High dimensional Hawkes processes for limit order books Modelling, empirical analysis and numerical calibration

Xiaofei Lu, Frédéric Abergel

To appear in *Quantitative Finance*, Vol. 00, No. 00, Month 20XX, 1–24

High dimensional Hawkes processes for limit order books

Modelling, empirical analysis and numerical calibration

Xiaofei LU† and Frédéric ABERGEL*

Laboratory MICS, CentraleSupélec

(*Received 00 Month 20XX; in final form 00 Month 20XX*)

High-dimensional Hawkes processes with exponential kernels are used to describe limit order books in order-driven financial markets. The dependencies between orders of various types are carefully studied and modelled, based on a thorough empirical analysis. The observation of inhibition effects is particularly interesting, and leads us to the use of non-linear Hawkes processes. A specific attention is devoted to the calibration problem, in order to account for the high dimensionality of the problem and the very poor convexity properties of the MLE. Our analyses show a good agreement between the statistical properties of order book data and those of the model.

Keywords: Hawkes processes, limit order books, high frequency data, non-convex optimization

JEL Classification: C58, C61, G14

1. Introduction

Limit order books have attracted a considerable amount of attention since the electronification of financial markets in the early '90s. The historical quote-driven markets, where designated market makers used to provide liquidity to all participants, have largely evolved into order-driven markets, where buy and sell orders are matched continuously in a double auction queueing system.

In an order-driven market, participants can submit orders of three basic types: limit order, market order and cancellation:

- **Limit order**: An order that specifies an upper/lower price limit (also called "quote") at which one (commonly called "liquidity provider") is willing to buy/sell a certain number of shares. The advantage of the limit order is that the transaction price is better than the instantaneous mid-price. However, there is no certainty that the limit order will be executed. Currently most markets adopt the "first in first out" rule, i.e. the priorities of limit orders are decided first according to price, and then to arrival time. A limit order can be entirely, partly or not executed.
- **Market order**: An order that triggers an immediate buy/sell transaction for a certain number of shares at the best available opposite quote(s). The advantage is to offer an immediate execution, however the price is worse than the mid-price. A market order can be executed with different limit orders as counterparties. The price is not necessarily the best limit price,

† Email: xiaofei.lu@centralesupelec.fr
*Corresponding author Email: frederic.abergel@centralesupelec.fr

if the quantity is big enough that the order eats up completely the first limit and hits the second or higher limits.

- **Cancellation** : An order that removes an existing limit order.

In addition to these three main types of orders, there exist various order services provided by the exchanges such as "stop loss", "good til' canceled"... Also note that some markets allow orders, such as "iceberg" orders, to provide hidden liquidity, making their presence difficult to infer from the order flow. Nevertheless, it is commonly agreed upon - and verified in practice - that the basic orders carry enough information for market microstructure studies.

An example is given in **Figure 1**.

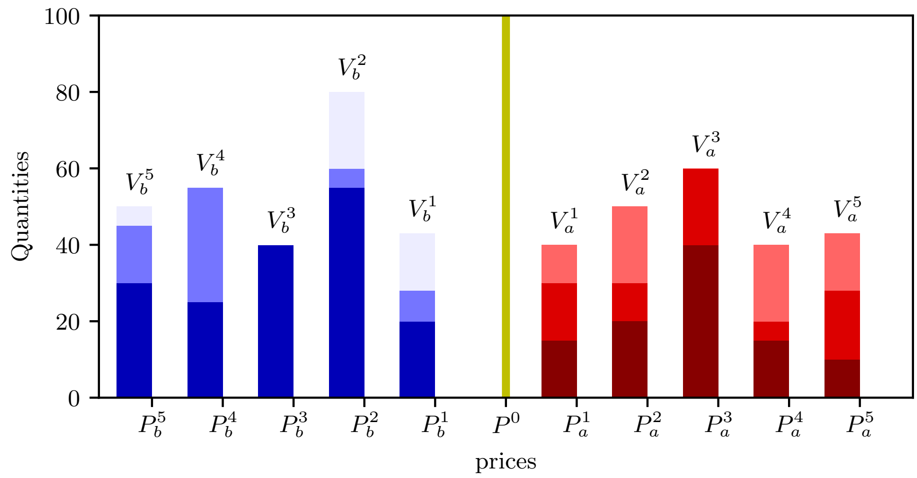

Figure 1. Illustrative order book example. Blue bars on the left half of the figure represent the available buy orders with prices $P_b^{\cdot}$ and total quantities $V_b^{\cdot}$. These correspond to the buyer side, also called the bid side. Participants in the bid side are providing liquidity with prices at which they are ready to buy some quantities of the stock. The right hand bars represent the sell side, commonly called the ask side or offer side, where participants willing to sell post their orders with the prices they are ready to sell the stock. The line in the middle corresponds to the mid-price level and is computed as the average between the best (highest) bid price and the best (lowest) ask price. A transaction occurs when a sell order and a buy order are at least partially matched. A queue of limit orders with the same price is called a limit. Different colors in the same limit represent orders with different priority with darker bars having higher execution priority.

Due to their obvious relevance to the understanding of financial markets, limit order books have been extensively studied in the past decade, both from empirical and mathematical points of view, see Abergel *et al.* (2016) for a survey of their properties. In particular, the mathematical modelling of limit order books is itself an active research area that has many useful and practical applications, and this paper is a contribution to the field.

A popular class of order book models is that of Markovian models, originating with the so-called *zero-intelligence* models as in Smith *et al.* (2003), then enriched with more complex and realistic contributions such as Cont *et al.* (2010) or Huang *et al.* (2015). In Markovian models the order flows are described by point processes with state-dependent conditional intensities.

In parallel with the development of Markovian models, and serving as an inspiration for the present work, many empirical studies have identified some **time dependency properties** of financial markets. To name a few, Gopikrishnan *et al.* (2000), Bouchaud *et al.* (2009) underline a significant positive autocorrelation and slow decay of the trade flow. Chakraborti *et al.* (2011) confirms that the Poisson hypothesis for the arrival of orders is not empirically satisfied, whereas Eisler *et al.* (2012) is an in-depth study of the correlation between, and price impact of, orders of all types. These findings advocate for a direct modelling of the temporal dependencies between order arrivals and in such a context, Hawkes processes have come up as a very adequate choice.

For this reason, Hawkes processes have triggered a lot of interest in recent studies on market microstructure and limit order books. To name a few, Bacry and Muzy (2014), Bacry *et al.* (2013, 2016) propose various models of price and order flow models, whereas Bacry *et al.* (2015) is an extensive survey of the application of Hawkes process in finance. In the specific context of limit order books, Large (2007) is an early study of Hawkes processes applied to order book modelling, Hawkes-process-based limit order book models are introduced and mathematically investigated in Abergel and Jedidi (2015), Zheng *et al.* (2014) and, in a sligthly different direction, Rambaldi *et al.* (2016) models the order volumes - in addition to their types - based on a multivariate Hawkes process.

In this paper, we contribute to this very fruitful and original strand of research by designing, analyzing and calibrating high-dimensional Hawkes processes describing limit order books. The effects that are most important to model are the mutual excitations between orders of various types, including some inhibition effects that are hightlighted in our empirical analysis.

The quality of various Hawkes-process-based order book models will be assessed using some objective criteria: a model will be deemed satisfactory if it can reproduce as many as possible of the stylized facts of financial data. Our approach starts with a precise empirical analysis of the dependencies between order arrivals of various types. Then, models built from multivariate, possibly nonlinear, Hawkes processes with multiple exponential kernels are introduced. Once a model is designed, it is evaluated. Following most papers involving Hawkes processes for order book modelling, the natural quantities of interest are the inter-event durations - or : inter-event forward recurrence times, which will be the main objects under scrutiny in the present work. The distribution of forward recurrence times, as well as the signature plot[1], are used as selection criteria. With this approach, we are able to discriminate between various Hawkes-process-based models, and provide a financial interpretation of the more successful ones in terms of their behaviour at various time scales, and the presence of inhibition as well as excitation effects.

In a slightly different, yet very important perspective, we also address carefully the estimation problem: it is well-known that the Maximum Likelihood Estimator of a Hawkes process is not the solution to a convex optimization problem, and therefore, great care must be taken when using it to obtain the values of the parameters. Actually, we found this part of our work so relevant and influential in practice that we decided it was worth a full section of the paper, hopefully sharing some useful knowledge with other potential users of high dimensional Hawkes processes.

The paper is organized as follows: Section 2 presents our most relevant empirical findings, laying the ground for the modelling based on linear and nonlinear Hawkes processes discussed in Section 3. In Section 4, the numerical aspects of model calibration are discussed in detail. Finally, some concluding remarks are presented in Section 5.

2. Empirical findings: the interdependencies of order book events

In this section, we present our main empirical findings on the dependencies between order arrivals. These findings pave the way for the modelling avenues followed in the next sections.

2.1. *Data and Framework*

This paper focuses on the DAX listed 30 stocks trading in XETRA - the electronic trading venue of the Frankfurt Stock Exchange. Three months (February to April 2016) of tick-by-tick data are used in this study. The data consist in the list of all trades and order book states any time a modification or a transaction occurs - with a resolution of $1\mu s$ ($10^{-6}s$). As is classical for high frequency financial data, see e.g. Muni Toke (2017) for a recent survey on order book reconstruction,

[1] A characterization of the realized price volatility at various frequencies.

Table 1. Event types definitions.

Notation	Definition
M, L, C, O	market order, limit order, cancellation, any order.
M^0, L^0, C^0, O^0	market order, limit order, cancellation, any order, that does not change the mid-price.
M^1, L^1, C^1, O^1	market order, limit order, cancellation, any order, that changes the mid-price.
M_{buy}, M_{sell}	buy/sell market order.
M^0_{buy}, M^0_{sell}	buy/sell market order that does not change the mid-price: i.e. order quantity $<$ best ask/bid available quantity.
M^1_{buy}, M^1_{sell}	buy/sell market order that changes the mid-price: ie. order quantity $\geq$ best ask/bid available quantity.
L_{buy}, L_{sell}	buy/sell limit order.
L^0_{buy}, L^0_{sell}	buy/sell limit order that does not change the mid-price: i.e. order price $\leq$ / $\geq$ best bid/ask price.
L^1_{buy}, L^1_{sell}	buy/sell limit order that changes the mid-price: ie. order price $>$ / $<$ best bid/ask price.
C_{buy}, C_{sell}	buy/sell cancellation.
C^0_{buy}, C^0_{sell}	buy/sell cancellation that does not change the mid-price: i.e. partial cancellation at best bid/ask limit or cancellation at another limit.
C^1_{buy}, C^1_{sell}	buy/sell cancellation that changes the mid-price: ie. total cancellation of best bid/ask limit order.

some data cleaning was involved in order to identify limit orders, market orders and cancellations given the states of the order book and the list of trades.

