

# Hawkes processes for credit indices time series analysis: How random are trades arrival times?

Achraf Bahamou, Maud Doumergue, and Philippe Donnat

Hellebore Capital Ltd  
Michelin House, London

**Abstract.** Targeting a better understanding of credit market dynamics, the authors have studied a stochastic model named Hawkes process. Describing trades arrival times, this kind of model allows for the capture of self-excitement and mutual interactions phenomena. The authors propose here a simple yet conclusive method for fitting multidimensional Hawkes processes with exponential kernels, based on a maximum likelihood non-convex optimization. The method was successfully tested on simulated data, then used on new publicly available real trading data for three European credit indices, thus enabling quantification of self-excitement as well as volume impacts or cross indices influences.

**Keywords:** Point and counting processes, multidimensional Hawkes process, financial time series analysis, non-convex optimization, maximum likelihood optimization, credit indices.

## 1 Introduction

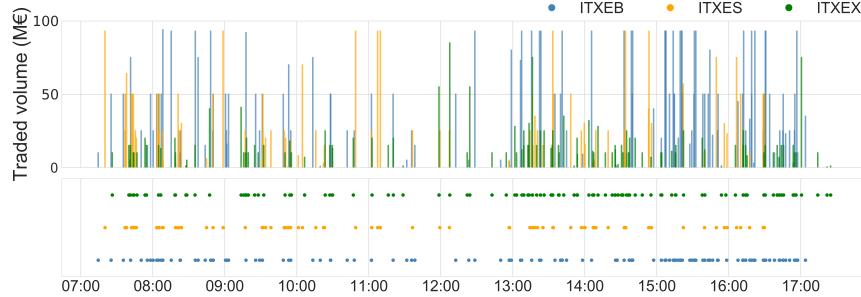
### 1.1 From credit derivative indices to Hawkes processes

Credit indices are financial instruments comprised of a set of credit securities, mainly used to hedge credit default risk. Each new index series is issued every 6 months and expires after a defined maturity. Though liquid, those indices are traded at a rather "mid-frequency" rate, with trades occurring at a minute scale. For European indices, the market is continuously open from 7:00 to 17:00, with a total of five to ten billion euros reported each day on average. Fig 1 gives an idea of the activity on a sampled day of trading. If recent regulations have led to a greater public reporting of trading activities, thus releasing more amount of traded data, this data is by essence quite sparse. Picturing such market behaviour over time represents a challenging opportunity: it originally motivated the work presented in this article.

This study focuses on three principal European credit indices, of the most traded 5-year maturity<sup>1</sup>: the *Main Index*, noted here ITXEB, a combination of 125 equally weighted investment grade entities; the *Crossover Index*, noted ITXEX, with 75 sub-investment grade names; and the *Senior Financial*, noted ITXES, a subset of 30 financial entities from Main Index, referencing senior debt.

---

<sup>1</sup> Data provided by otcstreaming.com, and available on demand.

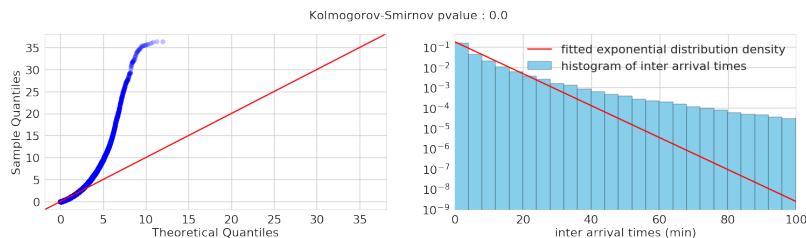


**Fig. 1.** Publicly reported trading activity on 22/05/2018 (*times, volumes*) for European credit indices ITXEB, ITXES & ITXEX.

When trying to capture the underlying dynamics of this market, one of the first questions that comes to mind is: how random is the timing of trades? A very naive approach is to model trades arrival times as purely random and uncorrelated processes, for example a Poisson process by trying to fit an exponential law over the distribution of inter-arrival times. Yet, such a model fails to fit the data: from Fig 2, we can notice a high density of very short inter-arrival times, suggesting that some self-excitation phenomenon occurs, with trades triggering more trades; an intuition that every practitioner of the field would confirm.

