure 5.11 presents the Root Mean Squared Error on the detected IB location relative to the average distance between events $\delta = T/N$ as a function of the ratio κ/τ.

Naturally, for small values of α, when detection of the exogenous IB is difficult, error of the estimation of z could be very large. However, we observe that for relatively high values of α the error drops to the value comparable with inter-event interval δ, and the best performance is attained when κ is of the same order of magnitude of τ. When κ is much larger than τ the performance deteriorates also for high values of α. This suggests that if the typical relaxation time of the IB is known then the best choice is to select κ to be of the same order. Alternative but more complicated way could be to employ the multi-scale analysis for multiple values of κ in a spirit of the original paper (Almgren, 2012). However, given that (i) our method is parametric and pre-identification is followed by the maximum likelihood estimation and (ii) due to stochastic nature of the Hawkes process (5.2), the peak of local intensity does not necessarily coincides with the start time z and can be delayed depending on the kernel $\phi(t)$, we consider the multi-scale approach to be an over-complication.

5.4.4 Single IB detection with the complete procedure

We now test our complete procedure in the case where one IB is present. We follow the same procedure as in Section 5.4.2. However, now we do not impose the search space for z, applying the pre-identification method instead. We set $\kappa = 100$ and we set the search window size in (5.14)–(5.16) to $w = 300$, i.e. while doing maximum likelihood estimation, we perform a constrained search for z_j in windows $W_j = [\bar{z} - 150, \bar{z} + 150]$. We consider the actual IB to be correctly identified by our procedure when one of the detected IBs is within a tolerance of 60 time units from its occurrence time. This includes also the cases where the actual IB is not the first one to be detected.

Table 5.2 reports the percentage of correctly classified IBs (true positives) for the different combinations of n, $f = \alpha\tau$, and τ. The performance of the method is shown to be comparable with the case where the correct search interval for z was known in advance (compare with Table 5.14), indicating that the pre-identification algorithm works fairly well in most cases. The lower rate of correct detections are found unsurprisingly for high values of τ and small values of f, i.e. small values of α. In these cases, the immediate shock to the intensity by the IB is smaller and the cumulative effect of the IB is more diluted over time, which makes it harder to localize.

$n = 0.3$						$n = 0.5$					
f	τ					f	τ				
	10	50	100	500	1000		10	50	100	500	1000
50	76	16	4	0	0	50	69	17	4	0	0
75	94	65	27	0	0	75	95	74	30	0	0
100	99	96	71	2	0	100	100	98	79	0	0
250	100	100	100	72	12	250	100	100	100	78	16
500	100	100	100	100	88	500	100	100	100	100	89
750	100	100	100	100	100	750	100	100	100	100	98
1000	100	100	100	100	100	1000	100	100	100	100	100

$n = 0.7$						$n = 0.9$					
f	τ					f	τ				
	10	50	100	500	1000		10	50	100	500	1000
50	62	35	13	0	0	50	9	28	18	2	0
75	97	86	50	0	0	75	26	64	56	5	0
100	96	98	84	4	0	100	55	83	79	15	2
250	100	100	100	88	32	250	99	100	100	72	39
500	100	100	100	100	89	500	97	100	99	98	78
750	100	99	100	100	98	750	100	95	97	96	90
1000	98	99	99	98	100	1000	96	98	89	100	97

Table 5.2: Percentage of correctly classified IBs for different combinations of the true IB parameters (α, τ) expressed in terms of $f = \alpha\tau$ and τ. The results refer to a sample size of roughly 5,000 events.

Further we need to explore the rate of false positives, i.e. the cases when more than one IB is detected in these simulations. Table 5.3 reports the number and the percentage of false positive identifications as a function of the branching ratio n. As it is seen, the number of cases where more than one IB is detected is however limited and concerns mainly the simulation of high values of n: $n = 0.7$ and especially $n = 0.9$, which is expected, since in these cases fluctuations of the intensity of the background Hawkes process are very strong and the probability of spontaneously generating a cluster which might me classified as IB is high. So reliable detection here requires sufficiently long time series, which ensures the efficient estimation of parameters of the Hawkes process when n is high.

Next, we examine the error on the estimation of the IB starting time z. In Table 5.4 we report the ratio between the root mean squared error $RMSE[(z - \hat{z})^2]$ and the average inter event time $\delta = T/N$. As before, the results are much better when the parameter $\alpha = f/\tau$ is large, while the performance deteriorates for small α and comparatively high τ albeit still remaining satisfactory.

In our procedure, we use the pre-identification algorithm discussed above to come up with a first guess of the IB location z_g. It is worth comparing this initial guess with the

	Sample size = 5000			
	$n = 0.3$	$n = 0.5$	$n = 0.7$	$n = 0.9$
Total FP	1	3	12	63
Worst case incidence (%)	0	0	2	18
	Sample size = 10000			
Total FP	0	0	9	37
Worst case incidence (%)	0	0	4	14

Table 5.3: Total number of false positives and percentage of false positives in the worst case resulting from our procedure applied on simulations where one IB is present. The worst case for a given value of n is the single combination (f, τ) where the most false positives are recorded.

final estimate $\hat{z}$ provided by our procedure. We consider the differences:

$$\frac{|z_g - z| - |\hat{z} - z|}{\delta}$$

where z represents the true IB location, for the correctly classified IB. We observe that in 91% of the cases the final estimate is not worse than the pre-estimaed guess (i.e. as close or closer to the true value). Moreover, when our method improves on the pre-identification method, the improvement is on average more than twice the average error when our method does not improve. In other words, the maximum likelihood optimization of the location parameter z appears to be worthwhile to do, even in such cases when the exogenous kernel (5.6) is strictly decaying and has maximum at (which leads to the expected peak of the intensity to be at $t = z$). In more complicated case when the peak of the reaction is delayed the optimization step will be essential.

In Appendix 5.C, Tables 5.17, 5.18, 5.19, and 5.20 report the relative mean squared errors on the IB parameters. For IBs in the detectable region the MSE is typically of the order of 10% for considered sample sizes.

Finally, it is worth mentioning that our procedure can be useful even in cases when the detection of the IB is not the main objective, but one is interested in the parameters of the base Hawkes model, such as the branching ratio n. It is known that the presence of non-stationarities and bursts dramatically inflate the branching ratio estimated under the assumption of stationarity (i.e. within the model (5.1)). This problem has been discussed already in Filimonov and Sornette (2015) and Rambaldi et al. (2015) and here we illustrate it with the Table 5.15 where we compare the values of n obtained from the base Hawkes model and from the best model selected by our procedure when one IB is present. Indeed in case of an intensity burst, the base model overestimates the branching ratio up to the critical level of 1 even for small values of the actual branching ratio ($n = 0.3$). In contrast, properly accounting for such exogenous event via the extended model (5.7) allows us to recover the correct value in all cases.

5.4.5 Iterative procedure of IB detection

In order to assess the performance of our procedure in the case when multiple IBs are present, we examine eight scenarios where two IBs occur in the same $T = 3,600$ time window. Specifically, we consider two kinds of IBs, one that we name "Small" (S) with parameters $(\alpha = 1.0, \tau = 350)$ and another that we name "Large" (L) with parameters $(\alpha = 1.5, \tau = 700)$. The distance between the IBs is also relevant, thus we test separately a configuration where the IBs are close (C), namely $z_2 - z_1 = 350$ units of time apart, and

$n = 0.3$						$n = 0.5$					
f	τ					f	τ				
	10	50	100	500	1000		10	50	100	500	1000
50	3.1					50	1.9				
75	0.6	15.4				75	0.7	12.3			
100	0.5	5.3	21.0			100	0.5	5.2	16.1		
250	0.2	1.3	3.0	22.8		250	0.1	0.9	2.5	26.0	
500	0.1	0.4	1.1	9.9	26.2	500	0.1	0.5	1.2	7.8	29.8
750	0.0	0.3	0.6	6.6	18.7	750	0.0	0.2	0.6	5.1	15.0
1000	0.0	0.2	0.3	4.4	13.0	1000	0.0	0.1	0.5	2.9	7.9

$n = 0.7$						$n = 0.9$					
f	τ					f	τ				
	10	50	100	500	1000		10	50	100	500	1000
50	1.6					50					
75	1.2	7.1	17.1			75		5.6	14.6		
100	0.4	3.5	10.7			100	0.5	2.9	13.4		
250	0.2	1.0	2.6	27.9		250	0.3	0.9	3.9	23.0	
500	0.1	0.4	0.6	8.5	17.0	500	0.1	1.0	0.8	9.9	20.0
750	0.0	0.3	0.5	3.6	13.7	750	0.0	0.4	0.8	4.4	16.1
1000	0.0	0.2	0.3	1.9	4.5	1000	0.0	0.4	1.0	8.1	10.4

Table 5.4: Ratio between the root mean squared error on the IB location parameter z and the average inter-event time T/N. The results shown are for a sample size of roughly 5,000 events. The error is computed only for cases where the IB is detected at least on 50 out of 100 simulations.

a second one where the IBs are far (F) apart, namely $z_2 - z_1 = 1,400$ units. These eight scenarios are summarized in Table 5.5.

We simulate 100 realizations for each scenario, for four different values of the endogenous parameter n, namely $n = (0.3, 0.5, 0.7, 0.9)$; other endogenous parameters are fixed as in the other experiments to $\tau_0 = 0.1$ and $p = 2.0$ and parameter μ is adjusted so to keep the expected sample size equal to 10,000 events. Table 5.6 presents the result of this test for the case $n = 0.7$, while the other cases are reported in Table 5.16.

Our results show that for intermediate values of n, our procedure attains very good results. As it is seen from the Table 5.6, in most of situations rate of true positives exceed 90–95% with a reasonably low numbers of false positive and false negative detections. The acceptance of more than two IBs in the test (false positive detection) has a maximum value below 10% and affecting almost exclusively the $n = 0.9$ case. And as expected, the probability of false negative increases with the decrease of size of the IB.

The most challenging scenarios for the identification are unsurprisingly those where a small IB follows a bigger one and thus gets overshadowed by the latter. This is especially pronounced when the IBs are close (CLS): in these cases the true positive rate is about 50% in case of $n = 0.7$, and similar in the case when small IB is followed by the small IB (CSS). Further, large values of the branching ratio are more difficult to handle, as we have already noted: when n approaches the critical value of 1, the variance of the cluster size becomes very large, thus making it extremely difficult to differentiate between an endogenously generated burst in intensity and an exogenously generated one.

The relative mean squared error on the IB parameters is generally under 10% for $n = (0.3, 0.5, 0.7)$. Errors on τ tend to be larger than those on α, especially on small IBs

	α_1	τ_1	z_1	α_2	τ_2	z_2
CSS	1.0	350	1625	1.0	350	1975
CSL	1.0	350	1625	1.5	700	1975
CLS	1.5	700	1625	1.0	350	1975
CLL	1.5	700	1625	1.5	700	1975
FSS	1.0	350	1100	1.0	350	2500
FSL	1.0	350	1100	1.5	700	2500
FLS	1.5	700	1100	1.0	350	2500
FLL	1.5	700	1100	1.5	700	2500

Table 5.5: Summary of the eight tested scenarios for cases with two IBs. C stands for close, F for far, L for large, and S for small (see text for details).

	$n = 0.7$							
	CLL	CLS	CSL	CSS	FLL	FLS	FSL	FSS
First correct	100	100	97	100	98	98	91	91
Second correct	96	50	100	56	98	90	98	90
Both correct	96	50	97	56	98	90	91	88
No shock detected	0	0	0	0	2	2	2	6
More than two	1	0	0	1	0	0	0	0

Table 5.6: Results of the tests on simulation with two IBs for the $n = 0.7$ case. All quantities are expressed in percent. Parameters of simulations are presented in Table 5.5.

following a large one where error on τ can be above 60%. Errors rise significantly for the case of $n = 0.9$, where values of more than 30% on α are not uncommon.

From the above tests we have validated our procedure and confirmed that the method is capable of obtaining meaningful results, provided that the underlying process is not too close to being critical (i.e. $n = \int_0^\infty \phi(t)\mathrm{d}t$ is sufficiently smaller than one). Naturally the quality of detection and estimation of parameters increases with the increase of amplitude and distance between shocks.