Due to the large quantity of data, problems such as mismatches of quantities and lack of synchronization were expected. However, such anomalies represent less than 3% of the data, and our results are thus reliable.

2.2. *Event definitions*

In this study, any change that modifies the best limits of the order book is called an "event"[1]. More precisely, an event can be a limit order, a market order, or a cancellation, and can affect the best bid or best ask. Moreover, events will be tagged according to whether they change the mid-price or not. **Table 1** summarizes the definitions and notations for the various event types considered in this paper.

2.3. *Statistical dependencies between order book events*

Table 2 represents the empirical probabilities of occurrence of an event of type j (in column), **conditioned** on the fact that the last observed event is of type i (in row). The last row represents the unconditional probabilities of each type of events.

To simplify the interpretation of the results, **Table 3** represents the ratio of conditional probabilities to unconditional probabilities, rounded to one decimal. It aims at revealing the mutual

[1]This simplifying choice essentially means that a level-1 order book is considered.

Table 2. Conditional probabilities of occurrences per event type.

	L^0_{buy}	L^0_{sell}	C^0_{buy}	C^0_{sell}	M^0_{buy}	M^0_{sell}	L^1_{buy}	L^1_{sell}	C^1_{buy}	C^1_{sell}	M^1_{buy}	M^1_{sell}
L^0_{buy}	27.94	9.73	26.24	20.99	1.09	0.50	4.98	1.78	0.08	3.60	2.65	0.44
L^0_{sell}	9.63	28.43	20.36	26.50	0.58	1.01	1.83	4.94	3.55	0.08	0.43	2.68
C^0_{buy}	21.64	24.72	29.33	7.51	0.66	0.57	2.92	4.28	4.97	1.32	0.97	1.11
C^0_{sell}	24.04	21.97	7.45	29.82	0.64	0.58	4.30	2.85	1.32	5.00	1.08	0.95
M^0_{buy}	20.69	8.02	6.96	11.18	9.08	0.59	9.64	0.86	1.12	6.65	24.27	0.94
M^0_{sell}	7.19	20.72	10.18	6.90	0.64	9.63	0.72	9.39	6.67	1.10	0.93	25.92
L^1_{buy}	32.48	10.83	1.17	26.57	0.92	0.94	4.38	1.68	8.89	4.93	2.43	4.77
L^1_{sell}	10.24	33.61	26.27	1.14	0.94	0.90	1.71	4.32	4.88	8.97	4.54	2.47
C^1_{buy}	14.46	12.40	51.27	4.59	0.26	0.10	8.42	4.06	2.83	0.96	0.51	0.15
C^1_{sell}	11.85	14.60	4.32	52.25	0.10	0.22	3.96	8.41	0.91	2.77	0.14	0.48
M^1_{buy}	12.23	6.18	4.56	30.18	1.04	0.64	24.94	4.39	1.16	8.70	3.35	2.64
M^1_{sell}	5.93	12.67	29.80	4.63	0.71	1.09	4.36	24.68	8.88	1.13	2.59	3.52
O	19.93	20.42	20.23	20.74	0.79	0.73	4.02	3.99	2.93	2.95	1.62	1.65

Table 3. Conditional probability leverage.

	L^0_{buy}	L^0_{sell}	C^0_{buy}	C^0_{sell}	M^0_{buy}	M^0_{sell}	L^1_{buy}	L^1_{sell}	C^1_{buy}	C^1_{sell}	M^1_{buy}	M^1_{sell}
L^0_{buy}	1.4	0.5	1.3	1.0	1.4	0.7	1.2	0.4	**0.0**	1.2	1.6	0.3
L^0_{sell}	0.5	1.4	1.0	1.3	0.7	1.4	0.5	1.2	1.2	**0.0**	0.3	1.6
C^0_{buy}	1.1	1.2	1.4	0.4	0.8	0.8	0.7	1.1	1.7	0.4	0.6	0.7
C^0_{sell}	1.2	1.1	0.4	1.4	0.8	0.8	1.1	0.7	0.5	1.7	0.7	0.6
M^0_{buy}	1.0	0.4	0.3	0.5	**11.5**	0.8	**2.4**	0.2	0.4	**2.3**	**15.0**	0.6
M^0_{sell}	0.4	1.0	0.5	0.3	0.8	**13.2**	0.2	**2.4**	**2.3**	0.4	0.6	**15.7**
L^1_{buy}	1.6	0.5	**0.1**	1.3	1.2	1.3	1.1	0.4	**3.0**	1.7	1.5	**2.9**
L^1_{sell}	0.5	1.6	1.3	**0.1**	1.2	1.2	0.4	1.1	1.7	**3.0**	**2.8**	1.5
C^1_{buy}	0.7	0.6	**2.5**	0.2	0.3	**0.1**	**2.1**	1.0	1.0	0.3	0.3	**0.1**
C^1_{sell}	0.6	0.7	0.2	**2.5**	**0.1**	0.3	1.0	**2.1**	0.3	0.9	**0.1**	0.3
M^1_{buy}	0.6	0.3	0.2	1.5	1.3	0.9	**6.2**	1.1	0.4	**2.9**	**2.1**	1.6
M^1_{sell}	0.3	0.6	1.5	0.2	0.9	1.5	1.1	**6.2**	**3.0**	0.4	1.6	**2.1**

relationships between events, and ratios greater than two are highlighted.

Results of **Table 3** are quite symmetric and no significant differences are observed between the buy and the sell side. Therefore, only the buy side is interpreted in detail below:

- $\boldsymbol{L^0_{buy}}$: adds liquidity to the first limit, signalling an increase of market demand at the current price level. This stimulates L^1_{buy} and M^1_{buy} events, based on the new consensus for a higher price. The corresponding probabilities for orders of type 0 are also increased, based on a similar reasoning but in a less aggressive way. On the other hand, the selling activity decreases in general except for C^1_{sell}, because some newly added limit orders may be cancelled shortly after. One notable thing is the sharp decrease of C^1_{buy}, as the newly added limit order probably comes from another trader, making it very unlikely that the first limit should be cancelled.
- $\boldsymbol{C^0_{buy}}$: decreases liquidity on the buy side. It triggers successive cancellations C^0_{buy} and C^1_{buy}: cancellations tend to follow themselves. M^0_{buy}, M^0_{sell}, M^1_{buy} and M^1_{sell} become less likely, revealing the influence of low liquidity on the participants' willingness to generate executions.
- $\boldsymbol{M^0_{buy}}$: largely increases the probability of M^0_{buy} and M^1_{buy}. This is commonly attributed to order splitting and the momentum effect (other participants following the move). L^1_{buy} and C^1_{sell} are also stimulated as a new price consensus emerges.
- $\boldsymbol{L^1_{buy}}$: improves the offered price to buy. The first effect is a strong increase in the probability of M^1_{sell}, i.e., participants entirely consume the new liquidity as the offered price has become higher. The second effect is an increase in the probability of C^1_{buy}, i.e., the new liquidity is rapidly cancelled. This is consistent with a similar observation made for L^0_{buy} orders, and might reflect some sort of market manipulation where agents are posting fake orders. Not

surprisingly, the conditional probability of C^0_{buy} is almost zero, because after one limit order is submitted, it cannot be partially cancelled. The fact that probability is not exactly 0 may be due to poor data synchronization, or the existence of hidden liquidity.

- $\boldsymbol{C^1_{buy}}$: a total cancellation of the best buy limit increases the probability of C^0_{buy} - order cancellations come in succession as market makers lose interest to provide liquidity even at the new best limit - and that of L^1_{buy} events, as traders may re-offer at the previous best price to gain priority. Events of other types become less frequent.
- $\boldsymbol{M^1_{buy}}$: consumes all the offered liquidity at the best ask. It stimulates L^1_{buy} and C^1_{sell} as a higher price consensus emerges among market participants. The probability of M^1_{buy} increases, indicating a short term momentum effect, and order splitting.

As a conclusion to this empirical section, let us just say that strong temporal dependencies between events are identified. Some orders actually trigger other events, a fact that can be seen as self- or cross-excitation phenomena. There are also some inhibition effects, when incoming orders prevent other events to occur. These two important features will be the target of the modelling approach presented in the next section.

3. Modelling dependencies using Hawkes processes

It is now widely accepted in the high frequency and market microstructure community that limit order books are worth modelling, and that the price dynamics can easily be extracted from that of the order book. In fact, the complexity of inter-event dependencies is so high that most significant features of the price dynamics : co-existence of time scales, leverage effect, signature plot, long term diffusivity... can be derived from advanced order book models.

In this section, point-process-based order book models are studied, building on the 12 event types previously introduced: $E = \{L^0_{buy}, L^0_{sell}, C^0_{buy}, C^0_{sell}, M^0_{buy}, M^0_{sell}, L^1_{buy}, L^1_{sell}, C^1_{buy}, C^1_{sell}, M^1_{buy}, M^1_{sell}\}$, where events with superscript 1 have an instantaneous price impact. In particular, it is clear that events in $E_{up} = \{L^1_{buy}, C^1_{sell}, M^1_{buy}\}$ lead to a price increase, whereas those in $E_{down} = \{L^1_{sell}, C^1_{buy}, M^1_{sell}\}$ result in price decrease.

The arrival of order book events is therefore modelled by a 12-variate simple point process $N(t) = (N_{L^0_{buy}}(t), \ldots, N_{M^1_{sell}}(t))$. Of interest is the associated intensity process $(\lambda_{L^0_{buy}}(t), \ldots, \lambda_{M^1_{sell}}(t))$. Assuming that the process is simple means that two events cannot occur at the same time, a fairly realistic assumption due to the high time resolution of modern stock exchanges.