	$E(\Delta t)$	$\sigma(\Delta t)$	$Q_1(\Delta t)$	$Q_2(\Delta t)$	$Q_3(\Delta t)$	Number of trades
ITXEB	7 min	10 min	1 min	3 min	8 min	22 227
ITXEX	8 min	12 min	2 min	4 min	10 min	18 152
ITXES	17 min	24 min	3 min	9 min	21 min	7 194

**Table 1.** Statistics on inter-arrival times  $\Delta t$  (period: 03/01/2017 to 14/12/2017)



**Fig. 2.** ITXEB, from 15/01/2017 to 15/05/2018: *right*: log-scaled distribution of inter-arrival times; *left*: Q-Q plot. The red lines represent the best exponential law fitting. With a p-value of 0.0, this model is clearly inappropriate.

Literature review reveals that one model has recently driven considerable attention and a growing interest from both the scientific and quantitative communities: Hawkes processes. This model, introduced in 1971 by Hawkes ([6], [7])

to characterize earthquake tremors, can be used to describe timing of trades and their cross influences from a point process perspective.

After summarizing related work for Hawkes processes, we will recall the theoretical framework around this stochastic model. We will then focus on optimization methods for fitting such a process and describe Two Stage Hawkes Likelihood Optimization (2SHLO), a maximum likelihood based algorithm to fit multidimensional Hawkes processes with a parametric exponential kernel. Lastly, we will present our results, firstly on simulated data, then on credit derivative indices mentioned above. Hawkes processes in finance have mainly been applied to high frequency data. Is such a model appropriate for "mid-frequency" credit data? Can we specify the impact of volumes in trades or describe mutual influences between different indices?

## 1.2 Related work

Hawkes processes were originally designed for seismology analysis and deeply studied for that purpose by Ogata [12], [10], [11]. The concept has been utilized to model other effects where self and cross ignitions happen, such as social media tweets cascading [14], crime occurrences due to gang retaliations [9], or financial market events. As regards to financial applications, Bowsher [4] proposed in 2002 a generalized Hawkes process model taking into account night gaps, inter-day dependencies as well as intraday seasonality (with a piece-wise baseline intensity), which was used to model interactions between trades and price changes and also estimate of the price volatility based on mid-quote intensity for NYSE stock. Market price microstructure & market impact (influence of market orders on forthcoming prices) was modelled in [3], price impact was also studied in [1]. [13] focused on the impact of volume and order types on the limit order book. [2] (2015) exposes a full review of Hawkes process applications to finance, such as price & volatility modeling, market reflexivity measurement, order book modeling or risk contagion modeling.

## 2 Modeling Self and Cross Excitement with Hawkes Processes

This section re-frames the required formalism, as very clearly stated in [2], [8] or [15], starting from point and counting process definitions, through the key concept of intensity function. Hawkes model formulation is detailed for the reader, with a peculiar focus on exponential kernel structure.

### 2.1 Core concepts: from counting & point processes to intensity functions

**Definition 1 (Point Process).** Let  $(\Omega, \mathcal{F}, P)$  be a probability space.

Let  $(t_k)_{k \in \mathbb{N}^*}$  be a sequence of non-negative random variables such that  $\forall k \in \mathbb{N}^*$ ,  $t_k < t_{k+1}$ .  $(t_k)_{k \in \mathbb{N}^*}$  is called a (simple) point process on  $\mathbb{R}_+$ .

**Definition 2 (Counting Process).** Let  $(t_k)_{k \in \mathbb{N}}$  be a point process. The right-continuous stochastic process defined for all  $t \in \mathbb{R}_+$  as  $N(t) = \sum_{k \in \mathbb{N}^*} 1_{t_k \leq t}$  is called the counting process associated with  $(t_k)_{k \in \mathbb{N}^*}$ .

The study of point processes goes through a single mathematical object, the *intensity function*, which is defined as the conditional probability density of occurrence of an event in the immediate future.

**Definition 3 (Intensity Function).** Let  $N$  be a counting process adapted to a filtration  $\mathcal{F}_t$ . The left-continuous intensity is heuristically defined as

$$\lambda(t|\mathcal{F}_t) = \lim_{h \rightarrow 0} \mathbb{E} \left[ \frac{N(t+h) - N(t)}{h} \middle| \mathcal{F}_t \right] \quad (1)$$

The intensity function depends on the choice of filtration  $\mathcal{F}_t$ , which represents the amount of information available until time  $t$  (See [5] for a rigorous definition of the intensity function). In the context of Hawkes processes, we will simply use the natural filtration, with all previous information being available, and consider  $\lambda(t)$ .