5.4.6 Misspecification of the endogenous kernel

Finally, we address an important practical aspect of the model error. In real life we do not know a priori what functional dependence of the kernels is the best to describe the dynamics of the system. As it was mentioned above, the discussion about the specification of the kernel of the Hawkes model (5.1) is still open, and for example in financial applications along with approximate power law kernel (5.5) a short-memoried exponential alternatives are considered and advocated. The question of the statistical tests for selecting the best candidate as well as the question of non-parametric estimation of the kernel are well beyond the scope of the present work. Here we will address this problem by considering the estimation error in case when the memory kernel is wrongly specified.

For this we simulate the process with a different endogenous kernel (namely — exponential and double exponential), while performing estimations using the same methodology as before that uses the $\phi(t)$ in the form of 5.5. We conduced both experiments where no IBs is present and where one IB is present. Detailed results are presented in Appendix 5.D. Overall our results are in line with those of Section 5.4.2 and 5.4.4, and show that our

procedure is fairly robust with respect to the kernel misspecification: as long as the exact parameters of the underlying Hawkes process is not in the main focus of attention, the exogenous IBs could be identified reliably even when the background process is not well specified. This is an important argument, as it relaxes the efforts needed for the practical application of the procedure.

5.5 Application to high-frequency FX data

We apply our methodology to study intensity bursts in the spot Foreign Exchange (FX) markets. Currency market is a complex decentralized system of trading platforms and venues, with co-existing Electronic Communication Networks (ECNs), that operate in a similar way to regular stock exchanges, and OTC dealing over the terminal chats or even voice calls. At the center of the market system there exist two inter-dealer electronic platforms: Electronic Broking Services (EBS) and Reuters. They are used by major banks and brokers as a source of interbank liquidity and also as a trading platform for large HFT players. Both platforms have a high requirement for the lot size to be at least one million units. Smaller market participants and retail clients operate on alternative trading venues, such as ECNs, or via brokers. EBS is the main venue for all the USD, EUR, CHF and JPY crosses, while Reuters is a main platform for trading of all the crosses for Commonwealth currencies and Scandinavian currencies (Golub et al., 2013).

5.5.1 Descriptive statistics

In our work we will analyze tick data from the EBS Live data feed for three currency pairs, namely EURUSD, EURJPY and USDJPY. EURUSD is by far the most liquid currency pair in the world with the average daily turnover of 1,289 billions of USD in 2013 according to the BIS triennial survey (Bank for International Settlements, 2013), with the second one being USDJPY with the turnover of 978 billions. EURJPY is also among the most actively traded contracts with the average daily turnover of 147 billions of USD (having only USDGBP, USDAUD, USDCAD and USDCHF before).

Our dataset spans the period from January 1, 2012 to December 18, 2012. FX markets are active 24 hours a day and in principle operate also 7 days a week; activity during weekends is however negligible. EBS Live is a premium feed which provides snapshot of the order book every 100 milliseconds (in contrast to a regular feed with quotes every 250 ms). We construct the point process of the events when either of best bid or best ask price changes. The dynamics of such process has been used as a proxy of high frequency volatility and has been modeled with Hawkes processes by several papers (Filimonov and Sornette (2012); Hardiman et al. (2013); Rambaldi et al. (2015)). The limitation of the time resolution of 100ms and absence of records within these sub-intervals might cause bias in the calibration of the point process. The implications of such coarse time resolution on the Hawkes process fitting was discussed in (Lallouache and Challet, 2016). To address this issue we use the same randomization procedure of (Filimonov and Sornette, 2012; Rambaldi et al., 2015), namely we subtract from each timestamp a random number uniformly distributed in the interval $[0, 0.1)s$.

Our proposed methodology is explicitly designed to detect and model localized "shock-like" non-stationarities (IBs) within the Hawkes process framework. However, the methodology is susceptible to other forms of non-stationarity, such as regime shifts or slow changes in base intensity (e.g. day-to-day fluctuations or time-of-day effects). In order to limit the side effects from such dynamics, we limit the window size of the analysis to one hour.

	# shocks	Shocks per window	Fraction no shock	Max shocks
EURJPY	792	0.51	0.61	5
EURUSD	720	0.46	0.63	4
USDJPY	558	0.36	0.70	3

Table 5.7: Total number of IBs detected for the three currency pairs, average number of IBs per window, fraction of windows with no IB and maximum number of IBs detected in a single window.

Specifically, from the best quote change time series, we extract windows of one hour starting at 00:20:00 UTC of Jan 01 2012. We keep only those windows where a sufficient number of events is present, namely those with at least 2,000 events. Furthermore, since we want to compare results across the three pairs, we retain only windows for which enough events are present in all pairs. This leaves us with 1,551 windows for each currency pair and in the following we will always refer to this subset of the original data. On average, we have 6,394 events per window in the case of the EURUSD pair and 4,169 and 3,187 for EURJPY and USDJPY cases respectively. We apply our complete procedure using $\kappa = 100$ and $w = 300$.

The total number of IBs detected for each currency pair is summarized in Table 5.7. On average we find about 0.3–0.5 IBs per window. We note however that even by selecting small windows on just one hour, we detect a certain number of IBs with very long (much larger than one hour) decay times τ. Most of them are found in the intervals 06:30 - 08:00 and 12:00 - 15:00. The first interval corresponds to the beginning of the European session on the FX market (07:00 London time) and to the opening of European stock markets (08:00 London time). The second one instead is when the American session begins (12:00 London time) and US stock markets open (14:30 London Time)[2]. Thus, these shocks appear to capture mostly time-of-day effects, so we have decided to remove them from the analysis. Specifically, we disregard bursts with $\tau > 5400s$, since the decay time is much longer than the detection time (this amounts to less than 10% of detected IBs).

Figure 5.3 presents the distribution of the estimated IB parameters α and τ, together with the corresponding fertility f and fertility normalized by the total number of events in the estimation window (f/N). As it is seen, the parameters ranges have very broad ranges reflecting the diversity of the shocks that we encounter. The shapes of the distributions for all currency pairs are similar. It is interesting to note, that the distributions of normalized fertilities are remarkably close among all currency pairs, as well as the distributions of the relaxation time τ. Here one can read the typical values of 200–400s for τ and 1%-4% of the total sample size (N) for f.

5.5.2 Simultaneous bursts in several markets

Here we consider IBs that are common between currency pairs, which we defined as events whose estimated starting time z is located within a tolerance window of $60s$. If we consider simultaneous shocks in all three markets at the same time, they will constitute 15–20% of all detected IBs. However, this figure increases significantly if we consider only two pairs as presented in Table 5.8. Since we perform the calibration of the model independently on

[2]Note also that daylight saving times are not synchronized between Europe and North America so there are some weeks of the year when US markets open at 13:30 London time.

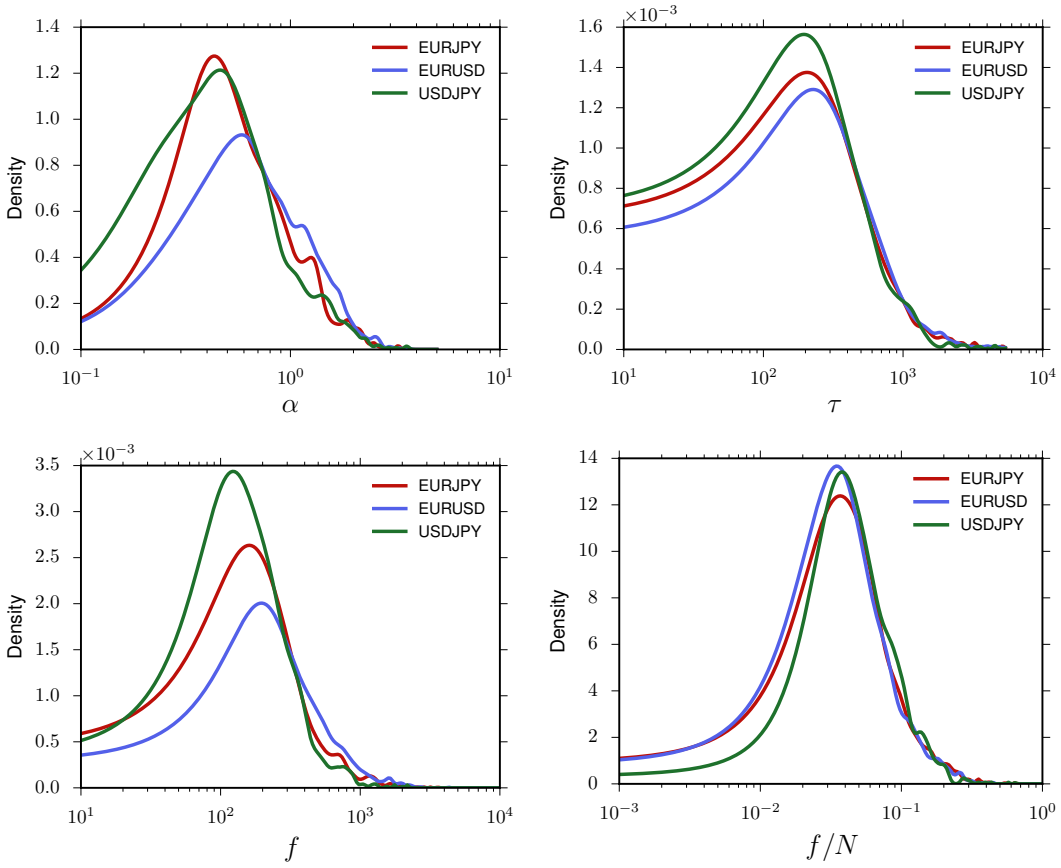

Figure 5.3: Density of the estimated IB parameters across the three currency pairs.

	# common	% pair 1	% pair 2
EURJPY/USDJPY	252	31.82	45.16
EURUSD/EURJPY	324	45.00	40.91
EURUSD/USDJPY	161	22.36	28.85

Table 5.8: Common IBs between two currency pairs: total number and a fraction of the detected IBs for each pair.

all three FX pairs[3], the detection of a large number of common IBs could be considered as a sanity check for the procedure and an indirect evidence that the detected IBs are genuine events. Indeed, in our analysis any two FX rates share one currency in common, hence it is natural to expect that they would react to common drivers. Moreover, all three FX pairs are linked in a triangular fashion and thus are attractive for high-frequency traders who arbitrage away a possible transient mispricing between rates.

It is interesting to look at the difference in the estimated initial times z in the case of common IBs. In Figure 5.4 we plot the histogram of the time delay of the IBs in EURJPY and USDJPY relative to the shock in EURUSD in case when all three pairs have a common shock. We observe that the distribution is quite broad; however, in both cases the peak of the density is found between 0 and $500ms$. This suggests that many times EURUSD is the leading pair for an intensity burst, which seems reasonable since it

[3] A small bias however could be introduced by the fact that we perform our analysis on a set of windows with at least 2,000 events in all three pairs.

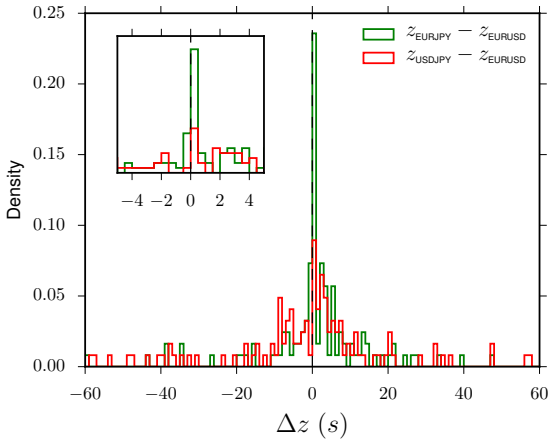

Figure 5.4: Histogram of the differences in detected IB times in the three rates for IBs common to all the three pairs. The time z_{EURUSD} is used as difference. The bin size is 1s in the main plot and $500ms$ in the inset.