Since the focus in this paper is on temporal interdependencies, N is actually a 12-variate **counting process**, and the marks determining the price jump when an event of type 1 occurs are not modelled. Rather, a simplifying assumption is made, namely, that the jump of the best bid or ask price following an event of type 1 is always one tick. This approximation reduces the dimensionality of the point process, while being consistent with the real behaviour of the chosen data set, for which the average jump size of the best bid and ask prices is 1.08 ticks[1]. Under this assumption, the reconstructed mid-price dynamics easily obtains as a by-product of event arrivals:

$$S(t) = S(0) + \left(\sum_{e \in E_{up}} N_e(t) - \sum_{e' \in E_{down}} N_{e'}(t) \right) \times \frac{\eta}{2}, \quad t > 0$$

where $\eta > 0$ is the tick size.

This simplification will be taken into account when comparing the performances of the model with the behaviour of real data.

[1]Actually, for some large tick stocks, the average is even smaller than 1.01.

Clearly, events of type 0 do not directly influence the price, rather, their impact will come from their influence on the intensities of the type 1 event arrival process.

As already said in the introduction, it has long been recognized that the class of Hawkes processes is particularly well suited to the modelling of point processes interacting *via* their conditional intensities. In the context of limit order book and high-frequency financial data, the paper Bacry *et al.* (2016) is a very interesting and relevant application of Hawkes processes to level-one limit order book modelling.

Let us now build on the results of Section 2 and study two classes of models, based respectively on linear and nonlinear Hawkes processes, that capture well the main characteristics of market dynamics. The performances of the models are presented in this section, while some more technical aspects pertaining to their numerical calibration are deferred until Section 4.

3.1. *Linear Hawkes process models*

In this short paragraph, we recall some essential definitions and results on the particularly interesting class of point processes introduced in Hawkes and Oakes (1974). We refer the interested readers to Brémaud and Massoulié (1996), Massoulié (1998) for a more in-depth presentation of these processes, and to Zheng *et al.* (2014), Abergel and Jedidi (2015) Bacry *et al.* (2016) for recent applications to order book modelling.

A multivariate point process $((T_i, X_i))_{i\in\mathbb{N}_*}$, associated to a counting process $(N(t))_{t\in\mathbb{R}_+} = (N_1(t), \ldots, N_M(t))_{t\in\mathbb{R}_+}$ with conditional intensity process $(\lambda(t))_{t\in\mathbb{R}_+} = (\lambda_1(t), \ldots, \lambda_M(t))_{t\in\mathbb{R}_+}$, is called a (linear, multivariate) **Hawkes process** Hawkes and Oakes (1974), Massoulié (1998) if there holds for $m \in \{1, \ldots, M\}$:

$$\lambda_m(t) = \mu_m + \sum_{n=1}^{M} \int_0^t \phi_{mn}(t-s)dN_n(s)$$

where μ_m are positive real numbers and ϕ_{mn} are nonnegative functions.

The μ_m are the *base intensities* and can be viewed as background intensities. Whenever an event occurs, the intensities increase, making subsequent events arrive at a higher frequency. Such effects are controlled by ϕ_{mn}. The functions ϕ_{mn}, the *kernel functions*, control the instantaneous increases and the relaxation speeds of the intensities in response to excitations.

For a multivariate Hawkes process, ϕ_{mm} describe the self-excitations, while ϕ_{mn} for $m \neq n$ measure the cross- (or: mutual) excitations, that is, the impact of an event of type n on the arrival of an event of type m.

A convenient, alternate way to express the intensity process is provided by the following equation:

$$\lambda(t) = \mu + \Phi \star dN \tag{1}$$

where $\Phi(t)$ is the $M \times M$ matrix whose entries are $\phi_{mn}(t)$, "$\star$" denotes the "matrix convolution"

$$\Phi \star dN = \int_{\mathbb{R}} \phi(t-s)dN(s)$$

and $\phi(t-s)dN(s)$ stands for the standard matrix-vector product.

It is clear that Hawkes processes are fully determined by their baseline intensity μ and the matrix Φ of kernel functions. In the following, we will concentrate on **exponential** kernels. This particular choice is classical, one of its main advantages being the Markovianity of the joint process (N, λ), see e.g. Massoulié (1998). For the models considered in this work, the intensities follow Equation (1), where Φ is a 12×12 kernel function matrix describing the excitation between events of various types:

$$\Phi = (\phi_{ij})_{i,j \in E}.$$

What we call *1-exponential* and *2-exponential* Hawkes models simply differ in the number of exponential functions used to define each kernel, namely:

- For the 1-exponential Hawkes model, $\phi_{ij}(t) = \alpha_{ij} \exp(-\beta_{ij} t)$
- For the 2-exponential Hawkes model, $\phi_{ij}(t) = \sum_{p=1}^{2} \alpha_{ijp} \exp(-\beta_{ijp} t)$

3.2. *Performances of the linear Hawkes models*

The adequacy of a linear Hawkes-process-based order book model is now evaluated, according to two criteria: a **goodness-of-fit** criterion for the distribution of forward recurrence times, and a criterion based on the **signature plot** generated by the model.

As a matter of fact, it is generally agreed upon that such statistical properties of the price process as the unconditional distribution of returns or the diffusive behaviour at large time scale, can easily be reproduced even with simpler models, whereas the signature plot and the inter-event durations offer a better challenge to discriminate among order book models.

3.2.1. **Goodness of fit.** It is well-known, see e.g. Bowsher (2007) that the transformed durations $\{\tau\}_i$ of a Hawkes process

$$\tau_i^m = \int_{T_i}^{T_{i+1}} \lambda^m(s) ds$$

are i.i.d. exponential random variables with parameter 1. This property is used to test the goodness-of-fit of the model to the data, by drawing Q-Q plots of the empirical quantiles with respect to the theoretical exponential distribution quantiles.

Though a global test can be conducted by concatenating all the transformed durations, plotting each dimension separately provides more information. This can be viewed as a marginal distribution fit test, i.e.: Given the law of other types of orders, how well can we fit the order under scrutiny ?

The procedure is as follows: first, the parameters for several order book models (Poisson, 1-exponential linear Hawkes, 2-exponential linear Hawkes) are calibrated, for each day in the study period. Then the transformed durations in the model are computed, and a Q-Q plot test is then performed. The results are shown in **Figure 2**.

As a first conclusion, one can easily see that a Poisson-process-based model globally fails to capture the distributional properties of recurrence times. The performances of the 1- and 2-exponential Hawkes models are similar, except for orders of type 0: the 2-exponential model significantly outperforms the 1-exponential model for L^0 and, to a lesser extent, for C^0 events. However, what is annoying is the behaviour for C^1 events: the distributions of the transformed durations in 1- and 2-exponential models are extremely close to one another, but neither is close to the theoretical exponential distribution.

This is an important, negative feature of the linear Hawkes models that will be revisited in the upcoming Subsection 3.3

3.2.2. **Signature plots.** The *signature plot* reveals some of the most important stylized facts about high frequency financial data. It is a plot of the realized variance as a function of the sampling frequencies.

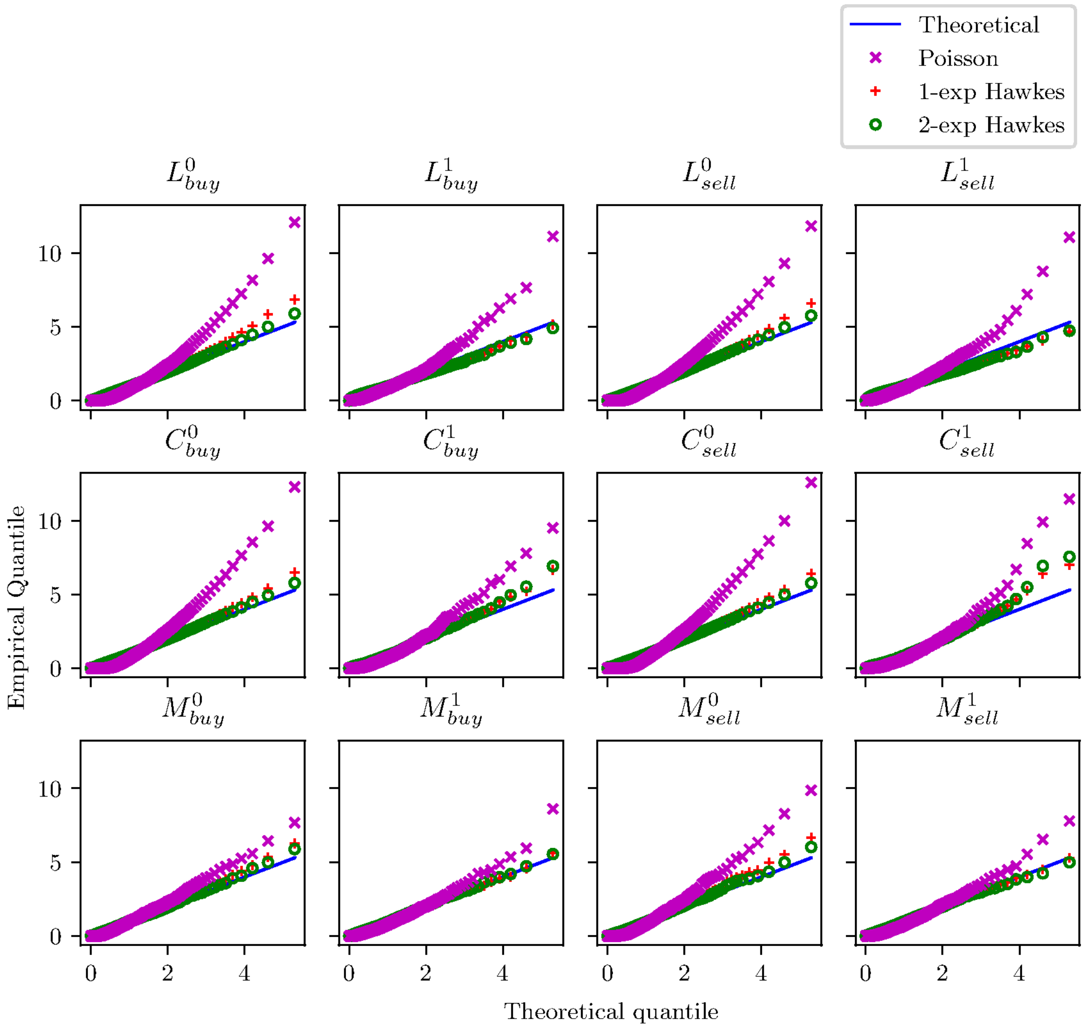

Figure 2. Q-Q plot goodness of fit tests of order book models.