#### Homogeneous Poisson process

One of the simplest point processes is the homogeneous Poisson process, for which the intensity function is constant over time:  $\forall t \geq 0$ ,  $\lambda(t) = \lambda$ . In that case, durations (or inter-event waiting times) are independent and identically distributed (following an exponential distribution of hazard rate  $\lambda$ ). As presented in the introduction, credit trades cannot be modeled by this memory-less model. Let's introduce Hawkes processes, a peculiar kind of non-homogeneous Poisson processes with linear dependencies over functions of past events.

## 2.2 Self-excitement: one-dimensional Hawkes processes

**Definition 4 (One-dimensional Hawkes Processes).**

Let  $(t_k)_{k \in \mathbb{N}}$  be a point process and  $N$  the associated counting process, such that its intensity function  $\lambda$  is defined for each time  $t \geq 0$  as

$$\lambda(t) = \mu(t) + \int_{-\infty}^t \phi(t-\tau) dN(\tau) = \mu(t) + \sum_{t_i < t} \phi(t-t_i) \quad (2)$$

where  $\mu : \mathbb{R} \mapsto \mathbb{R}_+$  is an exogenous base intensity and  $\phi : \mathbb{R}_+ \mapsto \mathbb{R}^+$  is a non-negative, measurable function such that  $\|\phi\|_1 = \int_0^\infty \phi(s) ds < 1$ .

$N$  is called a Hawkes process with baseline  $\mu$  and kernel  $\phi$ .

The kernel  $\phi$  expresses the positive influence of past events on the current value of the intensity. Each jump  $dN(\tau) \neq 0$  increases the probability of future events through the kernel  $\phi$ . Clustering effects and branching structure are well depicted by Hawkes processes, with the baseline activity generating immigrant events and descendant events enhancing the intensity.

For this study, we focus on *exponential kernels*, defined as  $\phi(t) = \alpha\beta e^{-\beta t} \mathbf{1}_{t \geq 0}$ , with parameters  $\alpha, \beta \geq 0$ , which allows easy interpretation:  $\alpha$  or *adjacency* represents the weight of previous events while  $\beta$  is the *decay* / typical duration of influence for a past event.

The *branching ratio*  $n = ||\phi||_1 = \int_0^\infty \phi(s)ds$  represents the average number of descendants for any event. For exponential kernel,  $n = \alpha$ .

### 2.3 Including mutual-excitement with multidimensional Hawkes processes

**Definition 5 (Multidimensional Hawkes Processes).**

Let  $M \in \mathbb{N}^*$  and  $\{(t_k^i)\}_{i=1,\dots,M}$  be a  $M$ -dimensional point process.

We denote by  $N_t = (N_t^1, \dots, N_t^M)$  the associated counting process such that its vector intensity function  $\lambda : \mathbb{R}_+ \mapsto \mathbb{R}_+^M$ , is defined as, for all  $t \geq 0$ ,  $i \in 1, \dots, M$ :

$$\lambda_i(t) = \mu_i + \sum_{j=1}^M \int_0^t \phi_{i,j}(t-\tau) dN_j(\tau) = \mu_i + \sum_{j=1}^M \sum_{n=1}^{N_j(t)} \phi_{i,j}(t-t_{j,n}), \quad (3)$$

with  $\mu = (\mu_i)_{i=1,\dots,M}$  an exogenous base intensity vector

and  $\phi(t) = (\phi_{i,j})_{i,j=1,\dots,M}$  a matrix-valued kernel that is component-wise positive and causal (null values when  $t < 0$ ) and with each component belonging to the space of  $L^1$ -integrable functions.

$N_t$  is called a multidimensional (or multivariate) Hawkes process with baseline  $\mu$  and kernel  $\phi$ .

The choice of an exponential kernel can be generalized for multivariate Hawkes processes with kernel components being expressed as  $\phi_{i,j}(t) = \alpha_{i,j}\beta_{i,j}e^{-\beta_{i,j}t} \mathbf{1}_{t \geq 0}$ .

This article focuses on Hawkes processes with exponential kernel and constant baselines, to which we propose a maximum likelihood based fitting method.