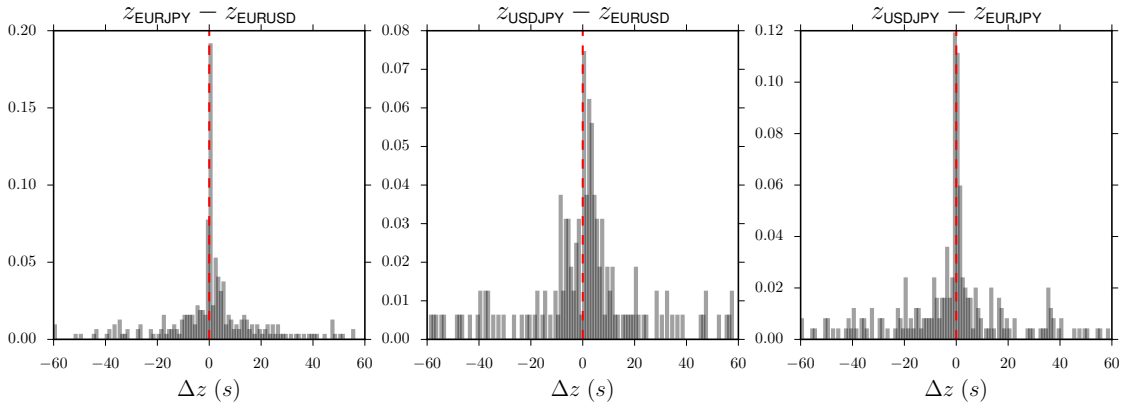

Figure 5.5: Histogram of the differences in detected IB times in two rates for IBs common to two pairs. The bin size is $1s$.

is by far the most liquid and important currency cross. Performing a bootstrap test on the one-sided t-statistic under the null $z_{xxxyyy} - z_{EURUSD} <= 0$ we obtain t-statisitics (p-values) 0.59 (0.28) and -0.72 (0.77) for EURJPY and USDJPY respectively. If we focus on close matches with absolute difference smaller than 5 seconds, we obtain t-statistics (p-values) 2.07 (0.02) and 2.24 (0.02) respectively. In Figure 5.5 we plot analogous histograms for the common IBs between two pairs. Again, we note that EURUSD appears to lead both the other crosses. Instead, a clear leader does not emerge between EURJPY and USDJPY.

Figure 5.6 presents the distribution of the normalized fertilities f/N for IBs that are (1) common to all three pairs, (2) common to only two pairs, and (3) detected in a single pair only. While the distributions in the cases (1) and (2) are very close to each other, one can see that the distribution for the third case has much heavier right tail implying that the IBs common to all three pairs have markedly higher fertility. This is confirmed by t-test results in Table 5.9, while the difference between idiosyncratic IBs and those common to exactly two pairs is not as significant. At the same time Kolmogorov-Smirnov tests are significant at the 1% level in all cases (including the test for the equality of distributions of (1) and (2)).

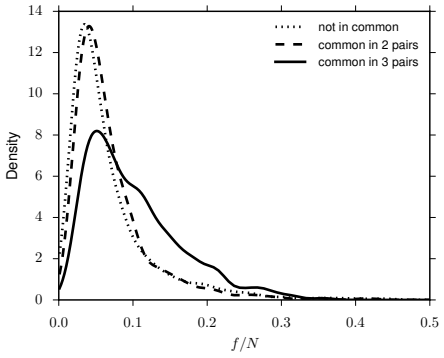

Figure 5.6: Distribution of the ratio f/N for IBs detected simultaneously on all three pairs (solid line) on exactly two pairs (dashed line) and for idiosyncratic IBs (dotted line).

	t-statistic	p-value	KS-statistic	p-value
2 vs 1	0.66	0.51	0.094	0.0011
3 vs 1	8.1	2e-15	0.3	1.3e-21
3 vs 2	7.3	8.1e-13	0.27	2.1e-16

Table 5.9: Results of Welch's t-test for equality of the means and Kolmogorov-Smirnov test for the null hypothesis that the samples come from the same distribution for fertility values of IBs common to all three pairs, to only two pairs and found in a single pair.

5.5.3 Intensity bursts and scheduled announcements

FX markets are usually very sensitive to economic announcements such as interest rate decisions, reports on the inflation, GDP and employment data, etc. Most of these announcements are scheduled in advance and market participants prepare for them to account for the uncertainty in the reported numbers. Here we study how the activity responds to such announcements.

We use a list of scheduled economic announcements retrieved from the website *www.dailyfx.com*. The dataset consists of economic data releases such as inflation, GDP and employment readings, as well as rate decisions communications and press conferences by central banks, for a total of 3,352 separate events. We focus on those intensity bursts that occur within $60s$ interval of a planned announcement time in the dataset. Such window size is a good proxy in most of cases, though not always: for example the FOMC rate decision is reported orally by the Chair of the Board of Governors of the Federal Reserve System, and the actual figure of the target interest rate can be pronounced several minutes after the start of the press conference (see for example the reaction time for the fixed income markets in Table 3 of Almgren (2012)).

In Table 5.10 we report the descriptive statistics of the IBs corresponding to scheduled announcements. We find that about 10% of the announcements correspond to a detected IB and conversely about 20% of the detected IBs are related to an economic announcement. The database of *www.dailyfx.com* also provides a suggestion for the currency most affected by the announcement and an indication of the typical importance of that announcement in a three level scale (Low, Medium, High). Using these indications, we unsurprisingly find that high-importance announcements have a markedly higher detection rate of above 25%, while figures for low-importance ones are around 2%.

Within detected IBs that match scheduled announcements, we note the prevalence of announcements related to the US economy. To check for over- and under- representation

	# matches	% news detected	% IBs related to news
EURUSD	118	10.35	16.39
EURJPY	137	12.02	17.30
USDJPY	121	10.61	21.68

Table 5.10: Number of detected IBs that match a scheduled announcement within a $60s$ interval. The fraction of announcements in the database that triggered an IB according to our procedure is also reported.

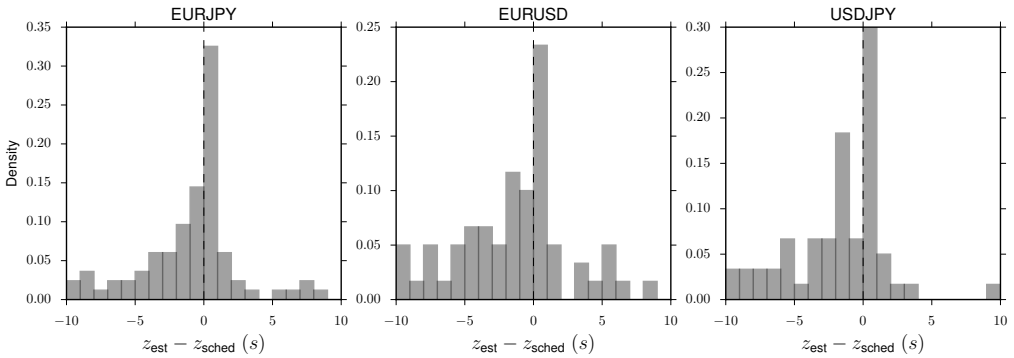

Figure 5.7: Histogram of the time differences between the detected IB time resulting from our procedure and the scheduled release time of economic announcements for which a match was found.

of the subpopulations we performed an hypergeometric test. We find that USD-related announcements are heavily over represented compared to a random matching (p-values lower than 10^{-12}), while all the other geographic origins are under represented. We remark that this test does not take into account the time-of-day distribution of the announcements. For example, most announcements concerning JPY are released between 23:00 and 8:00 London time, while our test assumes the same probabilities irrespectively of the time of day. It appears nevertheless that, at least for these currency pairs, US-related economic announcements are far more likely to result in an IB compared to announcements on other countries/currency.

It is interesting to note that the shock in market activity (an intensity burst) does not always follow the scheduled announcement, but sometimes precede it. Figure 5.7 shows the distribution of the time difference between the estimated IB starting time z and the scheduled release time for the announcement for which a match was identified. We observe the peak of the density within a first second after the announce time with a few cases of lags more than several cases. However, there exist a substantial number of IBs that have started well before the announcement time, indicating the preparation of the market participants (presumably, market makers specifically) for the upcoming release. In the following section we will observe such anticipation in the dynamics of the spread that reflect the risk perception of the market makers.

5.5.4 Intensity bursts and spread dynamics

Bid-ask spread is an important metric of the trading costs and the liquidity of the market. Being set up usually by a market makers it often reflects their anticipation of the uncertainty in price moves. FX traders know that right before an important announcement liquidity often "evaporates" from the order book leaving a wide spread and a small

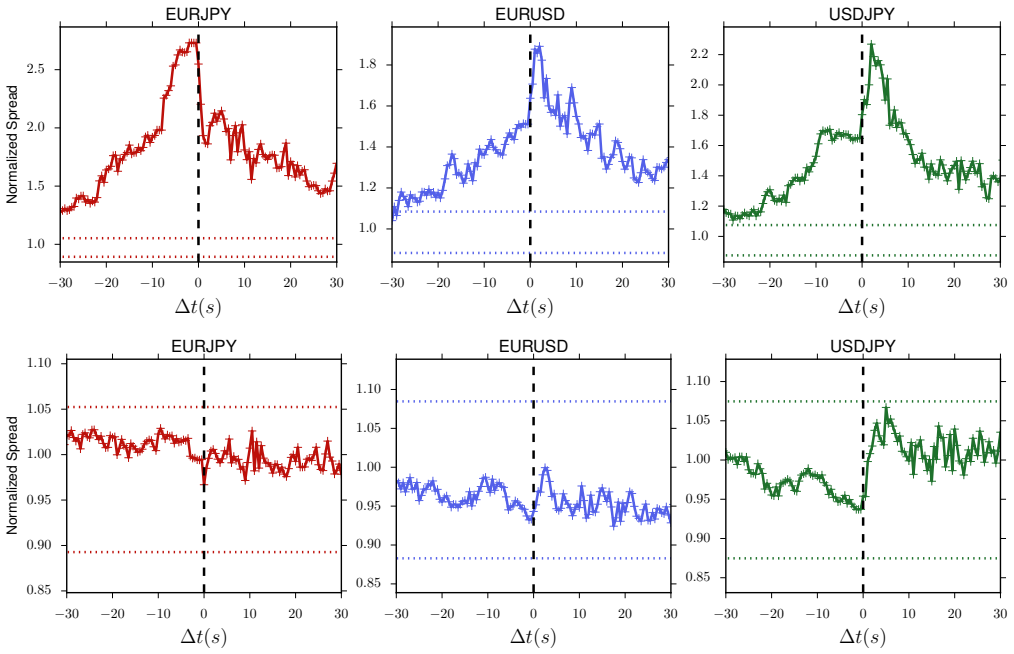

Figure 5.8: Dynamics of the average spread around the detected IBs. Detected IBs matching (not matching) a scheduled announcement are considered in the top (bottom) panels. The dotted lines represents 5 and 95 percentiles obtained on 5000 random intervals which are at least 15 min far from an IB.

amount of limit orders in the first levels, as no one wants to be affected by the possible adverse price move. Here we provide a quantitative analysis for such cases.

We consider the bid-ask spread dynamics in the windows of 2 minutes before and after the detected IBs (i.e. in window $[z - 120, z + 120]$). In order to aggregate data from a different periods with potentially different market conditions we normalize the spread values by the average value of the spread in a 3 minutes window ending 5 minutes before the detected IB (i.e. $[z - 480, z - 300]$). Taking all the normalized spread series for each FX pair, we finally compute an evenly spaced time series by averaging the values in $500ms$ bins.

Figure 5.8 presents the average spread dynamics separately for IBs that match scheduled announcements (top panel) and IBs that do not (bottom panel). We see that in case of announcements, the spread is persistently increasing towards the detected starting time of IB (as it was discussed in previous section, often this time lies within a few seconds around the announcement times), and sharply drops right after with a following decrease. Such dynamics is rather long-lasting (of the order of tens of seconds) and appear to be statistically significantly different compared to the spread dynamics without any announcements and shocks (plotted in dashed line in Figure 5.8).

In contrast, the spread dynamics in case of unexpected IB does not appear to be significantly different from the average dynamics obtained on random intervals. We can though note that a few seconds right before the IB the spread tends to drop and widen right after the IB. The most pronounced this pattern is in the USDJPY currency pair. This observation suggests that market is not able to anticipate the arrival of these type of IBs.