The realized variance for a stochastic process X_t over a time period $[0, T]$ at a sampling frequency h is simply

$$RV(h) = \frac{1}{T}\sum_{n=0}^{T/h}(X((n+1)h) - X(nh))^2. \tag{2}$$

An important stylized fact of financial markets is that the quantity RV generally increases when h becomes small. This phenomenon is associated to the mean reverting behaviour of the price at short time scales. It has long been observed and was already reported in Andersen et al. Andersen *et al.* (2000). It is noteworthy that the signature plot becomes even steeper when computed on transaction prices rather than mid-prices because of the *bid-ask bounce*, and we will focus on mid-prices to avoid this spurrious effect.

Once the model parameters are calibrated, the mid-price is easily simulated using Equation (3). Realized variances are calculated with sampling periods from 1 to 50 seconds, with a step of 1 second.

The results for the models and the real data are shown in **Figure 3**.

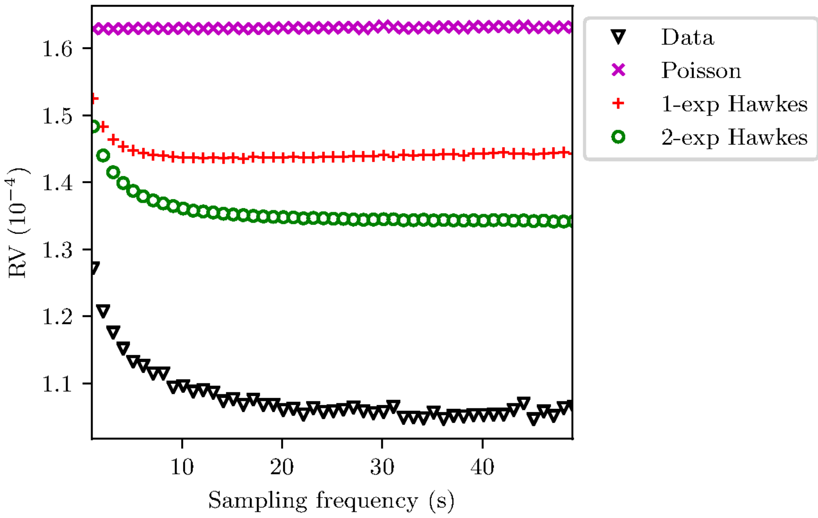

Figure 3. Mean signature plots of simulated price compared with real data.

Not surprisingly, the signature plot of the Poisson model is flat - this is expected, as the price dynamics in this model is that of a mid-price model with Poisson jumps, due to the mapping of orders that increase (resp. decrease) the price into upward (resp.downward) jumps.

The 1-exponential and 2-exponential Hawkes process models behave similarly: the realized volatility decreases when the sampling interval increases, but the long-term volatility level is too high compared to the data. Though reproducing the overall shape of the signature plot, the linear Hawkes-process-based order book models are not satisfactory.

3.3. *Nonlinear Hawkes process model*

This subsection addresses the shortcomings of linear Hawkes models in reproducing some characteristics of forward recurrence times and signature plots. Nonlinear Hawkes processes are introduced to overcome these difficulties, and their performances are studied.

3.3.1. **Order dependencies: inconsistencies between real data and linear Hawkes models.** The results presented in Paragraph 3.2.1 are now revisited in *event time*, temporarily ignoring the durations. When comparing the average conditional probability matrix of the 2-exponential Hawkes model with that of real data, one can check that most of the conditional probabilities are pretty close. However, for several pairs, there exist huge differences between the model and the real data, in particular for $C^1_{buy}|L^0_{buy}$, $M^0_{buy}|C^1_{buy}$ and $M^0_{sell}|C^1_{buy}$.

Table 3.3.1 below gives the list of all pairs (X, Y) for which the probability of an event of type X, conditioned on following an event of type Y, in the simulated order flow is either smaller than 50% or greater than 5 times the real conditional probability (only the buy side is shown, the sell side behaves similarly).

From a financial point of view, these discrepancies can easily be accounted for:

- A $C^1_{buy}|L^0_{buy}$ sequence almost never happens, because L^0_{buy} is a limit order added to the current first limit and it is highly unlikely that two orders should be cancelled at the same microsecond.
- The low probabilities of $L^1_{buy}|L^1_{buy}$ and $L^1_{sell}|L^1_{buy}$ comes from the constraint of the bid-ask spread: an aggressive limit order decreases the spread, and when the spread becomes one tick wide, other price-changing limit orders are no longer possible.

Table 4. Conditional probability comparison between simulated order flows and real data.

Pair	P_{simu}	P_{real}
$C^1_{buy}\|L^0_{buy}$	0.402	0.048
$L^1_{buy}\|L^1_{buy}$	1.628	0.141
$L^1_{sell}\|L^1_{buy}$	1.288	0.171
$M^0_{sell}\|C^1_{buy}$	0.545	0.068
$C^1_{buy}\|C^1_{buy}$	0.548	0.072
$M^1_{sell}\|C^1_{buy}$	0.854	0.037

Table 5. Medians of L_1 norm of kernels $\phi_{\cdot C^1_{buy}}$ in the 2-exponential model.

L^0_{buy}	L^0_{sell}	C^0_{buy}	C^0_{sell}	M^0_{buy}	M^0_{sell}	L^1_{buy}	L^1_{sell}	C^1_{buy}	C^1_{sell}	M^1_{buy}	M^1_{sell}
0.1563	0.2357	0.9392	0.0914	0	0	0.3845	0.1607	0	0	0.0013	0

- The remaining cases correspond to orders following a C^1_{buy} order that increases the bid-ask spread. There is no physical constraint preventing the spread from being wide, but participants in the market are not seemingly ready to sell when a cancellation order has already decreased the best bid price.

From a mathematical point of view, this poor fit comes from an inherent shortcoming of the linear Hawkes process model: the intensity for the arrival of an order of type e is written as

$$\lambda_e(t) = \mu_e + \sum_{e' \in E} \sum_{T_{e'} < t} \phi_{ee'}(t - T_{e'})$$

where $\phi_{e'} \geq 0$ and, by construction, $\lambda_e(t) < \mu_e$ cannot happen ! Consequently, **inhibition** effects, leading to a temporary decrease of certain short term conditional probabilities, are not modeled.

Note that, when calibrating the linear model (see Section 4 for details), the kernels corresponding to inhibitory behaviours are indeed forced to 0.

Below are the median values of the L_1 norms of the kernels stemming from the calibration results for C^1_{buy} stimulations in **Table 5**: clearly, kernels corresponding to the event pairs listed in **Table 3.3.1** have norms equal to 0.

Moreover, two other event pairs come out of the calibration with 0 kernel norms, $M^0_{buy}|C^1_{buy}$ and $C^1_{sell}|C^1_{buy}$. Although less obvious from the conditional probability matrix, this phenomenon is easy to interpret: a defensive cancellation on the bid side indicates a consensus of a fair price decrease in the market, therefore traders are less willing to buy at the previous ask price or cancel an existing ask order as it has already gained some queue priority with a profitable price.

3.3.2. **Model definition.** In order to incorporate inhibitory behaviours in the model, negative kernels are introduced in the Hawkes process. Then, a truncation is applied to avoid meaningless negative process intensities.

In the new model, the intensities satisfy the equation

$$\lambda(t) = (\mu + \Phi \star dN)_+, \tag{3}$$

where the entries of the matrix Φ are no longer constrained to take on positive values, and $()_+$ denotes the elementwise positive part function.

When enriched with the nonlinearity, the 2-exponential Hawkes process model retains its Markovian nature, see e.g. Brémaud and Massoulié (1996)Zhu (2015) for general results on nonlinear

Hawkes processes. The negative kernels are chosen under the following form

$$\phi_{mn} = \sum_{p=1}^{2} -\alpha_{mnp} \exp(-\beta_{mnp} t)$$

where the α's and β's are nonnegative real numbers. Note that we fix the same sign for the two exponentials, in order to avoid overfitting - it is actually unexpected for interdependencies to have different time regimes, for example an inhibitory effect in the short term that would become an excitation in the long run.

3.4. *Performances of the nonlinear Hawkes models*

3.4.1. **Goodness of fit.** Similarly to the analysis presented in Paragraph 3.2.1, the Q-Q plots of the empirical quantiles with respect to those of the theoretical exponential distribution are shown on Figure 4.