## 3 Hawkes Processes: Model Calibration

The most commonly used technique for parametric inference of Hawkes processes is a direct numerical optimization of the Maximum Likelihood, which was first introduced in the work of Ogata (1978) [10]. He proved the asymptotic consistency and efficiency of the Maximum Likelihood Estimator (MLE) under the assumption of stationary condition of the underlying point process. The negative log-likelihood function of a multidimensional Hawkes process over the time interval  $[0, t]$  is given in Daley & Vere-Jones [5] Proposition 13.1.VI by :

$$\mathcal{L}_t(\boldsymbol{\lambda}) := - \sum_{i=1}^M \left( \int_0^t \log \lambda_i(\tau) dN_i(\tau) - \int_0^t \lambda_i(\tau) d\tau \right). \quad (4)$$

Depending on the shape of kernels,  $\mathcal{L}_t$  is generally non-convex. Classic non-convex numerical optimization pitfalls are to be feared: a direct numerical optimization could converge to a merely local minimum. On the other hand, it may take too many iterations to converge as the negative log-likelihood function can be flat on some regions of the space of parameters. This problem is also faced when using an Expectation Maximization (EM) algorithm to find the MLE.

On the computational side, the repeated evaluation of the negative log-likelihood can be highly time consuming, essentially due to the nested sum in the first part of 4 where the conditional intensity is also expressed as a sum over the history.

### 3.1 Maximum Likelihood Estimation (MLE) for exponential multidimensional Hawkes processes

Let's consider a  $M$ -dimensional multivariate Hawkes model with constant base-lines  $\mu_i$  and kernel functions  $\phi_{i,j}$  as defined in 2.3. The choice of an exponential kernel has proved to be very interesting as it allows intuitive and meaningful interpretations of its parameters and also reduces the computational cost of the evaluation of the negative log-likelihood function as noted by Ogata (1981) [11] who exhibited a recursive formula that eliminated the nested sum evaluation problem.

The existing methods to fit exponential parameters through MLE directly use non-linear optimization algorithms such as Nelder-Mead (also called down-hill simplex method) or BFGS, with performance decreasing as the number of parameters increases. This represents quite an issue as  $M(2M + 1)$  parameters describe a  $M$ -dimensional Hawkes process with exponential kernel.

In addition, other existing methods often make the strong assumption that the decays  $\beta_{i,j}$  of the exponential kernel are given and fixed a priori. Consequently, the estimation of each kernel function is equivalent to the estimation of its adjacency coefficient  $\alpha_{i,j}$  and since the conditional intensity  $\lambda_i(t)$  is linear with respect to kernel functions  $\phi_{i,1}(t), \dots, \phi_{i,p}(t)$ , the negative log-likelihood function  $\mathcal{L}_t$  is convex with respect to all parameters. Therefore we can use the widely available convex optimization machinery to find the global minimum efficiently. The main drawback of this method is the possible inaccuracy of the decays parameters  $\beta_{i,j}$  which can lead to model mismatch.

### 3.2 Introducing Two Stage Hawkes Likelihood Optimization (2SHLO):

In order to combine the benefits of both approaches described above, we propose the following estimation method to fit exponential Hawkes processes.

The negative log-likelihood function for exponential multivariate Hawkes processes is given by (see Ogata[11] for the recursive formulation):

$$\begin{aligned} \mathcal{L}_t(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = & \sum_{m=1}^M \left( \mu_m T + \sum_{n=1}^M \alpha_{mn} \sum_{\{k: t_k^n < T\}} [1 - e^{-\beta_{mn}(T-t_k^n)}] \right. \\ & \left. - \sum_{\{k: t_k^n < T\}} \log[\mu_m + \sum_{n=1}^M \alpha_{mn} \beta_{mn} \sum_{\{k: t_k^n < t_i^m\}} e^{-\beta_{mn}(t_i^m - t_k^n)}] \right) \quad (5) \end{aligned}$$

As we mentioned earlier, if  $\beta$  is fixed, then  $\mathcal{L}_t(\cdot, \cdot, \beta)$  is convex with respect to parameters  $(\boldsymbol{\mu}, \boldsymbol{\alpha})$ . As a result it can be minimized using any convex optimization algorithm, for example an Accelerated Gradient Descent (AGD) which converges to a global minimum  $(\boldsymbol{\mu}^*(\boldsymbol{\beta}), \boldsymbol{\alpha}^*(\boldsymbol{\beta}))$  such that:

$$\mathcal{L}_t^*(\boldsymbol{\beta}) = \mathcal{L}_t(\boldsymbol{\mu}^*(\boldsymbol{\beta}), \boldsymbol{\alpha}^*(\boldsymbol{\beta}), \boldsymbol{\beta}) = \min_{(\boldsymbol{\mu}, \boldsymbol{\alpha})} \mathcal{L}_t(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \quad (6)$$