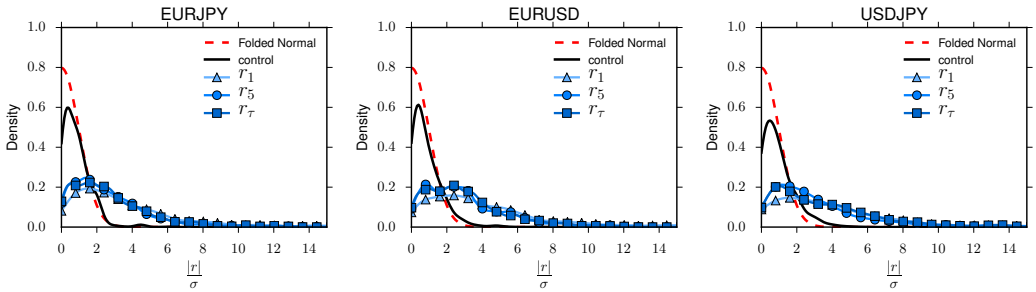

Figure 5.9: Kernel density estimation of the distribution of the absolute normalized returns $\frac{|r_i|}{\sigma_{\text{loc}}}$ during an IB. For comparison also the folded Normal distribution and the empirical distribution from a control group of 1,000 randomly chosen 1 minute returns are shown.

	EURJPY			EURUSD			USDJPY		
	mean	std	kurt	mean	std	kurt	mean	std	kurt
r_1	3.97	4.50	19.60	4.38	4.45	22.80	5.26	8.36	28.74
r_5	2.72	2.50	18.95	3.10	2.59	6.26	3.67	4.62	22.91
r_τ	2.78	2.48	16.64	3.08	2.41	3.87	3.87	4.27	16.94
folded normal	0.80	0.60	0.87	0.80	0.60	0.87	0.80	0.60	0.87

Table 5.11: Mean, standard deviation, and excess kurtosis of the empirical distributions of the absolute normalized returns $\frac{|r_i|}{\sigma_{\text{loc}}}$ during an IB. Theoretical values of the standard folded normal distribution are also reported for comparison.

5.5.5 Intensity bursts and price jumps

The second explanation of IBs is related to price jumps, since these could provide a possible mechanism that triggers IBs, particularly when the price jump is due to or associated with lack of liquidity. Vice versa the IBs might be a consequence of the activity of market-makers searching for a stable price level under the uncertainty of the future price moves, and a price shock or an established trend could be a resolution of such uncertainty.

To motivate the analysis of the relation between IBs and large price movements, in Figure 5.9 we plot the density estimation of the distribution of the absolute normalized returns associated with detected IBs. We consider three time scales for returns, one minute, five minutes, and τ minutes (see below for details). The corresponding empirical means and standard deviations as well as excess kurtosis are reported in Table 5.11. We also compare it with a control distribution obtained from the random sample. It is clear that the conditional distribution has much heavier tail with a large probabilities of extreme returns. The main conclusion is that IBs are often statistically associated with large price movements, thus it is important to consider the relation between IBs and jumps.

There is a vast econometric literature on price jumps and their detection (Andersen et al. (2007); Lee and Mykland (2008); Bollerslev et al. (2009); Aït-Sahalia et al. (2009) to name only a few), we refer to Bormetti et al. (2015) for literature survey and discussion about modeling co-jumps within the Hawkes process framework. Here, we are mostly interested in intraday abnormal returns on the same lines as Lee and Mykland (2008) and Joulin et al. (2008). More specifically we consider the series of absolute midprice log-returns $|r_i| = |\log(P_i/P_{i-1})|$ on a given time scale and we say that the ith return r_i is a $\theta\sigma$-jump

	EURJPY			EURUSD			USDJPY		
	$\theta = 3$	$\theta = 4$	$\theta = 5$	$\theta = 3$	$\theta = 4$	$\theta = 5$	$\theta = 3$	$\theta = 4$	$\theta = 5$
r_1	0.46	0.33	0.24	0.54	0.40	0.31	0.53	0.40	0.29
r_5	0.34	0.20	0.12	0.41	0.26	0.18	0.40	0.28	0.20
r_τ	0.35	0.23	0.13	0.43	0.26	0.17	0.48	0.37	0.25

Table 5.12: Fraction of times a price jump as defined in (5.18) is found in correspondence of an IB. Results for different return horizons and three values of the threshold θ are reported.

if the volatility-normalized return is larger than a certain fixed threshold θ:

$$\frac{|r_i|}{\sigma_{\text{loc}}} > \theta \tag{5.18}$$

where σ_{loc} is an estimate of the local volatility at time t_i. Here we have employed the Realized Bipower Variation (Lee and Mykland, 2008):

$$\sigma^2_{\text{RBV}}(t_i) = \frac{1}{K-2} \sum_{j=i-K+2}^{i-1} \left|\log\left(\frac{P(t_j)}{P(t_{j-1})}\right)\right| \left|\log\left(\frac{P(t_{j-1})}{P(t_{j-2})}\right)\right| \tag{5.19}$$

using one minute returns ($t_i - t_{i-1} = 60s$) on a $K = 120$ minutes window. This estimator has the advantage of being relatively robust to presence of the tail events within the estimation window. However, we found other estimators such as the Realized Variance to yield substantially similar results for our purposes.

We estimate the probability of observing a price jump given the presence of an IB at time z. For this we analyze price dynamics around z on three time scales: 1 minute (computing return r_1 between $z - 10s$ and $z + 50s$), 5 minutes (r_5 is computed between $z - 50s$ and $z+250s$) and a time-scale which is defined by the relaxation time τ of the IB (we calculate return r_τ between $z - \frac{\tau}{6}$ and $z + \frac{5\tau}{6}$). For the definition of the $\theta\sigma$-jumps (5.18) we use a suitably rescaled volatility (5.19) which is originally calculated on 1 minute intervals: namely we use $\sigma_{\text{loc},1} = \sigma_{\text{loc}}$, $\sigma_{\text{loc},5} = \sqrt{5}\sigma_{\text{loc}}$ and $\sigma_{\text{loc},\tau} = \sqrt{\tau}\sigma_{\text{loc}}$ respectively.

Our results are reported in Table 5.12, which presents the fraction of times where a price jump (5.18) is observed when an IB is detected by our procedure. It is seen that a large fraction of IBs is accompanied by significant price moves: for example 25%–30% of all IBs are accompanied with 5σ-jumps on 1 minute time scale, and often these price moves do not immediately mean-revert: 12%–20% of 5-minute returns exceed 5σ threshold.

We have also explored what indications can the model parameters give on the likelihood that an IB is accompanied by a price jump. For this we have performed a logistic-regression analysis, presented in 5.E. We found that the probability of obtaining a price jump simultaneously with an IB is controlled mostly by the amplitude α of the IB. Surprisingly, the relaxation time τ does not seem to be a significant contributor, while adding the branching ratio n as regressor improves the performance of the classifier. Overall, the classifier using both α and n is quite reliable if it predicts the presence of a jump, while it suffers of a high false negative rate.

So far we have investigated the presence of a price jump given a nearby IB. Now we consider the reverse question, namely whether the price jumps are accompanied by an intensity shocks. We again consider absolute one minute returns normalized by the local volatility estimated using Eq. (5.19). Table 5.13 shows the fraction of these jumps that match an IB within ± 5 min. Depending on threshold level θ, from 20 to 40% of all price jumps correspond to an IB as detected by our procedure.

	EURJPY	EURUSD	USDJPY
$\theta = 3$	0.23	0.22	0.18
$\theta = 4$	0.32	0.33	0.26
$\theta = 5$	0.42	0.44	0.34

Table 5.13: Fraction of price jumps that match an IB within ±5 min.

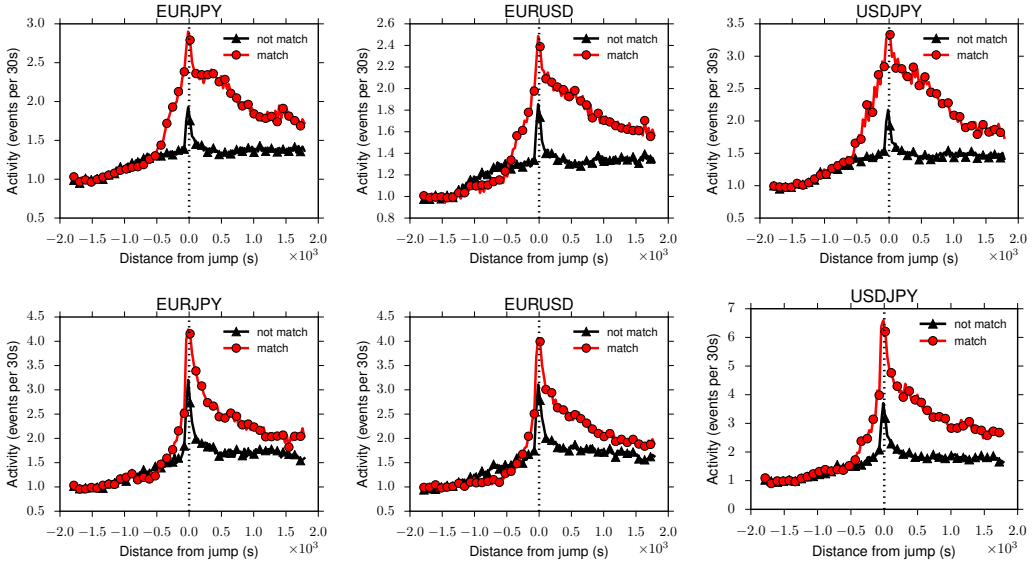

Figure 5.10: Mean normalized activity pattern around price jumps for jumps that match (do not match) with activity IBs detected by our procedure. The window is centered on price jump locations and is 1h wide. Before averaging, the activity is normalized by the mean activity in the first 10 minutes of the window. Top panels correspond to $3 < \theta < 5$, bottom panels - to $\theta \geq 5$.

Figure 5.10 presents the average activity around the price jumps for jumps that match an IB and for those that do not. We note that on average price jumps are associated with an increase of activity, though clearly not all of these activity spikes were identified as IB by our procedure (black lines in Figure): those activity spikes were not significantly different from the background endogenous process in terms of magnitude. Furthermore, the relaxation of activity after the price jump appears to be much faster for price jumps without IB that for those associated with an IB. It is interesting to note, that activity increases towards the price jumps on average in all cases. This suggests that most of detected price shocks are not exogenous, but either endogenously generated by the feedback mechanisms of the system, or anticipated exogenous shocks such as scheduled macro economical announcements.

One important result of our empirical analysis is that a significant fraction of IBs are not associated to macroeconomic news and jumps. These abrupt increases in high frequency volatility appear to be endogenously generated and not related to large price movements. Despite the fact that a full understanding of these phenomena is beyond the scope of this work, it is interesting to present a possible explanation for them, even if based on visual inspection of selected cases.

By looking at the bid-ask dynamics around these IBs not associated to price jumps or news we observe that, while before the start of the detected IB the midprice changes are relatively rare on the considered time scale, after it there is an intense readjustment process of best quotes, without a significant net price movement. This quote flickering is

frequently observed in many other IBs not associated to news or price jumps. Although we do not have a full explanation of this behavior, we note that it resembles a known strategy used by High Frequency Traders termed *quote stuffing* and consisting in flooding the market with huge numbers of orders and cancellations in rapid succession.

5.6 Conclusions

In this chapter we presented a novel procedure to detect intensity bursts in activity of a point process within a Hawkes processes framework. The ability to separate genuine external perturbation to the system from bursts generated by internal feedback mechanisms and correlations is a key feature of our procedure. From extensive numerical tests on both synthetic and real data we find that our procedure can be very effective in identifying sudden and short lived increases in activity that are not compatible with the endogenous or "normal" dynamics.

We applied the method to high frequency financial data describing the midprice dynamics in FX markets. We found a large number of IBs, both idiosyncratic to a specific rates and common to the three rates. Some of the IBs can be temporally related to price jumps and macroeconomic announcements. In the latter case we were able to compare the scheduled timing of the news with the inferred time of the start of the IB. In the former case we found that a large fraction of large jumps are associated with IBs. Interestingly a significant fraction of IBs appear to be unrelated with jumps and news, opening the question of the possible market events associated to them.