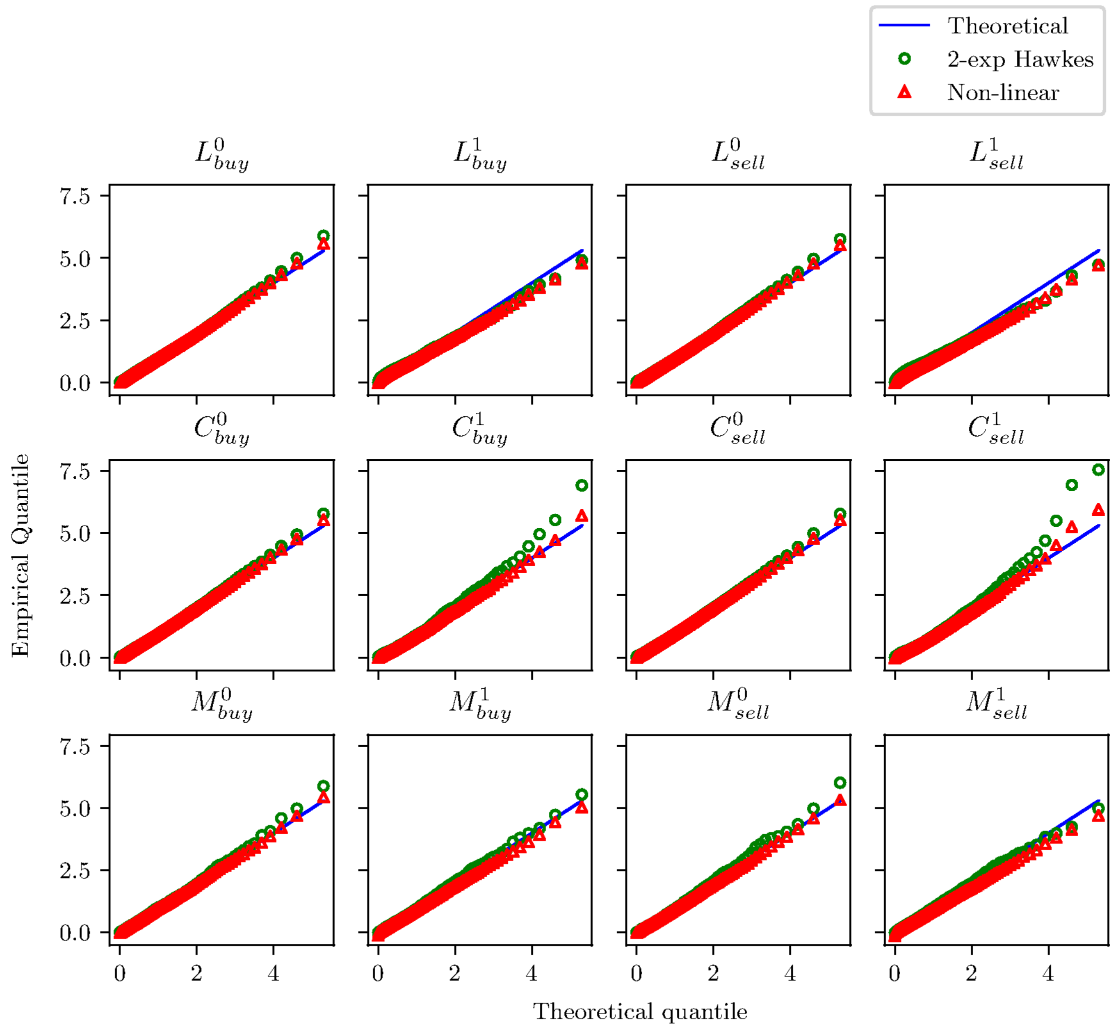

Figure 4. Q-Q plot goodness of fit tests of nonlinear Hawkes model.

It appears clearly, simply by eyeballing the graphs, that the nonlinear Hawkes model leads to a

Table 6. Improvement of conditional probabilities with the nonlinear Hawkes model.

Pair	P_{simu}	P_{simu}^{NL} [a]	P_{real}
$C_{buy}^1 \vert L_{buy}^0$	0.402	0.105	0.048
$L_{buy}^1 \vert L_{buy}^1$	1.628	0.254	0.141
$L_{sell}^1 \vert L_{buy}^1$	1.288	0.235	0.171
$M_{sell}^0 \vert C_{buy}^1$	0.545	0.135	0.068
$C_{buy}^1 \vert C_{buy}^1$	0.548	0.163	0.072
$M_{sell}^1 \vert C_{buy}^1$	0.854	0.108	0.037

[a] Conditional probabilities in nonlinear model.

Table 7. Medians of L_1 norm of kernels $\phi_{\cdot C_{buy}^1}$ in nonlinear model for those that were 0 in 2-exponential model.

M_{buy}^0	M_{sell}^0	C_{buy}^1	C_{sell}^1	M_{sell}^1
-0.0319	-0.1593	-0.0541	-0.1439	-0.1908

Table 8. Median of optimal likelihood functions for each type of order in 2-exponential and non-linear models.

Model	L_{buy}^0	L_{sell}^0	C_{buy}^0	C_{sell}^0	M_{buy}^0	M_{sell}^0
2-exp Hawkes	12862.2	14122.8	20018.4	21821.2	-2693.1	-1932.0
Non-linear Hawkes	12862.2	14122.8	20018.4	21821.2	-2584.9	-1784.6
	L_{buy}^1	L_{sell}^1	C_{buy}^1	C_{sell}^1	M_{buy}^1	M_{sell}^1
2-exp Hawkes	-1415.8	-1379.8	-1025.0	-1125.0	-1307.4	-1162.2
Non-linear Hawkes	-962.6	-999.4	-638.6	-803.0	-1099.0	-995.1

statistically more satisfactory fit than the linear 2-exponential Hawkes model previously studied. This better performance will be confirmed by the analysis of the signature plots and forward recurrence times.

In Tables 6 and 7, the calibration results of the nonlinear model are compared to those of the 2-exponential linear model. Clearly, the conditional probabilities are closer to real data, showing that a definite improvement is achieved by the nonlinear model. Also note that the kernels that were formerly set to 0 now take on quite significant negative norms, a fact which confirms that the inhibition effect plays an important role in the order dynamics.

Another interesting insight is provided in Table 8, by comparing the optimal values of the log-likelihood functions for both models. One can actually see where inhibition effects become more pregnant: orders of type 1 are clearly more influenced than orders of type 0, confirming the improvement already observed for the Q-Q plots in the goodness-of-fit test.

3.4.2. **Signature plots.** The signature plots of linear and nonlinear 2-exponential Hawkes models are shown in **Figure 5**, and compared to that of real data. The asymptotic volatility level significantly improves with the nonlinear model, and the resulting signature plot is overall a very good fit.

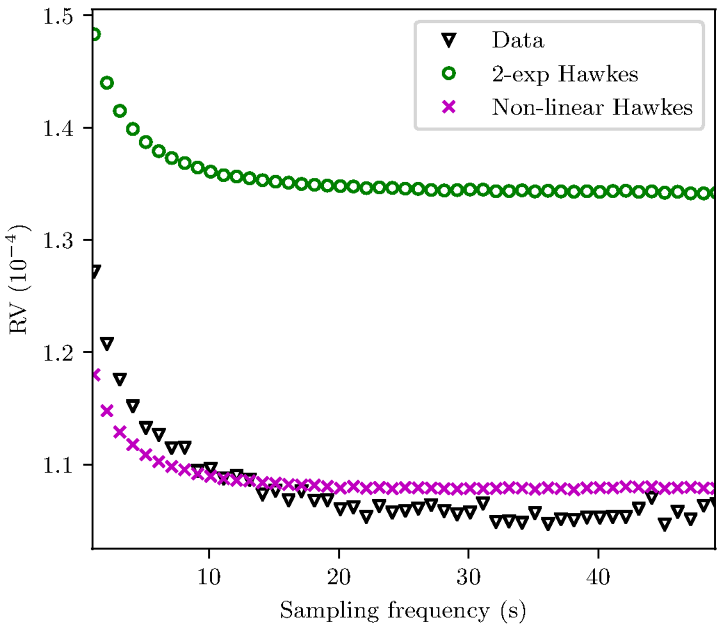

Figure 5. Mean signature plots of linear and nonlinear 2-exponential Hawkes process compared with real data.

3.4.3. **Analysis of self- and cross-excitation recurrence times.** The rationale behind the introduction of nonlinear Hawkes models was the empirically observed presence of inhibitory effects among events. As a consequence, one should hope that the inter-event recurrence times would behave in a more realistic way with these models.

Figure 6 and **7** show the cumulative distribution function (CDF) and the probability density of the (logarithm of) the forward recurrence times for all events of type 1 - that is, the forward recurrence times of (or: duration between) price jumps.

Specifically, define the inter-jump duration as

$$\Delta T_i = T_{i+1} - T_i$$

where T_i are the timestamps of the event arrivals.

According to the type of event causing the jump, these durations are furthermore separated into two subgroups: *self-excitation durations* $\Delta T^a \in \{\Delta T_i | X_i = X_{i+1}\}$ and *cross-excitation durations* $\Delta T^c \in \{\Delta T_i | X_i \neq X_{i+1}\}$.

Inter-jump durations predicted by the model are then computed, and compared to data: although the linear Hawkes model already performs well in reproducing the inter-jump duration distributions both for self- and cross-excitations, one can see that the nonlinear Hawkes process further improves the fit in the range between milliseconds and seconds ($\log_{10}(\Delta T) \in (-3, 1)$).

As a conclusion, one can say that the nonlinear Hawkes model provides a very satisfactory enhancement to the classical one, whether one uses Q-Q plots, signature plots or inter-jump recurrence times as benchmarks. This improvement is in fact quite natural, and is related to the empirical evidence presented in 3.3.1 on inhibition effects between events.

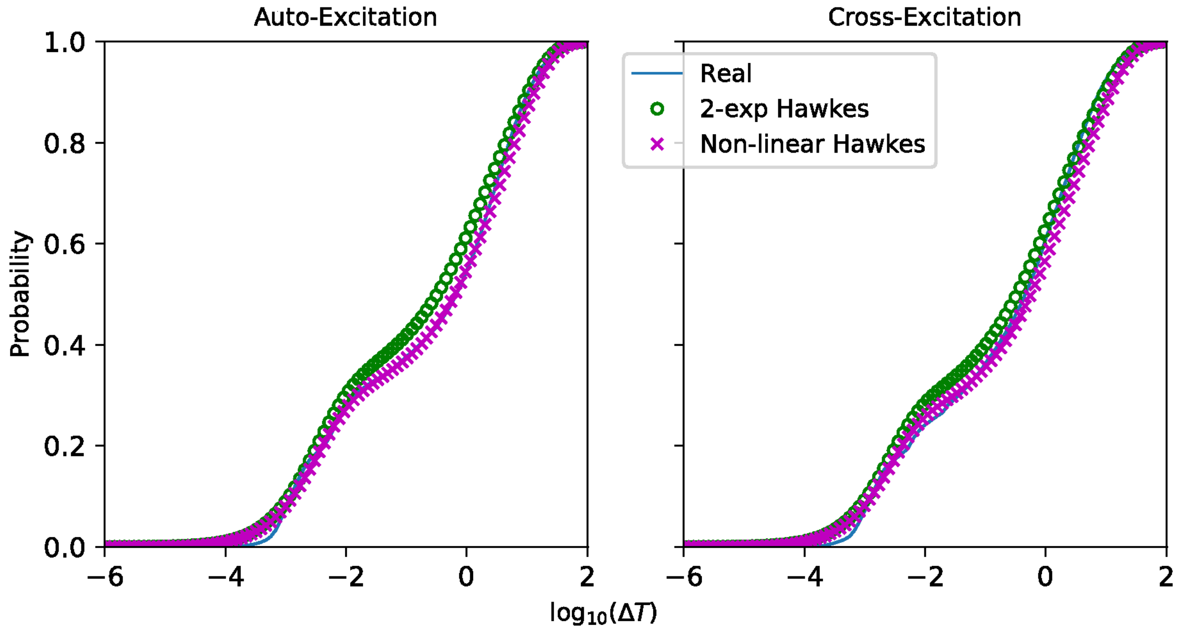

Figure 6. Cumulative distribution functions of log inter-jump durations for simulated price processes compared with real data.