$\mathcal{L}_t^*(\boldsymbol{\beta})$  is still a non-convex function but defined on a space with a lower dimension than the initial parameter space. Any non-linear heuristic algorithms can now be used to minimize  $\mathcal{L}_t^*(\boldsymbol{\beta})$  leading to the MLE as<sup>2</sup>:

$$\min_{(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta})} \mathcal{L}_t(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \min_{(\boldsymbol{\beta})} \min_{(\boldsymbol{\mu}, \boldsymbol{\alpha})} \mathcal{L}_t(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \quad (7)$$

We summarize the fitting method as follow :

---

**Algorithm 1** Two Stage Hawkes Likelihood Optimization (2SHLO)

---

```

Start from mean of inter arrival times  $\beta_0$ 
for each step of Nelder-Mead method until convergence do
    for each needed computation of  $\mathcal{L}_t^*(\beta_i)$  do
        Start from a random  $(\boldsymbol{\mu}_0, \boldsymbol{\alpha}_0)$ 
        Use Accelerated Gradient Descent to optimize  $\mathcal{L}_t(\boldsymbol{\mu}, \boldsymbol{\alpha}, \beta_i)$ 
        Retrieve resulting  $\mathcal{L}_t^*(\beta_i)$ 
    end for
end for
return last  $(\boldsymbol{\beta}, \boldsymbol{\mu}^*(\boldsymbol{\beta}), \boldsymbol{\alpha}^*(\boldsymbol{\beta}))$ 

```

---

### 3.3 Goodness of fit

The *compensator function*  $\Lambda$  of a point process with intensity  $\lambda$  is defined as  $\forall t \geq 0, \Lambda(t) = \int_0^t \lambda(s | \mathcal{F}_t) ds$ . To assess the goodness of fit of our estimation, we use the residual point process analysis theorem, as stated in [5]:

---

<sup>2</sup> if we call  $m = \min_{(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta})} \mathcal{L}_t(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ , the following inequality holds for all  $(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ :  $m \leq \mathcal{L}_t(\boldsymbol{\mu}^*(\boldsymbol{\beta}), \boldsymbol{\alpha}^*(\boldsymbol{\beta}), \boldsymbol{\beta}) \leq \mathcal{L}_t(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta})$  3.2 follows immediately by taking the minimum over  $(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta})$  in each side of the inequality.

*The transformed sequence  $\{t_k^*\} = \{\Lambda(t_k)\}$  is a realization of a unit rate Poisson process if and only if the original sequence  $\{t_k\}$  is a realization from the point process defined by  $\lambda$ .*

The goodness of fit can be tested by comparing the transformed estimated times to a standard Poisson process using QQ plots and comparison of density functions.

## 4 Results - 2SHLO in practice

We have been running our experiments in Python 3, relying on *tick*[16] library for Hawkes process simulations, AGD optimization and analytic tools. 2SHLO is also benefiting of well-known *scipy* scientific package.

To provide an idea of the computational performance of the fitting method, we mention that, on average, it takes 4.69 seconds to fit 2 dimensional exponential hawkes process with training length of  $T = 10000$ .

### 4.1 Validating 2SHLO performances over simulated data

To generate Hawkes processes, a few methods are available (as summarized in [2]), either based on thinning, time-change or cluster algorithm. We have here used *tick* functions for simulations which are based on the thinning algorithm described in (Ogata, 1981, p.25, Algorithm 2) [11].

To validate the performance of the fitting algorithm, we ran a series of 100 simulation and fitting procedure on the simulated 2D Hawkes time-stamps using different end times  $T$  to have different training sizes and using the following simulation parameters :

$$\mu = \begin{pmatrix} 0.1 \\ 0.2 \end{pmatrix} \quad \alpha = \begin{pmatrix} 0.5 & 0.00 \\ 0.4 & 0.3 \end{pmatrix} \quad \beta = \begin{pmatrix} 0.3 & 0.00 \\ 0.2 & 0.2 \end{pmatrix}$$

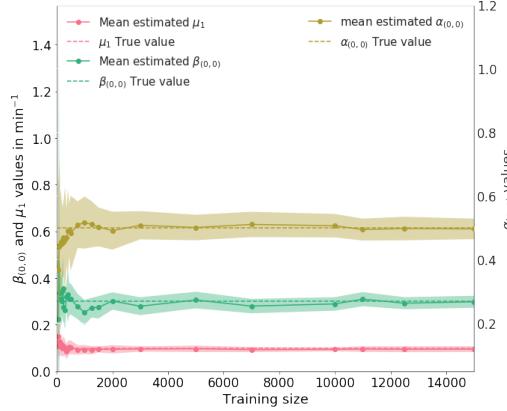
Fig. 3 confirms that the estimated parameters converge to their true optimal values for increasing  $N$  number of ticks in the training set. This reassures that when the Hawkes process is well specified and indeed recovers the optimal parameters asymptotically.