As a final remark we observe that the extension of the Hawkes model for incorporation of the explicit impact of IBs can be very relevant for mitigating the bias in the estimation of the branching ratio n in systems where exogenous shocks might be present. Indeed, we showed that when such shocks are not properly accounted for, conclusions based on standard Hawkes models can severely overestimate the branching ratio.

In conclusion, we stress that, although our methodology was developed with financial applications in mind, it can be applied wherever the dynamics is naturally bursty and correlated and the influence of sudden external shocks is relevant. Examples include tweets, email, web clicks, arrival of customers and many others.

Appendix

5.A Details of likelihood optimization

The log-likelihood (5.8) for model (5.2) with endogenous kernel of the form $\phi(t) = nh(t)$ and exogenous kernels of the form $\phi_S(t) = \alpha q(t)$ and given the observations $\{t_i\}_{i=1,\dots,N} \in [0, T]$ reads

$$\begin{aligned}\log \mathcal{L}(\mu, n, \psi, \{\alpha_k, z_k, \xi_k\}_{k=1,\dots,M}) = &- \mu T - nH_1(\psi) - \sum_{k=1}^{M} \alpha_k K_1(z_k, \xi_k) \\ &+ \sum_{t_i} \log\left(\mu + nH_2(\psi; t_i) + \sum_{k=1}^{M} \alpha_k K_2(z_k, \xi_k; t_i)\right)\end{aligned} \tag{5.20}$$

where μ is the baseline intensity parameter, (n, ψ) denotes the endogenous kernel parameters, (α_k, z_k, ξ_k) are the parameters of the k-th IB, M is the total number of IBs,

and

$$H_1 = \sum_{t_i} \left(\tilde{h}(T - t_i) - \tilde{h}(0) \right) \tag{5.21}$$

$$H_2 = \sum_{t_j < t_i} h(t_i - t_j) \tag{5.22}$$

$$K_1(k) = \tilde{q}(T - z_k) - \tilde{q}(0) \tag{5.23}$$

$$K_2(k; t_i) = q(t_i - z_k) \tag{5.24}$$

with $\tilde{g}$ denoting the antiderivative of the function g. We note that model (5.1) is recovered when $M = 0$. As noted in Filimonov and Sornette (2015), in the optimization of (5.20), one parameter can be obtained from the relation

$$n \frac{\partial \log \mathcal{L}}{\partial n} + \mu \frac{\partial \log \mathcal{L}}{\partial \mu} + \sum_k f_k \frac{\partial \log \mathcal{L}}{\partial \alpha_k} = -\mu T - n H_1 - \sum_{k=1}^{M} \alpha_k K_1(z_k, \xi_k) + N \tag{5.25}$$

which holds at the optimum. Hence, finding the maximum of (5.20) is equivalent to minimize

$$\begin{aligned} G(n, \psi, \{\alpha_k, z_k, \xi_k\}_{k=1,\dots,M}) = -\sum_{t_i} \log \Bigg[\frac{N}{T} + n \left(H_2(\psi; t_i) - \frac{H_1(\psi)}{T} \right) \\ + \sum_{k=1}^{M} \alpha_k \left(K_2(z_k, \xi_k; t_i) - \frac{K_1(z_k, \xi_k)}{T} \right) \Bigg] \end{aligned} \tag{5.26}$$

When $M = 0$ we perform the optimization using standard optimizers such as the L-BFGS-B method (Byrd et al., 1995). Since G is usually not convex, multiple starting points are tried in order to improve the chances of finding the global optimum (Filimonov and Sornette, 2015).

When adding one IB term we use a subordination procedure to perform the optimization in a similar spirit to Filimonov and Sornette (2015). Specifically, we proceed as follows. Let us indicate with θ_1 all the parameters of the model except the IB location z_1. Then, we separate the optimization into the two step

$$\hat{\theta}_1 = \arg\min_{\theta_1} S(\theta_1) \tag{5.27}$$

with

$$S(\theta_1) = \min_{z_1 \in \{t_i \in W_1\}} G(\theta_1, z_1) \tag{5.28}$$

Using the guess $\bar{z}_1$ from the pre-estimation procedure, we minimize G with respect to θ_1 while keeping fixed $z_1 = \bar{z}_1$. Then in step (5.28) we update the estimate of z_1 while keeping θ_1 fixed. We perform the optimization (5.27) using standard quasi-newton algorithms, while the optimization with respect to the IB location z_1 can be performed with a direct search over the values $t_i \in W_1$, as repeated evaluation of (5.26) for different values of z_1 is very cheap computationally.

When we add another IB to the model we proceed in a similar fashion. We optimize again all the parameters except the first IB location and we separate as before the optimization over z_2 from the one over the other parameters θ_2.

5.B Appendix to Section 5.4

5.C Errors on IB parameters

$n = 0.3$						$n = 0.5$					
f	τ					f	τ				
	10	50	100	500	1000		10	50	100	500	1000
50	92	18	5	0	0	50	92	22	4	0	0
75	100	66	27	0	0	75	100	83	35	0	0
100	100	98	75	2	0	100	100	100	82	0	0
250	100	100	100	86	22	250	100	100	100	100	26
500	100	100	100	100	100	500	100	100	100	100	100
750	100	100	100	100	100	750	100	100	100	100	100
1000	100	100	100	100	100	1000	100	100	100	100	100

$n = 0.7$						$n = 0.9$					
f	τ					f	τ				
	10	50	100	500	1000		10	50	100	500	1000
50	71	44	14	0	0	50	14	57	54	2	1
75	99	92	62	0	0	75	29	93	90	17	2
100	100	100	93	9	0	100	57	100	100	62	16
250	100	100	100	100	64	250	99	100	100	100	100
500	100	100	100	100	100	500	99	98	99	100	100
750	100	100	100	100	100	750	100	98	99	100	100
1000	100	99	100	100	100	1000	98	99	97	97	99

Table 5.14: Percentage of correctly classified IBs for different combinations of the true IB parameters (f, τ). In this experiment we do not use the pre-identification algorithm to limit the search space for the IB location but instead we directly provide the correct search interval. The results refer to a sample size of roughly 5,000 events.

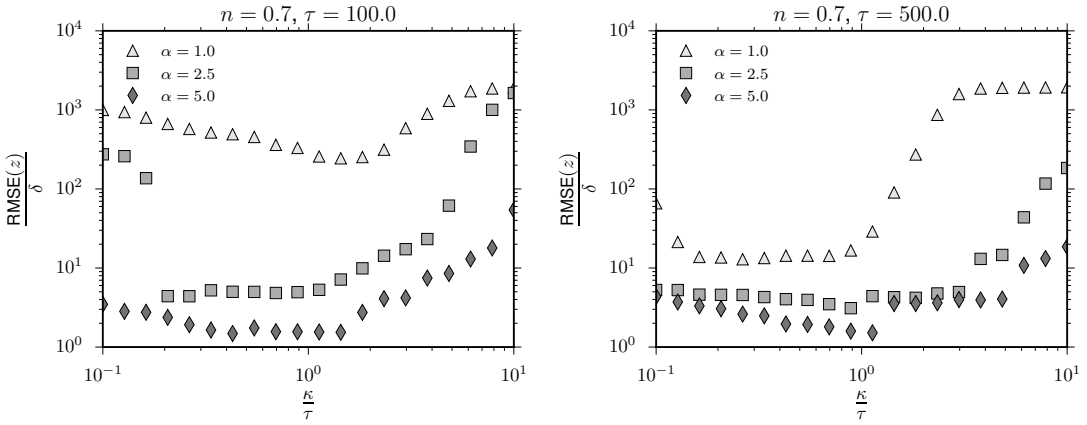

Figure 5.11: Root Mean Squared Error on the IB location z relative to the typical inter-event time $\delta = \frac{T}{N}$ as a function of $\frac{\kappa}{\tau}$. In the above figures $T = 3,600$, roughly $5,000$ events, and $n = 0.7$. 100 simulations were performed for each combination (α, τ).

$n = 0.3$					
f	τ				
	10	50	100	500	1000
50	0.34	0.31	0.30	0.30	0.30
75	0.37	0.32	0.32	0.30	0.30
100	0.42	0.35	0.33	0.31	0.30
250	0.64	0.66	0.60	0.33	0.31
500	0.74	0.92	0.93	0.65	0.42
750	0.77	0.98	1.00	0.95	0.71
1000	0.78	0.99	1.00	1.00	0.95

$n = 0.3$					
f	τ				
	10	50	100	500	1000
50	0.30	0.31	0.30	0.30	0.30
75	0.30	0.30	0.31	0.30	0.30
100	0.30	0.30	0.30	0.31	0.30
250	0.30	0.29	0.30	0.30	0.30
500	0.30	0.30	0.30	0.30	0.30
750	0.30	0.30	0.29	0.30	0.30
1000	0.30	0.30	0.30	0.29	0.29

$n = 0.7$					
f	τ				
	10	50	100	500	1000
50	0.71	0.70	0.70	0.69	0.69
75	0.72	0.71	0.70	0.69	0.69
100	0.73	0.72	0.71	0.70	0.69
250	0.77	0.78	0.77	0.72	0.70
500	0.83	0.85	0.86	0.80	0.75
750	0.87	0.90	0.91	0.89	0.82
1000	0.91	0.93	0.94	0.95	0.90

$n = 0.7$					
f	τ				
	10	50	100	500	1000
50	0.70	0.69	0.70	0.69	0.69
75	0.69	0.69	0.69	0.69	0.69
100	0.69	0.69	0.69	0.70	0.69
250	0.69	0.69	0.69	0.69	0.69
500	0.69	0.69	0.69	0.69	0.69
750	0.69	0.69	0.69	0.69	0.69
1000	0.70	0.70	0.69	0.70	0.69

Table 5.15: Average values of the branching ratio n of the endogenous kernel, obtained from the base Hawkes model (left) and from the best model selected by our procedure (right) when the simulated time series has a IB with parameters (f, τ). Here we present the case where the true values are $n = 0.3$ (top) and $n = 0.7$ (bottom).

	$n = 0.3$							
	CLL	CLS	CSL	CSS	FLL	FLS	FSL	FSS
First correct	99	100	97	93	78	100	87	73
Second correct	99	39	100	48	78	79	100	66
Both correct	98	39	97	42	78	79	87	63
No shock detected	0	0	0	0	22	0	0	24
More than two	0	0	1	0	0	0	0	0
	$n = 0.5$							
	CLL	CLS	CSL	CSS	FLL	FLS	FSL	FSS
First correct	100	100	97	97	97	99	90	86
Second correct	98	49	100	49	97	86	99	85
Both correct	98	49	97	47	97	86	90	81
No shock detected	0	0	0	0	3	1	1	10
More than two	0	0	0	0	0	0	0	0
	$n = 0.9$							
	CLL	CLS	CSL	CSS	FLL	FLS	FSL	FSS
First correct	78	94	76	93	64	86	68	92
Second correct	62	39	94	80	53	65	92	90
Both correct	61	39	73	74	52	65	68	86
No shock detected	21	5	3	0	35	12	8	4
More than two	3	0	6	7	5	2	4	5

Table 5.16: Results of the tests on simulation with two IBs for the $n = 0.3$, $n = 0.5$, and $n = 0.9$ cases. All quantities are expressed in percent. C stands for close, F for far, L for large, and S for small (see main text for details).

$n = 0.3$						$n = 0.5$					
f	τ					f	τ				
	10	50	100	500	1000		10	50	100	500	1000
50	33.7					50	43.3				
75	24.9	49.8				75	29.5	40.5			
100	22.8	29.1	30.8			100	24.1	30.8	43.8		
250	12.3	13.0	16.7	24.1		250	14.5	13.7	15.6	20.8	
500	8.3	8.6	10.2	15.0	20.1	500	9.8	10.2	11.6	13.2	15.8
750	6.5	7.9	8.5	9.1	13.2	750	8.5	7.9	8.4	11.1	13.9
1000	6.0	7.0	5.8	7.4	10.3	1000	6.2	7.7	7.7	9.1	9.4

$n = 0.7$						$n = 0.9$					
f	τ					f	τ				
	10	50	100	500	1000		10	50	100	500	1000
50	62.8					50					
75	37.1	38.8	70.0			75		101.8	66.8		
100	29.0	31.4	42.9			100	53.9	56.3	30.3		
250	18.5	17.5	17.5	22.0		250	32.5	27.6	20.3	41.2	
500	12.6	13.6	12.8	12.7	16.3	500	24.0	29.8	22.7	12.5	16.1
750	10.6	11.1	12.0	9.5	12.5	750	18.9	23.9	18.0	12.1	10.6
1000	7.4	11.0	11.1	9.1	8.6	1000	9.3	17.0	18.1	12.5	12.2

Table 5.17: Relative Root Mean Squared Error (%) on the α parameter obtained with our procedure on simulations where a single IB is present. One hundred simulation were generated for each combination of n, f, and τ. The MSE is shown only for cases where the IB was detected in at least 50 out of 100 repetitions. These tables refers to a sample size of approximately 5000 events.