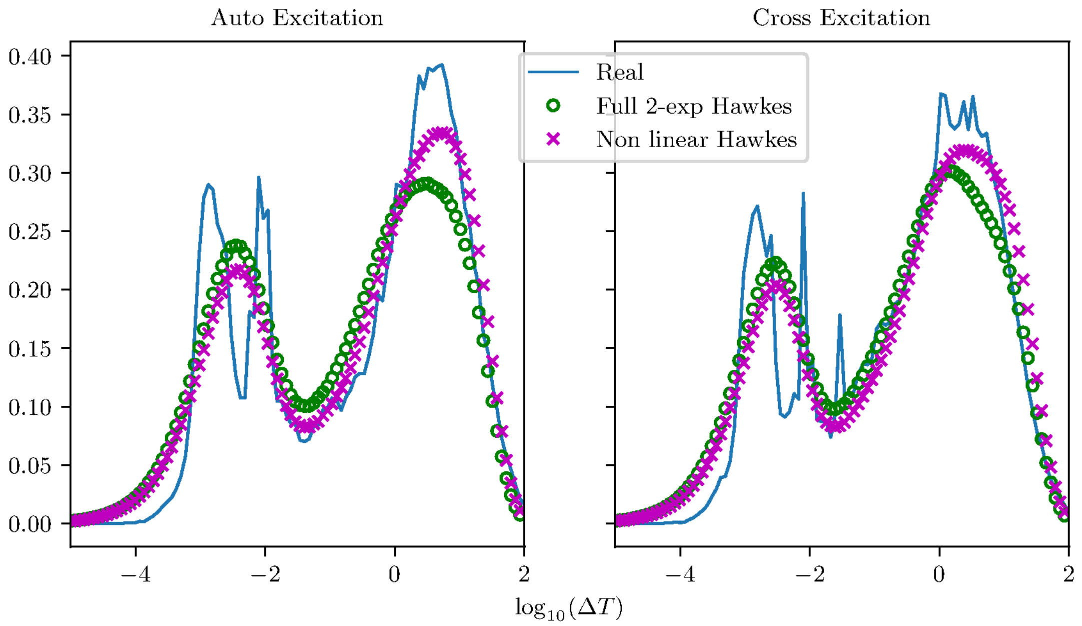

Figure 7. Probability density of log inter-jump durations.

4. Some numerical aspects of model calibration

This section is devoted to an analysis of the numerical algorithms used to calibrate the various models introduced in Section 3. Although rather technical, we think it is relevant - actually, very useful - for readers interested in calibrating high-dimensional Hawkes-processes to high frequency financial data (or other types of data).

Several optimization procedures are discussed and compared, and the best performer among those we have tested is thorougly investigated.

4.1. *Calibration with maximum likelihood estimation*

Let $((T_i, X_i))_{i\in\mathbb{N}^*}$ be a multivariate point process with associated counting process $(N_1(t), \ldots, N_M(t))$, whose intensities are to be estimated.

The log-likelihood, see Ozaki (1979)Rubin (1972), of given intensities $(\lambda_1(t), \ldots, \lambda_M(t))$, and a sample of observation $\{T_i, X_i\}_{i\in\{1,\ldots,M\}}$, is defined by the sum of the log-likelihood of each component:

$$\begin{aligned}\ln L(\lambda, \{T_i, X_i\}_{i\in\{1,\ldots,D\}}) &= \sum_m \ln L_m(\lambda_m, \{T_i, X_i\}_{i\le D}) \\ &= \sum_{m=1}^{M}\left[\int_0^T \ln\lambda_m(s)dN_m(s) + \int_0^T(-\lambda_m(s))ds\right].\end{aligned}$$

In the case of a Hawkes process with exponential kernels, a straightforward computation gives:

$$\int_0^T \ln\lambda_m(s)dN_m(s) = \sum_{T_i^m}\ln\left[\mu_m + \sum_{n=1}^{M}\alpha_{mn}A_{mn}(i)\right]$$

and

$$\int_0^T \lambda_m(s)ds = \mu_m T - \sum_{n=1}^{M}\sum_{T_k^n}\frac{\alpha_{mn}}{\beta_{mn}}\left(e^{-\beta_{mn}(T-T_k^n)} - 1\right),$$

where $A_{mn}(i) = \sum_{T_k^n<T_i^m} e^{-\beta_{mn}(T_i^m - T_k^n)}$ can be computed iteratively as

$$A_{mn}(i) = A_{mn}(i-1)e^{-\beta_{mn}(T_i^m - T_{i-1}^m)} + \sum_{T_{i-1}^m\le T_k^n<T_i^m} e^{-\beta_{mn}(T_i^m - T_k^n)}$$

so that

$$\ln L_m(\lambda_m, \{T_i, X_i\}_{i\le D}) = -\mu_m T + \sum_{n=1}^{M}\sum_{T_k^n}\frac{\alpha_{mn}}{\beta_{mn}}\left(e^{-\beta_{mn}(T-T_k^n)} - 1\right) + \sum_{T_i^m}\ln\left[\mu_m + \sum_{n=1}^{M}\alpha_{mn}A_{mn}(i)\right].$$

It is however clear, and quite unfortunate, that the likelihood function is not strictly concave. For example, in the 1-dimensional case, its expression simplifies to

$$\ln L(\lambda, \{T\}) = -\mu T + \sum_{T_i}\frac{\alpha}{\beta}\left(e^{-\beta(T_D - T_i)} - 1\right) + \sum_{T_i}\ln\left[\mu + \alpha\sum_{T_j<T_i} e^{-\beta(T_i - T_j)}\right],$$

and, letting β tend to ∞, there holds

$$\lim_{\beta\to+\infty}\ln L(\lambda, \{T\}) = -\mu T + N(T)\ln\mu,$$

which is finite. However, a strictly concave continuous function having a local maximum cannot tend to a finite limit at infinity.

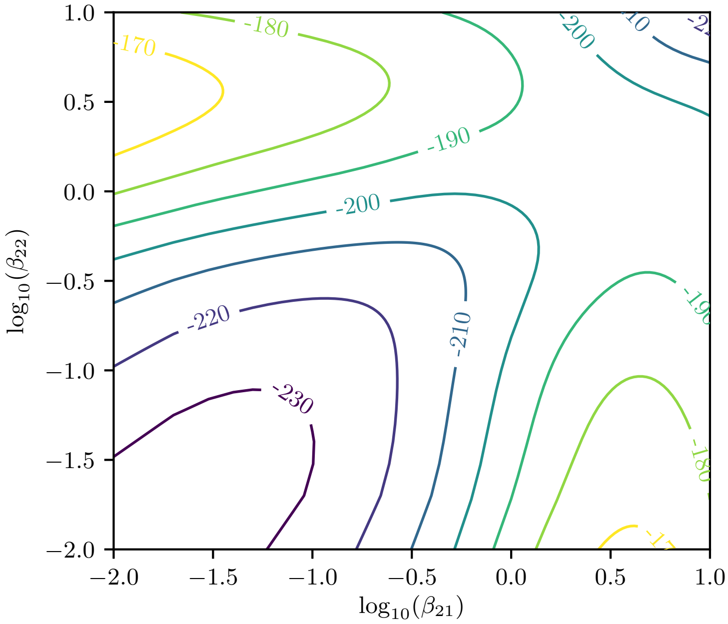

Figure 8. Example of local maxima in 2-d Hawkes process likelihood function $\ln L_2(\lambda_2)$

In fact, not only is the likelihood function not concave, but it actually has several local maxima. An illustrative example is given in **Figure 8** where we draw the contour plot of the partial likelihood function $\ln L_2$ of a simulated 2-dimensional Hawkes process. The kernels are exponential functions with parameters specified in Equation (4). While μ_2, α_{21} and α_{22} are kept fixed, the likelihood values are plotted as functions of β_{21} and β_{22}. The two axes are presented in logarithmic scale.

It is clear that there are at least two local minima in this example.

$$\mu = \begin{pmatrix} 0.1 \\ 0.2 \end{pmatrix} \qquad \alpha = \begin{pmatrix} 5.0 & 10.0 \\ 1.0 & 2.0 \end{pmatrix} \qquad \beta = \begin{pmatrix} 20.0 & 15.0 \\ 3.0 & 10.0 \end{pmatrix} \tag{4}$$

The existence of several local maxima make gradient-type algorithms less relevant for the maximum likelihood procedure and a global optimization algorithm appears necessary. The **Nelder-Mead** simplex algorithm (*NM*) has been widely used in previous works on the calibration of Hawkes processes; however we find it not stable enough when a good a priori guess is not available.

For these reasons, the **Differential Evolution** algorithm (*DE*) Storn and Price (1995) has been chosen to perform the optimization. *DE* is an efficient genetic evolutionary algorithm that has been adopted in various engineering domains such as electrical power systems, artificial neural networks, operation research, image processing... Starting from a population of randomly generated points, the algorithm performs a mutation-crossover-selection procedure, where the population is updated to have better objective function values and a large tentative space is scanned.

A pseudocode is given in Appendix A.

4.2. *Benchmarking the DE algorithm*

Simulation-based numerical experiments are performed in order to compare the efficiency of the *NM* and *DE* algorithms. More specifically, we consider a 2-dimensional Hawkes process

Table 9. Error rate (%) of calibration by Nelder-Mead algorithm and Differential Evolution algorithm.