### 4.2 Calibrating univariate Hawkes processes on "mid-frequency" credit derivative trades

We have fitted a modified one-dimensional Hawkes process on the period from 15/01/2017 to 15/12/2017 of reported trades data<sup>3</sup> on 3 indices ITXEB, ITXES and ITXEX. The modification consists of imposing a null value of the intensity in the gap between two days and the standard exponential kernel format during trading days so as to account for the absence of trading activity overnight. The model is thus slightly modified and described in the appendix section in a more

---

<sup>3</sup> The trading data have been cleaned by discarding transactions that do not reflect market signals such as roll and switch trades



**Fig. 3.** Estimated parameters of a 2D exponential Hawkes process trained with 2SHLO on simulated datasets with increasing lengths

detailed way. Before making this modeling choice, we experimented with a more elaborate model that accounts for the overnight spillover effect as described in the work of C. G.Bowsher [4] where the author defines the intensity as recursively depending on the level of the stochastic parts of the intensity at the end of the previous trading day and thus extending the effect of trading days over time. The fitting procedure resulted into a null parameter estimation of the spillover effect which motivated the choice of the model modification described above.

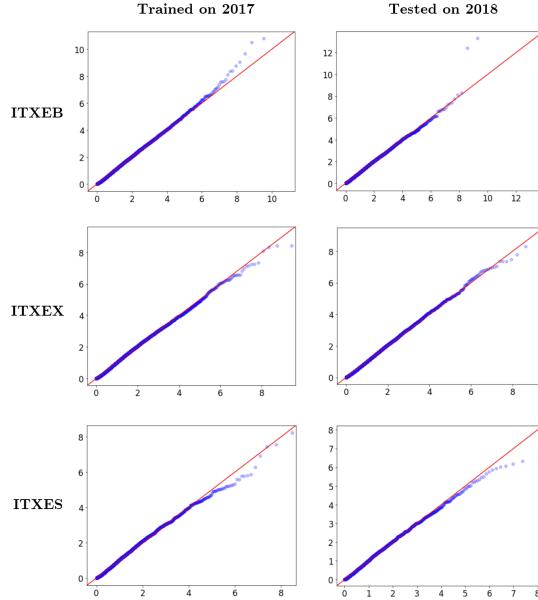
To validate the 'universality' of our estimates, we tested the goodness of fit of the Hawkes model with the estimated parameters on out of sample period from 15/01/2018-15/07/2018. The model fitted well for the 3 indices time series (see Fig. 4) suggesting that each series has stable intrinsic parameters that characterize the pattern of trades arrival times.

	Estimated $\mu^{-1}$	Estimated $\alpha$	Estimated $\beta^{-1}$
ITXEB	22 min	0.62	20 min
ITXEX	24 min	0.60	23 min
ITXES	58 min	0.65	36 min

**Table 2.** Estimated parameters for ITXEB, ITXES and ITXEX over the period (2017-01-15 to 2017-12-15)

### 4.3 Measuring the influence of traded volumes

When considering the distribution of trade volumes for credit indices, volume clusters are clearly outlined. For example, for index ITXEB, three bins can be distinguished: "small" trades, "medium" trades and "big" trades (See Fig. 5).



**Fig. 4.** Hawkes process fitted on 2017 data and tested on 2018 data for each index

Using multivariate Hawkes processes model, volume impacts have been modeled by considering 3 counting processes, splitting trades per volume size.

By using the same model as the last section to take overnight gaps into consideration, the fitting algorithm trained on 2018 data resulted in the following parameters estimates ( $\mu$  and  $\beta$  expressed in minute $^{-1}$ ) :