5.D Kernel misspecification

In this section we present the result obtained by our procedure when the endogenous kernel $\phi(t)$ used to generate the data differs from the one used in the estimation.

In particular we run simulations with a single exponential kernel

$$\phi_{\text{SE}} = nbe^{-bt} \tag{5.29}$$

and a double exponential kernel

$$\phi_{\text{DE}} = n\left(ab_A e^{-b_A t} + (1-a)b_B e^{-b_B t}\right). \tag{5.30}$$

The test procedure is exactly the same as outlined in Sections 5.4.2 and 5.4.4. The result presented are for a target sample size of 10,000. In the following, SE1 (SE2) will denote simulations with the single exponential as kernel and parameter $b = 0.1$ ($b = 1$), while DE will denote simulations with the double exponential kernel and parameters $a = 0.7, b_A = 2$, and $b_B = 0.1$.

In Table 5.21 we present the rate of false positives for simulations where no IB is present, while in Tables 5.22, 5.23, 5.24 the rate of correctly classified IBs for the SE1, SE2 and DE cases respectively. Our procedure of the IB identification appears to be robust with respect to miss-specifications of the model.

$n = 0.3$						$n = 0.5$					
f	τ					f	τ				
	10	50	100	500	1000		10	50	100	500	1000
50						50					
75	27.4					75	33.9				
100	26.6	38.8				100	23.6	50.6			
250	12.7	18.0	19.1			250	12.5	15.6	19.3		
500	8.4	9.7	11.2	19.1		500	10.8	12.2	12.9	19.6	29.1
750	7.0	8.1	8.4	12.0	16.8	750	7.8	7.8	9.6	11.8	17.8
1000	5.9	6.6	7.3	10.6	13.7	1000	6.0	7.5	8.8	10.1	14.1

$n = 0.7$						$n = 0.9$					
f	τ					f	τ				
	10	50	100	500	1000		10	50	100	500	1000
50						50					
75	40.4					75					
100	35.9	42.6				100		49.3			
250	16.0	22.5	26.0			250	29.2	37.8	26.6	30.6	
500	12.2	13.1	13.0	17.9	24.9	500	17.1	24.3	19.0	16.3	18.7
750	11.2	11.7	11.4	11.9	15.4	750	16.7	23.4	17.0	11.2	16.2
1000	7.4	10.3	10.3	10.7	11.3	1000	8.9	17.0	15.3	11.5	11.7

Table 5.18: Relative Root Mean Squared Error (%) on the α parameter obtained with our procedure on simulations where a single IB is present. One hundred simulation were generated for each combination of n, f, and τ. The MSE is shown only for cases where the IB was detected in at least 50 out of 100 repetitions. These tables refers to a sample size of approximately 10000 events.

$n = 0.3$						$n = 0.5$					
f	τ					f	τ				
	10	50	100	500	1000		10	50	100	500	1000
50	44.9					50	39.9				
75	23.9	67.4				75	23.1	55.7			
100	23.1	31.6	59.1			100	17.7	38.8	54.5		
250	11.2	15.0	15.1	28.2		250	11.9	13.8	16.8	29.8	
500	7.1	7.4	9.9	18.9	31.5	500	7.5	8.6	8.8	16.5	32.0
750	5.7	6.0	7.3	11.7	20.4	750	6.2	5.8	7.4	11.6	19.1
1000	4.4	5.4	5.7	8.6	15.2	1000	4.5	4.7	5.2	9.3	13.9

$n = 0.7$						$n = 0.9$					
f	τ					f	τ				
	10	50	100	500	1000		10	50	100	500	1000
50	44.4					50					
75	36.9	33.4	44.1			75		27.0	35.3		
100	18.7	29.4	36.5			100	26.1	23.1	26.6		
250	12.8	13.3	14.8	34.0		250	17.7	11.9	13.8	29.6	
500	9.4	8.1	9.8	14.4	28.3	500	15.2	12.2	9.4	10.5	24.6
750	6.3	6.1	6.6	8.7	18.5	750	11.5	7.7	7.2	7.1	17.3
1000	5.7	4.9	5.7	6.8	11.4	1000	13.8	10.7	6.3	6.9	17.2

Table 5.19: Relative Root Mean Squared Error (%) on the τ parameter obtained with our procedure on simulations where a single IB is present. One hundred simulation were generated for each combination of n, f, and τ. The MSE is shown only for cases where the IB was detected in at least 50 out of 100 repetitions. These tables refers to a sample size of approximately 5000 events.

$n = 0.3$						$n = 0.5$					
f	τ					f	τ				
	10	50	100	500	1000		10	50	100	500	1000
50						50					
75	52.1					75	41.3				
100	30.2	47.1				100	25.2	35.2			
250	11.9	15.9	26.7			250	11.4	18.2	22.8		
500	7.8	9.8	12.8	30.0		500	7.5	10.1	11.8	26.2	37.9
750	5.6	8.2	9.0	18.1	30.4	750	6.8	6.9	9.1	14.2	33.6
1000	5.0	5.6	6.5	12.0	23.2	1000	5.1	5.9	7.4	12.4	22.1

$n = 0.7$						$n = 0.9$					
f	τ					f	τ				
	10	50	100	500	1000		10	50	100	500	1000
50						50					
75	39.7					75					
100	29.5	50.0				100		26.1			
250	13.8	16.4	18.8			250	19.3	17.3	17.6	28.6	
500	7.4	11.0	11.7	22.0	34.0	500	13.2	11.0	11.8	15.0	28.1
750	7.8	7.5	8.3	14.0	21.4	750	10.4	8.5	7.3	11.4	23.2
1000	6.6	6.3	6.6	10.7	17.7	1000	14.0	7.4	6.0	6.2	16.8

Table 5.20: Relative Root Mean Squared Error (%) on the τ parameter obtained with our procedure on simulations where a single IB is present. One hundred simulation were generated for each combination of n, f, and τ. The MSE is shown only for cases where the IB was detected in at least 50 out of 100 repetitions. These tables refers to a sample size of approximately 10000 events.

	DE				SE1				SE2			
n	0.3	0.5	0.7	0.9	0.3	0.5	0.7	0.9	0.3	0.5	0.7	0.9
FPR	0.0	0.0	0.5	0.8	0.1	0.0	0.0	0.5	0.0	0.1	0.2	0.2

Table 5.21: Percentage of false positive using Bayesian Information Criterion when the simulated model has no IBs. 1000 simulations where performed in each case. All values expressed in percent (%).

$n = 0.3$						$n = 0.5$					
f	τ					f	τ				
	10	50	100	500	1000		10	50	100	500	1000
50	0	0	0	0	0	50	0	0	0	0	0
75	0	0	0	0	0	75	0	0	0	0	0
100	0	0	0	0	0	100	1	0	0	0	0
250	95	11	0	0	0	250	100	81	21	0	0
500	100	100	100	0	0	500	100	100	100	1	0
750	100	100	100	0	0	750	100	100	100	49	0
1000	100	100	100	37	0	1000	100	100	100	99	5

$n = 0.7$						$n = 0.9$					
f	τ					f	τ				
	10	50	100	500	1000		10	50	100	500	1000
50	1	1	0	0	0	50	38	8	1	0	0
75	14	1	0	0	0	75	84	11	5	0	0
100	41	5	1	0	0	100	97	44	11	0	0
250	100	100	90	1	0	250	100	99	97	8	2
500	100	100	100	64	2	500	100	100	98	71	14
750	100	100	100	100	35	750	97	96	93	90	47
1000	100	100	100	100	84	1000	98	98	97	92	67

Table 5.22: Percentage of correctly classified IBs for different combinations of the true IB parameters (α, τ) expressed in terms of $f = \alpha\tau$ and τ. The results refer to the case SE1 and 100 simulations for each case where performed.

5.E Price jumps classification

We now investigate whether the IB parameters α and τ are good predictors for the presence of a price jump. To this end we fix $\theta = 4$ and we consider a logistic model where the dependent variable is the classification jump/no jump (mapped to $\{1, 0\}$) and the regressors are the estimates of α and τ or α and the branching ratio n. For each currency pair, we fit the logistic model on a randomly chosen subsample of about 70% of the detected IBs (later we will perform an out of sample analysis with the remaining 30%). In Table 5.25 we report the obtained coefficients as well as their standard errors for EURUSD. The results for the other pairs are very similar. We note immediately that the τ parameter does not appear to bring any contribution to the classification and indeed Akaike criterion selects the model with only α as regressor. When the variable n is used in place of τ as a regressor we obtain a significant coefficient for n, indeed we find that this model is favored by AIC over the oneusing only α for all three pairs.

In Table 5.26 we report some performance metrics obtained by using the fitted models to predict the presence of jump in price on the remaining 30% of the sample, using 0.5 as threshold to predict a price jump. Overall, the prediction yielded has fairly high accuracy and precision but suffer from a high false negative rates. For completeness, in Figure 5.12 we plot the ROC curves obtained on the test samples. Overall, the classifier based on the parameters α and n of IB is quite reliable if it predicts the presence of a jumps, while it can lead to significant errors if it predicts that the jump is not there.

$n = 0.3$						$n = 0.5$					
f	τ					f	τ				
	10	50	100	500	1000		10	50	100	500	1000
50	0	0	0	0	0	50	0	0	0	0	0
75	0	0	0	0	0	75	1	0	0	0	0
100	0	0	0	0	0	100	1	0	0	0	0
250	95	18	0	0	0	250	100	92	40	0	0
500	100	100	100	0	0	500	100	100	100	6	1
750	100	100	100	1	0	750	100	100	100	67	1
1000	100	100	100	57	0	1000	100	100	100	100	13

$n = 0.7$						$n = 0.9$					
f	τ					f	τ				
	10	50	100	500	1000		10	50	100	500	1000
50	8	1	0	0	0	50	5	0	0	0	0
75	23	1	0	0	0	75	22	2	1	0	0
100	58	11	2	0	0	100	65	6	2	1	0
250	100	100	92	3	0	250	100	84	63	19	3
500	100	100	100	79	8	500	100	99	98	93	57
750	100	100	100	99	60	750	95	95	98	97	78
1000	100	100	100	100	91	1000	94	90	94	89	88

Table 5.23: Percentage of correctly classified IBs for different combinations of the true IB parameters (α, τ) expressed in terms of $f = \alpha\tau$ and τ. The results refer to the case SE2 and 100 simulations for each case where performed.

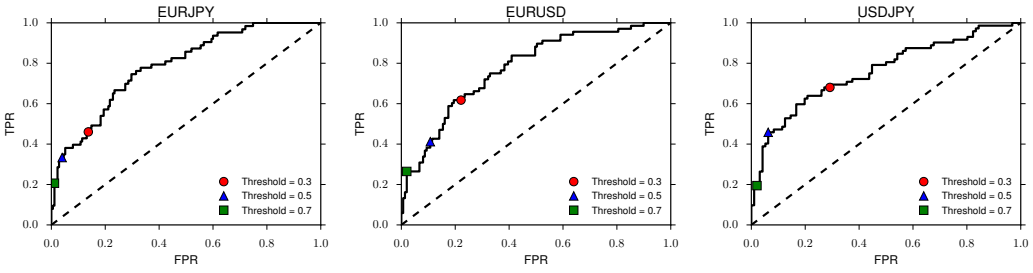

Figure 5.12: ROC curves for the logit classifier that uses α and n to predict whether a price jump ($\theta = 4$) will be present together with the IB.