Algorithm	T	μ_1	α_{11}	α_{12}	β_{11}	β_{12}	μ_2	α_{21}	α_{22}	β_{21}	β_{22}
DE	250	0.0	0.0	0.0	0.0	0.0	0.6	1.0	2.0	1.6	2.2
	2500	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.1
	25000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NM random	250	2.1	2.1	2.0	2.0	2.1	24.5	27.6	30.7	24.8	30.5
	2500	1.4	1.4	1.4	1.4	1.4	12.1	14.8	16.9	14.6	18.8
	25000	1.0	1.0	0.8	1.0	1.0	9.9	12.6	14.0	11.4	16.7
NM perfect	250	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	2500	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	25000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

where the parameters are specified in (4). 100 process paths are simulated for each $T \in \{100, 250, 500, 1000, 2500, 5000, 10000, 25000\}$, and the parameters are calibrated from each simulated path with various algorithms.

NM is used with different initialization methods. For *NM random*, the initial reference point is drawn from uniform distributions. Denoting by ρ the L_1 norm of the kernel ($\rho = \frac{\alpha}{\beta}$), we choose

$$\mu \sim \mathcal{U}(0,1) \qquad \rho \sim \mathcal{U}(0,1) \qquad \beta \sim \mathcal{U}(0,100) \tag{5}$$

and optimize with respect to ρ instead of α.

The algorithm *NM perfect* refers to *NM* where the true input parameters are used as reference point.

The empirical probability of error for each optimization algorithm is show in **Table 9** for $T \in \{250, 2500, 25000\}$:

Clearly, with the possible exception of short time horizon, *DE* almost always finds the optimal point, getting very close to the *NM perfect* algorithm.

4.3. *Improvement in high dimensions*

The local maximum problem is more severe when dealing with higher dimension and real data instead of simulated data. In this section, we present some treatments designed to mitigate the numerical issues and boost the convergence towards a global maximum.

4.3.1. **Some evolutions of the *DE* algorithm: a quick guided tour.** Thanks to its wide variety of applications, *DE* has attracted a lot of interest, and the recent survey paper Das *et al.* (2016) documents a host of novel ideas to improve its classical form. Below is a brief summary of some of the proposed improvements (notations are those used in Algorithm 1):

- **Mutation strategy.** The donner vector $v_{i,g}$ in mutation can be generated with different strategies. The classical algorithm adopts a so-called "DE/rand/1" strategy

$$v_{i,g} = x_{r_1,g} + F(x_{r_2,g} - x_{r_3,g})$$

where r_1^g, r_2^g and r_3^g are mutually exclusive integers randomly chosen in $[\![1, N]\!]$ $\{i\}$. It could be preferable to approach the current best value

$$v_{i,g} = x_{best,g} + F(x_{r_2,g} - x_{r_3,g})$$

or use more points for deviation

$$v_{i,g} = x_{r_1,g} + F(x_{r_2,g} - x_{r_3,g}) + F(x_{r_4,g} - x_{r_5,g})$$

Combinations of these ideas are of course possible, which create vast candidate strategies.

- **Crossover.** Apart from the idea of the *binomial/uniform* crossover, another method called *exponential* crossover is also considered. The trial vector u takes the value of the donner vector v for adjacent coordinates. The benefit is limited to special structures of problems where neighboring variables are linked but relatively independent of other variables. As a result the *binomial* crossover is more frequently used.
- **Adaptation of control parameter (F and CR) and strategy.** It aims at adding learning performances to the offspring generation. Either the strategies are randomly chosen from fixed ensemble of strategies and parameters, which are designed to aid the algorithm to converge or explore larger space so that the combination can balance the two effects; or the mutation strategy is fixed, but the parameters can adapt to the evolution.
- **Population control.** The most natural idea is the reduction of population as they approach to each other and concentrate in a small region. Such reduction can be pre-scheduled or dynamically controlled based on the computational budget. On the other hand, varied population (instead of monotonically decreasing) is also introduced as a choice to adapt to the evolution of the algorithm.

Other extensions actually go beyond the classical framework, for example using new initialization techniques, adding clustering technique for the sub-population topology, and so on. Hybridization opens another branch of research: on the one hand DE is combined with other heuristic methods to explore the advantages of exploration strategies, and on the other hand, local search methods are injected into the DE algorithm to boost convergence and precision.

In the interest of tractability, we choose to concentrate on the non-hybrid extensions. In Das *et al.* (2016), the algorithm L-SHADE is reported to have the "best competitive performance among non-hybrid algorithms at the CEC 2014 competition on real parameter single-objective optimization". Compared to the classical algorithm, L-SHADE combines adaptation in every respect - mutation, parameter control and population control:

- **Mutation** use the $current-to-pbest/1$ strategy, where the new donner vectors are obtained by

$$v_{i,g} = x_{i,g} + F_i(x_{pbest,g} - x_{i,g}) + F_i(x_{r_1,g} - x_{r_2,g}),$$

where $x_{pbest,g}$ is randomly selected from the best $\lfloor pN \rfloor$ members in generation g, where ($p \in [0,1]$) . This strategy exhibits some greediness towards the current best points, but the existence of p leaves the flexibility for tradeoff between exploitation and exploration.
- **Parameter control** In order to dynamically adapt the parameters F and CR, a record of past candidates is maintained. Two lists of size H, M_{CR} and M_F, are kept. For each generation, F_i and CR_i are drawn randomly with certain distributions depending on randomly chosen means from the lists:

$$F_i = randc_i(M_{F,r_i}, 0.1), \qquad CR_i = randn_i(M_{CR,r_i}, 0.1)\mathbb{1}_{\{M_{CR,r_i} \neq Null\}},$$

where $randn$ follows a normal distribution and $randc$, a Cauchy distribution. For each generation, the kth element ($k = g \bmod H$) of the list is updated, according to CR_i and F_i that succeed to find ameliorated points. Such mechanism introduces learning characteristics for the F and CR selection, in order to overcome the stagnation problem.
- **External archive introduction** To maintain diversity, a external archive is used so that parent vectors that are worse than the trial vectors are preserved in A. When generating

donner vectors, $x_{r_2,g}$ can be selected from $P \cup A$.

- **Linear population size reduction** The whole population N_g decreases according to the allowed total number of generations.

$$N_{g+1} = \text{round}\left(\left(\frac{N_{min} - N_{init}}{G}\right) * g + N_{init}\right),$$

where N_{init} is the classical initial population size, and N_{min} is the smallest possible population size for a mutation strategy.

The L-SHADE is a combination of interesting ideas. Roughly stated, the current-to-pbest/1 mutation helps approach the best candidates in the population, accelerating the convergence of the algorithm; the parameter control aims at learning the trade-off between exploration and exploitation; the external archive is to help keep diversity of the population so that exploration is partly internalized by the exploitation of the abandoned history; and the population size reduction saves computational cost to allow larger initial populations.

4.3.2. **Calibrating high dimensional Hawkes order book models.** Let us now turn towards the actual application of *L-shade* to the task at hand.

Starting from 100 different initial populations for each strategy with the same number of points and maximum generations, we plot the histograms of the final log-likelihood function values for one dimension of the 12-dimensional Hawkes model with real data in **Figure 9**, for different modifications of *DE*. The classical strategy, noted as "rand/1", serves as a reference for the suggested "current-to-pbest/1". The parameter adaptation is also combined with "rand/1" to provide better performances. We finally introduce a version with a refinement of the initial parameter intervals, noted as "better guess". The right subplot is a zoom of the one on the left, to further show the improvement due to "better guess".

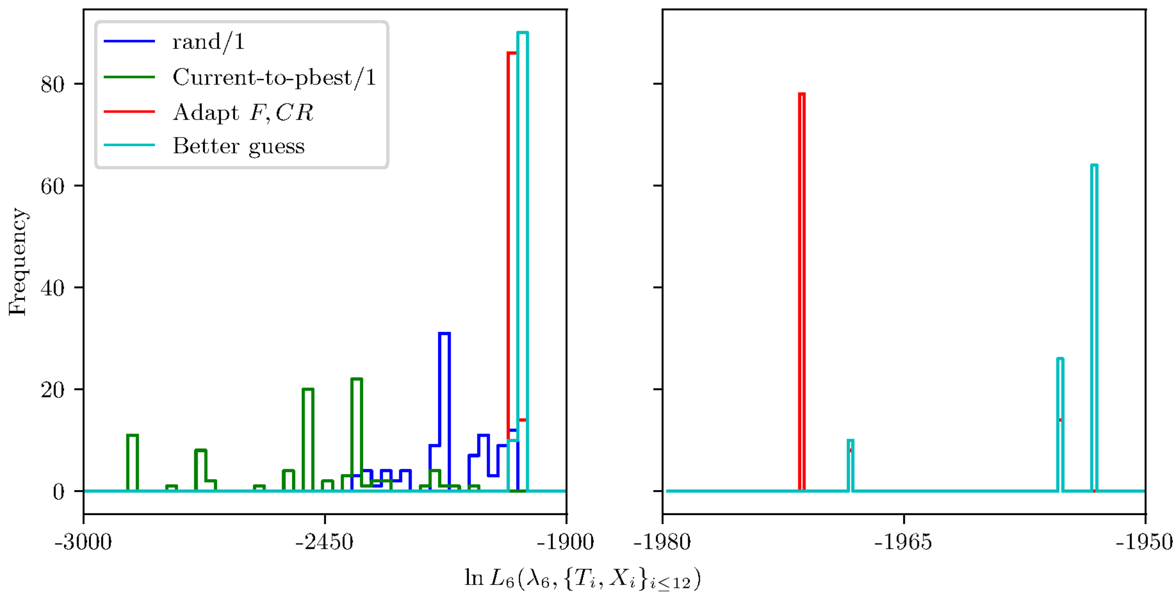

Figure 9. Distribution of optimal objective likelihood functions in different optimization strategy tests. The right one is zoomed at the optimal zone for further illustration.

Some comments are in order :

- In the "current-to-pbest/1" strategy, the closer p is to 0, the greedier the algorithm is, and the more probable it is that the optimization gets trapped at a local maximum. The closer p is to 1, the more the algorithm favors exploration.
- The learning mechanism for F and CR in the adaptative version leads to some improvements. Parameters are initialized according to the following distributions:

$$\mu_m \sim \mathcal{U}(0, \frac{0.2N_m}{T}), \quad \rho_{mn} \sim (0, \min(\frac{0.2N_m}{N_n}, 0.5)), \quad \beta_{mn} = u_1 \mathbb{1}_{\{u_0=0\}} + u_2 \mathbb{1}_{\{u_0=1\}}$$

$$\text{for } u_0 \sim \mathcal{B}(1, 0.5), \quad u_1 \sim \mathcal{U}(0,1) \quad u_2 \sim \mathcal{U}(0,100)$$

derived from the physical interpretation of μ as the baseline intensity, of ρ as the integrated intensity of the influence from event arrival, and based on the relation

$$\mathbb{E}[\lambda_{m,\infty}]T = \mu_m T + \mathbb{E}[\sum_n \rho_{mn} N_n].$$

- Although different runs starting from different initial populations do not converge to the global maximum, some improvement may be gained from a "better guess" of the initial intervals.