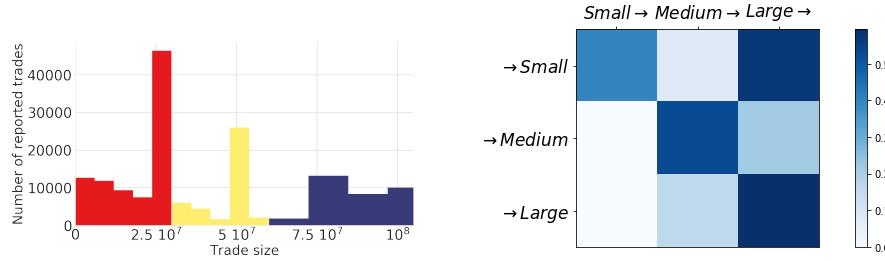
$$\frac{1}{\mu} = \begin{pmatrix} 40 \\ 42 \\ 38 \end{pmatrix} \quad \alpha = \begin{pmatrix} 0.40 & 0.07 & 0.57 \\ 0.00 & 0.53 & 0.21 \\ 0.00 & 0.16 & 0.59 \end{pmatrix} \quad \frac{1}{\beta} = \begin{pmatrix} 12 & 21 & 53 \\ 71 & 18 & 12 \\ 53 & 14 & 34 \end{pmatrix}$$

Visualizing in Fig. 5 the branching ratio which corresponds to the values of  $\alpha$  allows some intuitive interpretation of the fitted model as it shows that large trades are purely self exciting process and they have a major influence in triggering small trades, we also notice that all trade categories have a non negligible self excitation component.

#### 4.4 Measuring cross interactions between traded indices

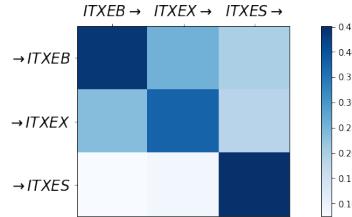
We have adopted again a multivariate Hawkes processes with null intensity overnight model to study the cross interaction between indices trades times. The fitting algorithm trained on 2017 data resulted in the following parameters estimates ( $\mu$  and  $\beta$  expressed in minute $^{-1}$ ) :

$$\frac{1}{\mu} = \begin{pmatrix} 29 \\ 30 \\ 125 \end{pmatrix} \quad \alpha = \begin{pmatrix} 0.44 & 0.25 & 0.20 \\ 0.23 & 0.37 & 0.18 \\ 0.07 & 0.08 & 0.45 \end{pmatrix} \quad \frac{1}{\beta} = \begin{pmatrix} 23 & 9 & 21 \\ 14 & 13 & 31 \\ 39 & 17 & 25 \end{pmatrix}$$



**Fig. 5.** Distribution of trade sizes over one week trading period 2018-05-21:2018-05-25 (left; adjacency matrix ( $\alpha_{i,j}$ ) of 3D exponential Hawkes fitted on 3 volume bin series (for index ITXEB).

From Fig. 6 we can notice that ITXES index is the least influenced by the other indices with the very low cross excitation components and a high self excitation component which was expected due to its low liquidity. We also notice that the estimated baselines of all indices are greater for this multivariate model compared to the uni-dimensional model which can be explained by the fact that we decreased the 'Poissonian' behaviour of the arrival times by including more information about the influence of other indices.



**Fig. 6.** adjacency matrix ( $\alpha_{i,j}$ ) of 3D exponential Hawkes fitted on 3 indices series (ITXEB, ITXES, ITXEX) for the period 2017-01-15 - 2017-12-15.

## 5 Conclusion

Through point process analysis, Hawkes processes allows a relatively simple and interpretable representation of time dependant events with self and mutual excitements. With 2SHLO; a maximum likelihood based algorithm mixing convex and non-convex optimization, we have been able to fit fastly and properly exponential kernel multivariate Hawkes processes. The algorithm has been applied to credit derivative trading data, an unexplored "mid-frequency" market for such processes. We have been able to emphasize self excitement for three index

trades and to measure impacts of traded volumes. With the same framework, we were also able to quantify cross influences between indices. Next prospects turns towards testing 2SHLO performances on high dimensional datasets, for instance studying both price & volume trade influences as well as market bid/ask dynamics on credit derivatives in a forecasting perspective.