$n = 0.3$						$n = 0.5$					
f	τ					f	τ				
	10	50	100	500	1000		10	50	100	500	1000
50	0	0	0	0	0	50	0	0	0	0	0
75	0	0	0	0	0	75	0	0	0	0	0
100	0	0	0	0	0	100	1	0	0	0	0
250	92	17	0	0	0	250	100	92	42	0	0
500	100	96	97	0	0	500	100	100	100	4	0
750	100	99	98	0	0	750	100	100	100	59	1
1000	100	100	100	53	0	1000	100	100	100	99	13

$n = 0.7$						$n = 0.9$					
f	τ					f	τ				
	10	50	100	500	1000		10	50	100	500	1000
50	6	0	0	0	0	50	14	0	1	0	0
75	27	5	0	0	0	75	45	4	0	0	0
100	61	21	5	0	0	100	87	22	1	0	0
250	100	100	94	1	0	250	100	98	83	1	0
500	100	100	100	63	4	500	100	100	100	56	4
750	100	100	100	100	45	750	96	96	98	91	50
1000	100	100	100	100	91	1000	87	96	93	91	77

Table 5.24: Percentage of correctly classified IBs for different combinations of the true IB parameters (α, τ) expressed in terms of $f = \alpha\tau$ and τ. The results refer to the case DE and 100 simulations for each case where performed.

	coef	**std err**	**z**	**P>|z|**	**[95.0% Conf. Int.]**
constant	-3.2572	0.352	-9.254	0.000	-3.947 -2.567
α	1.8699	0.247	7.557	0.000	1.385 2.355
τ	0.0001	0.000	0.893	0.372	-0.000 0.000
constant	-10.4279	1.685	-6.187	0.000	-13.731 -7.125
α	2.1019	0.251	8.374	0.000	1.610 2.594
n	10.0643	2.221	4.532	0.000	5.711 14.417

Table 5.25: Obtained parameters value and associated errors when α and τ (top) or α and n (bottom) are used as regressors. Results shown are for the EURUSD pair.

	EURJPY	EURUSD	USDJPY
TPR	0.33	0.40	0.46
TNR	0.96	0.89	0.94
Precision	0.75	0.63	0.85
Accuracy	0.79	0.74	0.73

Table 5.26: Out of sample performance metrics for the logistic model with α and n as regressors. True Positive Rate (TPR), True Negative Rate (TNR), Precision (True Positives/Predicted Positives) and Accuracy are reported.

Conclusions

In this thesis we presented some new applications of Hawkes point processes to high-frequency financial data. In the work of Chapter 3, we analyzed high frequency interactions between different order types and sizes. Our approach leveraged existing non-parametric estimation method developed for multivariate Hawkes processes to include the order size in the model by enlarging the dimension of the process. We found that the order size is an important variable to consider for a faithful modeling of the order book dynamics. We also confirmed previous findings and conjectures indicating that self-excitation observed in models that do not consider the order size is chiefly due to same size excitation. Furthermore, we observe a more dramatic and persistent impact of large orders, particularly on assets where they are relatively rare, in line with the idea that they carry more information. We also obtained a new theoretical result which affirm that the non-parametric estimation method of Bacry and Muzy (2016) will find negative-valued kernels whenever inhibition effects are present in the system. The possibility of investigating moderate inhibition effects alongside exciting with the aforementioned method further widens the scope and power of Hawkes processes analysis.

In the other two original works presented in Chapters 4 and 5, we tackled the modeling and detection of local non-stationarities in a Hawkes process framework. These works were originally inspired by the need to account for external punctual events like news announcements that can have a temporary but very significant effect on the order flow intensity, which in turn is related to volatility. We first developed an extension of the standard Hawkes model that is able to deal with cases where the occurrence time of the external event is known in advance. We showed that our model is able to reproduce very well both the magnitude and the time persistence of these news-driven intensity bursts and also to capture anticipation effects due to the fact that the announcement is scheduled in advance.

In our last work we devised a procedure that allows to apply the previous model also to cases where both the occurrence times and the number of burst is unknown, thus opening its use as an anomalies detection tool. We tested our procedure on FX data and we found that a large number of Intensity Bursts occurs, a sizable fraction of which are found simultaneously on different currency crosses. Although a significant part of them is related to scheduled economic announcements and price jumps, there is another relevant part which appears to be unrelated to either of those. Despite the lacking of a full understanding of these phenomena at the moment, preliminary investigations suggest that they could be related at least in some instances to high frequency strategies where a great number of orders and cancellations is submitted to the market (*quote stuffing*). We believe our methodology could be a valuable addition to the already available tools for anomaly detection, which is a central theme for market regulators as well as for risk managers.

Finally, we stress that the range of possible applications of the presented methodolo-

gies extends beyond financial ones. Indeed, all the methods we exposed are in principle amenable to be used whenever a point process description of the system is appropriate and mutual or self interaction are relevant.

Acknowledgments

Research presented in Chapter 3 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 Françaiseand from the chair of the Risk Foundation: Quantitative Management Initiative. We also thank Thibault Jaisson, Iacopo Mastromatteo and Jean-François Muzy for useful discussions.

For research presented in Chapter 4, we acknowledge the eRisk HSBC team for support and EBS for having provided the data. We also thank E. Bacry and J-P. Bouchaud for useful discussions.

For the work of Chapter 5, we are grateful to the Chair of Entrepreneurial Risks of ETH Zürich and specifically to Prof. Didier Sornette for fruitful discussions and financial support. We also thanks Spencer Wheatley and Prof. Frederic Abergel for useful discussions.

Bibliography

Abergel, F. and A. Jedidi (2013). A mathematical approach to order book modeling. *International Journal of Theoretical and Applied Finance 16*(05), 1350025.

Abergel, F. and A. Jedidi (2015). Long-time behavior of a hawkes process–based limit order book. *SIAM Journal on Financial Mathematics 6*(1), 1026–1043.

Abergel, F., C.-A. Lehalle, and M. Rosenbaum (2014). Understanding the stakes of high-frequency trading. *The Journal of Trading 9*(4), 49–73.

Aït-Sahalia, Y., J. Jacod, et al. (2009). Testing for jumps in a discretely observed process. *The Annals of Statistics 37*(1), 184–222.

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on 19*(6), 716–723.

Alfonsi, A. and P. Blanc (2016). Dynamic optimal execution in a mixed-market-impact Hawkes price model. *Finance and Stochastics 20*(1), 183–218.

Allen, D., Z. Lazarov, M. McAleer, and S. Peiris (2009). Comparison of alternative ACD models via density and interval forecasts: Evidence from the australian stock market. *Math. Comput. Simulat. 79*(8), 2535 – 2555.

Almgren, R. (2012). High-frequency event analysis in eurex interest rate futures. Technical report.

Almgren, R. and N. Chriss (2001). Optimal execution of portfolio transactions. *Journal of Risk 3*, 5–40.

Andersen, T. G. and T. Bollerslev (1998). Deutsche mark–dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies. *the Journal of Finance 53*(1), 219–265.

Andersen, T. G., T. Bollerslev, and D. Dobrev (2007). No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and iid noise: Theory and testable distributional implications. *Journal of Econometrics 138*(1), 125–180.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica 61*(4), 821–856.

Avellaneda, M. and M. D. Lipkin (2003). A market-induced mechanism for stock pinning. *Quantitative Finance 3*(6), 417–425.

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

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

Bacry, E., A. Iuga, M. Lasnier, and C.-A. Lehalle (2015). Market impacts and the life cycle of investors orders. *Market Microstructure and Liquidity 1*(02), 1550009.

Bacry, E., T. Jaisson, and J. Muzy (2016). Estimation of slowly decreasing Hawkes kernels: application to high-frequency order book dynamics. *Quantitative Finance 16*(8), 1179–1201.

Bacry, E., I. Mastromatteo, and J.-F. Muzy (2015). Hawkes processes in finance. *Market Microstructure and Liquidity 1*(01), 1550005.

Bacry, E. and J.-F. Muzy (2014). Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance 14*(7), 1147–1166.

Bacry, E. and J.-F. Muzy (2016). First-and second-order statistics characterization of Hawkes processes and non-parametric estimation. *IEEE Transactions on Information Theory 62*(4), 2184–2202.

Bank for International Settlements (2013). *Triennial Central Bank Survey. Foreign exchange turnover in April 2013: preliminary global results.*

Bauwens, L., P. Giot, J. Grammig, and D. Veredas (2004). A comparison of financial duration models via density forecasts. *Int. J. Forecasting 20*(4), 589–609.

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

Bauwens, L., W. B. Omrane, and P. Giot (2005). News announcements, market activity and volatility in the euro/dollar foreign exchange market. *Journal of International Money and Finance 24*(7), 1108 – 1125.

Bessembinder, H., M. Panayides, and K. Venkataraman (2009). Hidden liquidity An analysis of order exposure strategies in electronic stock markets. *Journal of Financial Economics 94*(3), 361–383.

Biais, B., P. Hillion, and C. Spatt (1995). An empirical analysis of the limit order book and the order flow in the paris bourse. *J. Finance 50*(5), 1655–1689.

Birz, G. and J. R. J. Lott (2011). The effect of macroeconomic news on stock returns: new evidence from newspaper coverage. *Journal of Banking & Finance 35*, 2791–2800.

Blanc, P., J. Donier, and J.-P. Bouchaud (2015). Quadratic Hawkes processes for financial prices. *arXiv:1509.07710v1*.

Blundell, C., J. Beck, and K. A. Heller (2012). Modelling reciprocating relationships with Hawkes processes. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25*, pp. 2600–2608. Curran Associates, Inc.

Bollerslev, T., U. Kretschmer, C. Pigorsch, and G. Tauchen (2009). A discrete-time model for daily s & p500 returns and realized variations: Jumps and leverage effects. *Journal of Econometrics 150*(2), 151–166.

Bormetti, G., L. M. Calcagnile, M. Treccani, F. Corsi, S. Marmi, and F. Lillo (2015). Modelling systemic price cojumps with Hawkes factor models. *Quantitative Finance 15*(7), 1137–1156.

Bouchaud, J. P. (2010). Price impact. In R. Cont (Ed.), *Encyclopedia of Quantitative Finance*. John Wiley & Sons Ltd.

Bouchaud, J.-P., M. Mézard, M. Potters, et al. (2002). Statistical properties of stock order books: empirical results and models. *Quantitative finance 2*(4), 251–256.

Bowsher, C. G. (2002). Modelling security market events in continuous time: Intensity based, multivariate point process models.

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

Brémaud, P. and L. Massoulié (1996). Stability of nonlinear hawkes processes. *Ann. Probab. 24*(3), 1563–1588.

Brémaud, P. and L. Massoulié (2001). Hawkes branching point processes without ancestors. *Journal of applied probability 38*(1), 122–135.

Byrd, R. H., P. Lu, J. Nocedal, and C. Zhu (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing 16*(5), 1190–1208.

Chaboud, A. P., B. Chiquoine, E. Hjalmarsson, and C. Vega (2014). Rise of the machines: Algorithmic trading in the foreign exchange market. *The Journal of Finance 69*(5), 2045–2084.

Chakraborti, A., I. M. Toke, M. Patriarca, and F. Abergel (2009). Econophysics: Empirical facts and agent-based models. *arXiv preprint arXiv:0909.1974*.

Chakraborti, A., I. M. Toke, M. Patriarca, and F. Abergel (2011). Econophysics review: I. Empirical facts. *Quantitative Finance 11*(7), 991–1012.

Challet, D. and R. Stinchcombe (2001). Analyzing and modeling 1+ 1d markets. *Physica A: Statistical Mechanics and its Applications 300*(1), 285–299.

Chandola, V., A. Banerjee, and V. Kumar (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR) 41*(3), 15–58.

Chornoboy, E. S., L. P. Schramm, and A. F. Karr (1988). Maximum likelihood identification of neural point process systems. *Biological Cybernetics 59*(4), 265–275.

Claeskens, G. and N. Hjort (2008). *Model Selection and Model Averaging*. Cambridge University Press.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance 1*(2), 223–236.

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

Cont, R. and P. Tankov (2004). *Financial Modelling With Jump Processes*. Chapman and Hall/CRC, London.