Clearly, an increase of the population size plays a major role in boosting the convergence: a larger population prevents points from getting trapped around the same local maximum. On the other hand, it is useless to keep all the population as the algorithm approaches the end of its iterations, since points tend to form clusters. As a consequence, it makes sense to consider effective population reduction techniques and use the saved computational budget to cover a larger search space.

Building on the linear population reduction method inspired by the combination of *DE* with clustering algorithms in Li and Zhang (2011) and the use of pairwise Euclidean distance for dynamic population control in Yang *et al.* (2013), we propose an additional reduction mechanism which allows, not only to decrease the function evaluation times, but also to avoid convergence to local maxima.

The algorithm is said to have converged if each coordinate of all the points in the population has converged. The convergence conditions of the coordinates are

$$\sigma(\mu_m) < e_r\langle\mu_m\rangle \quad \text{or} \quad \max \mu_m - \min \mu_m < e_a,$$

$$\sigma(\rho_{mn}) < e_r\langle\rho_{mn}\rangle \quad \text{or} \quad \max \rho_{mn} - \min \rho_{mn} < e_a,$$

$$\sigma(\beta_{mn}) < e_r\langle\beta_{mn}\rangle \quad \text{or} \quad \langle\rho_{mn}\rangle < e_a,$$

where $\sigma(.)$ and $\langle\cdot\rangle$ are the standard deviation and the mean value respectively, and e_r and e_a are the relative and absolute error tolerance. At each generation, we eliminate points that are close to the current b best ones, using a criterion similar to the termination conditions for the population: suppose the points are sorted according to their objective function values by descending order. For a given point x_i, if $\exists j \in [\![1, b]\!]/\{i\}$ such that all the following conditions are satisfied:

$$|\mu_{mi} - \mu_{mj}| < e_r\mu_{mj} \quad \text{or} \quad |\mu_{mi} - \mu_{mj}| < e_a$$

$$|\rho_{mni} - \rho_{mnj}| < e_r\rho_{mnj} \quad \text{or} \quad |\rho_{mni} - \rho_{mnj}| < e_a$$

$$|\beta_{mni} - \beta_{mnj}| < e_r \beta_{mnj} \quad \text{or} \quad \rho_{mnj} < e_a,$$

then x_i is eliminated from the population. In practice, it is convenient to select a small value for b. The decrease of population size saves some computational budget for the algorithm, which is very beneficial as the computation of the likelihood function is costly.

The combination of these population reduction techniques allows to increase the initial population by a factor of 5 to 10 with no significant impact on the total computation time, and the convergence is largely improved.

As a conclusion, one can say that the improved version *L-SHADE* of the *DE* algorithm drastically enhances the performances of the calibration, but despite all these efforts, we are still left with an average failure rate of approximately 5%.

5. Conclusion

This paper is a study of Hawkes processes applied to high frequency limit order book data. Suitably designed nonlinear Hawkes processes that include inhibitory effects and a co-existence of time scales are shown to successfully model the dependencies between the arrival of order book events. Thanks to the particularly well-suited distinction between events that trigger, or do not trigger, an immediate change in the current price, the dynamics of the model fully reflect that of the price. Such a description helps cope with some shortcomings of order book models that were previsouly observed, particularly concerning the realized spot price volatility.

The paper also gives a detailed analysis on a very important, albeit technical, topic: the choice of the optimization algorithm for the Maximum Likelihood Estimator. The *L-SHADE* algorithm is a significant improvement over the classical *Differential Evolution* algorithm, thanks to better initializations and population control.

As a conclusion, one can say that nonlinear Hawkes processes capture well such fundamental features of market dynamics as conditional probabilities, forward recurrence times, or the signature plot. They provide an accurate description of the order book in the high frequency realm, as well as a realistic behaviour of more macroscopic quantities. While leading to a better understanding of the mechanisms driving the markets, their use in the simulation of order driven markets can also lead to a host of potential applications.

References

Abergel, F., Anane, M., Chakraborti, A., Jedidi, A. and Muni Toke, I., *Limit order books*, 2016, Cambridge University Press.

Abergel, F. and Jedidi, A., Long-Time Behavior of a Hawkes Process-Based Limit Order Book. *SIAM Journal on Financial Mathematics*, 2015, **6**, 1026–1043.

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P., Great realizations. *Risk*, 2000, **13**, 105–108.

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

Bacry, E., Jaisson, T. and Muzy, J.F., Estimation of slowly decreasing Hawkes kernels: application to high-frequency order book dynamics. *Quantitative Finance*, 2016, **16**, 1179–1201.

Bacry, E., Mastromatteo, I. and Muzy, J.F., Hawkes processes in finance. *Market Microstructure and Liquidity*, 2015, **1**.

Bacry, E. and Muzy, J.F., Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance*, 2014, **14**, 1147–1166.

Bouchaud, J.P., Farmer, J.D. and Lillo, F., How Markets Slowly Digest Changes in supply and Demand. In *Handbook of Financial Markets: Dynamics and Evolution*, edited by T. Hens and K.R. Schenk-Hoppé, pp. 57–156, 2009, Elsevier, San Diego.

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

Brémaud, P. and Massoulié, L., Stability of nonlinear Hawkes processes. *Ann. Probab.*, 1996, **24**, 1563–1588.

Chakraborti, A., Muni Toke, I., Patriarca, M. and Abergel, F., Econophysics review: I. Empirical facts. *Quantitative Finance*, 2011, **11**, 991–1012.

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

Das, S., Mullick, S.S. and Suganthan, P., Recent advances in differential evolution–An updated survey. *Swarm and Evolutionary Computation*, 2016, **27**, 1–30.

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

Gopikrishnan, P., Plerou, V., Gabaix, X. and Stanley, H.E., Statistical properties of share volume traded in financial markets.. *Physical Review E*, 2000, **62**, R4493.

Hawkes, A.G. and Oakes, D., A Cluster Process Representation of a Self-Exciting Process. *Journal of Applied Probability*, 1974, **11**, 493–503.

Huang, W., Lehalle, C.A. and Rosenbaum, M., Simulating and analyzing order book data: The queue-reactive model. *Journal of the American Statistical Association*, 2015, **110**, 107–122.

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

Li, Y.l. and Zhang, J., A differential evolution algorithm with dynamic population partition and local restart. In *Proceedings of the Proceedings of the 13th annual conference on Genetic and evolutionary computation*, pp. 1085–1092, 2011.

Massoulié, L., Stability results for a general class of interacting point processes dynamics, and applications. *Stochastic Processes and their Applications*, 1998, **75**, 1–30.

Muni Toke, I., Reconstruction of Order Flows using Aggregated Data. *Market microstructure and liquidity*, 2017.

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

Rambaldi, M., Bacry, E. and Lillo, F., The role of volume in order book dynamics: a multivariate Hawkes process analysis. *Quantitative Finance*, 2016, pp. 1–22.

Rubin, I., Regular point processes and their detection. *IEEE Transactions on Information Theory*, 1972.

Smith, E., Farmer, J.D., Gillemot, L. and Krishnamurthy, S., Statistical theory of the continuous double auction.. *Quantitative Finance*, 2003, **6**, 481–514.

Storn, R. and Price, K., *Differential evolution-a simple and efficient adaptive scheme for global optimization over continuous spaces*, Vol. 3, , 1995, ICSI Berkeley.

Yang, M., Cai, Z., Li, C. and Guan, J., An improved adaptive differential evolution algorithm with population adaptation. In *Proceedings of the Proceedings of the 15th annual conference on Genetic and evolutionary computation*, pp. 145–152, 2013.

Zheng, B., Roueff, F. and Abergel, F., Modelling Bid and Ask Prices Using Constrained Hawkes Processes: Ergodicity and Scaling Limit. *SIAM Journal on Financial Mathematics*, 2014, **5**, 99–136.

Zhu, L., Large deviations for Markovian nonlinear Hawkes processes. *The Annals of Applied Probability*, 2015, **25**, 548–581.

Appendix A: Pseudocode of basic Differential Evolution

Algorithm 1 Differential Evolution algorithm

Input. Maximum total generation G, population size $N \geq 4$, mutation factor $F \in (0, 2)$, crossover rate $CR \in (0, 1)$, parameter domain Ω, termination criteria.
Output. optimal point (optimal function value, termination generation etc.).
// *Initialization phase*
g=1; Initialize the initial population $(x_{1,1}, \ldots, x_{N,1})$ randomly such that $x_{i,1} \in \Omega$;
while g $\leq$ G and termination criteria not met **do**
    for $i \leftarrow 1, N$ **do**
        // *Mutation*
        Choose randomly r_1, r_2 and r_3 in $[\![1, N]\!]$ such that i, r_1, r_2 and r_3 are distinct;
        Construct donner $v_{i,g+1} \leftarrow x_{r_1,g} + F(x_{r_2,g} - x_{r_3,g})$;
        // *Crossover. Construct trial element* $u_{i,g+1}$
        I_{rand} is a random integer from $[\![1, D]\!]$;
        for $j \leftarrow 1, D$ **do**
            $rand_{j,i} \sim \mathcal{U}(0, 1)$;
            if $rand_{j,i} \leq CR$ or $j = I_{rand}$ **then**
                $u_{j,i,g+1} \leftarrow v_{j,i,g+1}$;
            else
                $u_{j,i,g+1} \leftarrow x_{j,i,g}$;
            end if
        end for
        // I_{rand} *ensures that* $u_{i,g+1} \neq x_{i,g}$
        // *Selection*
        if $f(u_{i,g+1}) \leq f(x_{i,g})$ **then**
            $x_{i,g+1} \leftarrow u_{i,g+1}$;
        else
            $x_{i,g+1} \leftarrow x_{i,g}$;
        end if
    end for
    $g \leftarrow g + 1$;
end while