Crane, R. and D. Sornette (2008). Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences of the United States of America 105*(41), 15649–15653.

Cutler, D. M., J. M. Poterba, and L. H. Summers (1989). What moves stock prices? *Journal of Portfolio Management 15*, 4–12.

Da Fonseca, J. and R. Zaatour (2014). Hawkes process: Fast calibration, application to trade clustering, and diffusive limit. *Journal of Futures Markets 34*(6), 548–579.

Daley, D. J. and D. Vere-Jones (2003). *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods*. Springer, New York.

Daley, D. J. and D. Vere-Jones (2008). *An Introduction to the Theory of Point Processes, Volume II: General Theory and Structure*. Springer, New York.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika 64*(2), 247–254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika 74*(1), 33–43.

Donnay, K. and V. Filimonov (2014). Views to a war: systematic differences in media and military reporting of the war in Iraq. *EPJ Data Science 3*(1), 25.1–29.

Easley, D. and M. O'hara (1987). Price, trade size, and information in securities markets. *Journal of Financial Economics 19*(1), 69–90.

Ederington, L. H. and J. H. Lee (1993). How markets process information: News releases and volatility. *The Journal of Finance 48*(4), 1161–1191.

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

Ellul, A., C. W. Holden, P. Jain, and R. Jennings (2003). Determinants of order choice on the new york stock exchange. *manuscript, Indiana University*.

Embrechts, P., T. Liniger, L. Lin, et al. (2011). Multivariate Hawkes processes: an application to financial data. *Journal of Applied Probability 48*, 367–378.

Engle, R. F. (2000). The econometrics of ultra-high-frequency data. *Econometrica 68*(1), 1–22.

Engle, R. F., R. Ferstenberg, and J. R. Russell (2006). Measuring and modeling execution cost and risk.

Engle, R. F. and V. K. Ng (1993). Measuring and testing the impact of news on volatility. *Journal of Finance 48*, 1749–1778.

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

Farmer, J. D., F. Lillo, et al. (2004). On the origin of power-law tails in price fluctuations. *Quantitative Finance 4*(1), 7–11.

Farmer, J. D., P. Patelli, and I. I. Zovko (2005). The predictive power of zero intelligence in financial markets. *Proceedings of the National Academy of Sciences of the United States of America 102*(6), 2254–2259.

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

Filimonov, V. and D. Sornette (2015). Apparent criticality and calibration issues in the Hawkes self-excited point process model: application to high-frequency financial data. *Quantitative Finance 15*(8), 1293–1314.

Foucault, T. (1999). Order flow composition and trading costs in a dynamic limit order market. *Journal of Financial markets 2*(2), 99–134.

Glosten, L. R. and P. R. Milgrom (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of financial economics 14*(1), 71–100.

Golub, A., A. Dupuis, and R. Olsen (2013). High-Frequency Trading in FX Markets. In D. Easley, M. L. de Prado, and M. O'Hara (Eds.), *High-Frequency Trading. New Realities for Traders, Markets and Regulators*, pp. 65–90. Risk Books.

Gould, M. D., M. A. Porter, S. Williams, M. McDonald, D. J. Fenn, and S. D. Howison (2013). Limit order books. *Quantitative Finance 13*(11), 1709–1742.

Gresnigt, F. and P. H. Franses (2015). Specification Testing in Hawkes Models.

Groß-Klußmann, A. and N. Hautsch (2011). When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions. *Journal of Empirical Finance 18*(2), 321–340.

Hall, A. D. and N. Hautsch (2006). Order aggressiveness and order book dynamics. *Empirical Economics 30*(4), 973–1005.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica 64*(2), 413.

Hansen, N. R., P. Reynaud-Bouret, and V. Rivoirard (2015). Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli 21*(1), 83–143.

Hardiman, S. J., N. Bercot, and J.-P. Bouchaud (2013). Critical reflexivity in financial markets: a Hawkes process analysis. *The European Physical Journal B 86*, 1–9.

Hardiman, S. J. and J.-P. Bouchaud (2014). Branching-ratio approximation for the self-exciting Hawkes process. *Physical Review E 90*(6), 062807.

Harris, L. and J. Hasbrouck (1996). Market vs. limit orders: the superdot evidence on order submission strategy. *Journal of Financial and Quantitative analysis 31*(02), 213–231.

Harris, T. E. (2002). *The Theory of Branching Processes*. Dover Phoenix Editions.

Hasbrouck, J. (1988). Trades, quotes, inventories, and information. *Journal of Financial Economics 22*(2), 229–252.

Hasbrouck, J. (1991). Measuring the information content of stock trades. *The Journal of Finance 46*(1), 179–207.

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

Hawkes, A. G. and D. Oakes (1974). A Cluster Process Representation of a Self-Exciting Process. *Journal of Applied Probability 11*(3), 493–503.

Hendershott, T., C. M. Jones, and A. J. Menkveld (2011). Does algorithmic trading improve liquidity? *The Journal of Finance 66*(1), 1–33.

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

Hollifield, B., R. A. Miller, and P. Sandås (2004). Empirical analysis of limit order markets. *The Review of Economic Studies 71*(4), 1027–1063.

Huang, W., C.-A. Lehalle, and M. Rosenbaum (2015). Simulating and Analyzing Order Book Data: The Queue-Reactive Model. *Journal of the American Statistical Association 110*(509), 107–122.

Jaisson, T. (2015). *Market activity and price impact throughout time scales*. Theses, Ecole Polytechnique.

Jaisson, T., M. Rosenbaum, et al. (2015). Limit theorems for nearly unstable hawkes processes. *The Annals of Applied Probability 25*(2), 600–631.

Joulin, A., A. Lefevre, D. Grunberg, and J. Bouchaud (2008). Stock price jumps: News and volume play a minor role. *Wilmott Magazine*, 1–7.

Jovanović, S., J. Hertz, and S. Rotter (2015). Cumulants of Hawkes point processes. *Physical Review E 91*, 042802.

Kirchner, M. (2015). An estimation procedure for the Hawkes process. *arXiv:1509.02017*.

Kirilenko, A. A., A. S. Kyle, M. Samadi, and T. Tuzun (2015). The flash crash: The impact of high frequency trading on an electronic market. *Available at SSRN 1686004*.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, 1315–1335.

Lallouache, M. and D. Challet (2016). The limits of statistical significance of Hawkes processes fitted to financial data. *Quantitative Finance 16*(1), 1–11.

Lange, T. and A. Rahbek (2009). An Introduction to Regime Switching Time Series Models. In *Handbook of Financial Time Series*, pp. 871–887. Berlin, Heidelberg: Springer Berlin Heidelberg.

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

Lee, S. S. and P. A. Mykland (2008). Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics. *Review of Financial Studies 21*(6), 2535–2563.

Lewis, E. and G. Mohler (2011). A nonparametric em algorithm for multiscale hawkes processes. *Joint Satistical Meetings 2011*.

Lewis, E., G. Mohler, P. J. Brantingham, and A. L. Bertozzi (2012). Self-exciting point process models of civilian deaths in iraq. *Security Journal 25*(3), 244–264.

Lillo, F., S. Micciché, M. Tumminello, J. Piilo, and R. N. Mantegna (2015). How news affects the trading behaviour of different categories of investors in a financial market. *Quantitative Finance 15*(2), 213–229.

Linger, T. J. (2009). *Multivariate Hawkes processes*. Theses, ETH Zurich.

Lo, I. and S. G. Sapp (2010). Order aggressiveness and quantity: How are they determined in a limit order market? *Journal of International Financial Markets, Institutions and Money 20*(3), 213–237.

MacKinlay, D. (2015). Estimating self-excitation effects for social media using the Hawkes process. Master's thesis, ETH Zurich.

Manganelli, S. (2005). Duration, volume and volatility impact of trades. *J. Financ. Mark. 8*(4), 377–399.

Martins, R. and D. Hendricks (2016). The statistical significance of multivariate Hawkes processes fitted to limit order book data. *arXiv:1604.01824*.

Minc, H. (1988). *Nonnegative matrices*. Wiley.

Mitchell, M. (2009). *Complexity: A guided tour*. Oxford University Press.

Mitchell, M. L. and J. H. Mulherin (1994). The impact of public information on the stock market. *The Journal of Finance 49*(3), 923–950.

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

Muni Toke, I. and F. Pomponio (2011). Modelling trades-through in a limited order book using Hawkes processes. *Economics discussion paper* (2011-32).

Møller, J. and J. G. Rasmussen (2005). Perfect simulation of Hawkes processes. *Advances in Applied Probability 37*(3), 629–646.

Obizhaeva, A. A. and J. Wang (2013). Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets 16*(1), 1–32.

Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics 30*(1), 243–261.

Ogata, Y. (1981). On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory 27*(1), 23–31.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical association 83*(401), 9–27.

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

Pacurar, M. (2008). Autoregressive conditional duration models in finance: a survey of the theoretical and empirical literature. *J. Econ. Surv. 22*(4), 711–751.

Parlour, C. A. (1998). Price dynamics in limit order markets. *Review of Financial Studies 11*(4), 789–816.

Parlour, C. A. and D. J. Seppi (2008). Limit order markets: A survey. *Handbook of financial intermediation and banking 5*, 63–95.

Pierre Bremaud, L. M. (1996). Stability of Nonlinear Hawkes Processes. *The Annals of Probability 24*(3), 1563–1588.

Rambaldi, M., E. Bacry, and F. Lillo (2016). The role of volume in order book dynamics: a multivariate Hawkes process analysis. *(submitted)*.

Rambaldi, M., V. Filimonov, and F. Lillo (2016). Detection of intensity bursts using Hawkes processes: an application to high frequency financial data. *(submitted)*.

Rambaldi, M., P. Pennesi, and F. Lillo (2015). Modeling foreign exchange market activity around macroeconomic news: Hawkes-process approach. *Physical Review E 91*, 012819.

Reynaud-Bouret, P., V. Rivoirard, and C. Tuleau-Malot (2013). Inference of functional connectivity in neurosciences via Hawkes processes. In *1st IEEE Global Conference on Signal and Information Processing*.

Reynaud-Bouret, P., S. Schbath, et al. (2010). Adaptive estimation for hawkes processes; application to genome analysis. *The Annals of Statistics 38*(5), 2781–2822.

Roşu, I. (2009). A dynamic model of the limit order book. *Review of Financial Studies 22*(11), 4601–4641.

Rubin, I. (1972). Regular point processes and their detection. *IEEE Transactions on Information Theory 18*(5), 547–557.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics 6*(2), 461–464.

Shiller, R. J. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *The American Economic Review 71*(3), 421–436.

Smith, E., J. D. Farmer, L. s. Gillemot, S. Krishnamurthy, et al. (2003). Statistical theory of the continuous double auction. *Quantitative finance 3*(6), 481–514.

Sornette, D. and G. Ouillon (2005). Multifractal scaling of thermally activated rupture processes. *Physical Review Letters 94*, 038501.

Toke, I. M. (2011). Market making" in an order book model and its impact on the spread. In F. Abergel, B. K. Chakrabarti, A. Chakraborti, and M. Mitra (Eds.), *Econophysics of Order-driven Markets*, New Economic Windows, pp. 49–64. Springer Milan.

Toke, I. M. (2015). The order book as a queueing system: average depth and influence of the size of limit orders. *Quantitative Finance 15*(5), 795–808.

Toke, I. M. and N. Yoshida (2016). Modelling intensities of order flows in a limit order book. *arXiv:1602.03944*.

Wong, C. S. and W. K. Li (2001). On a Mixture Autoregressive Conditional Heteroscedastic Model. *Journal of the American Statistical Association 96*(455), 982–995.

Zhao, L. (2010). *A model of limit-order book dynamics and a consistent estimation procedure*. Ph. D. thesis, Carnegie Mellon University.

Zheng, B., F. Roueff, and F. Abergel (2014). Modelling bid and ask prices using constrained hawkes processes: Ergodicity and scaling limit. *SIAM J. Finan. Math. 5*(1), 99–136.

Zhou, K., H. Zha, and L. Song (2013). Learning social infectivity in sparse low-rank networks using multi-dimensional hawkes processes. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, pp. 641–649.