## References

1. Amaral, L., Papanicolaou, A.: Price Impact of Large Orders Using Hawkes Processes. In: NYU Tandon Research Paper No. 2874042 (2017)
2. Bacry, E., Mastromatteo, I., Muzy, J.F.: Hawkes Processes in Finance. In: Market Microstructure and Liquidity, vol. 01 (2015)
3. Bacry, E., Muzy, J.F.: Hawkes models for price and trades high-frequency dynamics. In: Quantitative Finance, vol. 14, pp. 1147–1166 (2013)
4. Bowsher, C. G.: Modelling Security Market Events in Continuous Time: Intensity Based, Multivariate Point Process Models. In: Nuffield Economics Working Paper (2003)
5. Daley, D.J., Vere-Jones, D.: An introduction to the theory of point processes, vol. 2. Springer (1988)
6. Hawkes, A.G.: Spectra of Some Self-Exciting and Mutually Exciting Point Processes. In: Biometrika, vol. 58, pp. 83–90 (1971)
7. Hawkes, A.G.: Point Spectra of Some Mutually Exciting Point Processes. In: Journal of the Royal Statistical Society. Series B (Methodological), vol. 33, pp. 438–443 (1971)
8. Laub, P., Taimre, T., Pollett, P.: Hawkes Processes. In: arXiv preprint arXiv:1507.02822 (2015)
9. Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., Tita, G. E.: Self-Exciting Point Process Modeling of Crime. In: Journal of the American Statistical Association, vol. 106, pp. 100–108 (2011)
10. Ogata, Y., The asymptotic behaviour of maximum likelihood estimators for stationary point processes. Annals of the Institute of Statistical Mathematics, vol. 30, pp. 243–261 (1978)
11. Ogata, Y., On Lewis simulation method for point processes, IEEE Transactions on Information Theory, vol. 27, pp. 23–31 (1981)
12. Ogata, Y.: Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes. In: Journal of the American Statistical Association, vol. 83, pp. 9–27 (1988)
13. Rambaldi, M., Bacry, E., Lillo, F.: The role of volume in order book dynamics: a multivariate Hawkes process analysis. In: Quantitative Finance (2016)
14. Rizoiu, M., Lee, Y., Mishra, S., Xie, L.: A Tutorial on Hawkes Processes for Events in Social Media. In: Frontiers of Multimedia Research, pp. 191–218 (2017)
15. Toke, I. M., An Introduction to Hawkes Processes with Applications to Finance (2011) [http://lamp.ecp.fr/MAS/fiQuant/ioane\\_files/HawkesCourseSlides.pdf](http://lamp.ecp.fr/MAS/fiQuant/ioane_files/HawkesCourseSlides.pdf)
16. Bacry E., Bompaire M., Gaffas S., Poulsen S., Tick: a Python library for statistical learning, with a particular emphasis on time-dependent modelling. <https://x-datainitiative.github.io/tick/#>

## Appendix: Hawkes Processes Formulas

### 5.1 The Univariate Bowsher Hawkes process model

**Definition:** As described in [4], the intensity of the univariate Bowsher process is defined recursively depending on the level of the stochastic parts of the intensity function at the end of the  $(d - 1)$ th trading day and the contributions of the events occurring on day  $d$ . Thus it is necessary to perform the following data transformation: the real half-line is partitioned into intervals of length  $l$  corresponding to the different trading days. This partition is written as :

$$(0, \infty) = (0, \tau_1] \cup (\tau_1, \tau_2] \cup \dots \cup (\tau_{d-1}, \tau_d] \dots$$

The model is thus defined by the (scalar) stochastic intensity :

$$\forall t \geq 0, \quad \lambda(t) = \mu + \tilde{\lambda}(t)$$

such that :

$$\forall t \in ]\tau_{d-1}, \tau_d] \quad \tilde{\lambda}(t) = \pi \tilde{\lambda}(\tau_{d-1}) e^{-\rho(t-\tau_{d-1})} + \int_{[\tau_{d-1}, t)} \alpha e^{-\beta(t-u)} dN(u)$$

where the parameters used are :

- $]\tau_{d-1}, \tau_d]$  is the interval defining the day number  $d$
- $\mu$  is the baseline intensity
- $\pi$  is the spillover adjacency coefficient
- $\rho$  is the spillover decay
- $\alpha$  is the self excitement adjacency coefficient
- $\beta$  is the self excitement decay
- $dN$  is the process differentiate

**The Loglikelihood:** The minus loglikelihood of a univariate point process is defined as :

$$\mathcal{L}_T(\lambda) = - \left( \int_0^T \log \lambda(t) dN_i(t) - \int_0^T \lambda(t) dt \right)$$

Computing the minus loglikelihood in the Hawkes Bowsher model results into the following closed form :

$$\begin{aligned} \mathcal{L}_T(\lambda) &= \mu T + \sum_{d \in days} \pi \tilde{\lambda}(\tau_{d-1}) (1 - e^{-\rho(\tau_d - \tau_{d-1})}) \\ &\quad + \alpha \sum_{t \in ]\tau_{d-1}, \tau_d]} (1 - e^{-\rho(t - \tau_{d-1})}) - \sum_{t_i} \log \lambda(t_i) \end{aligned}$$

**The Fitting procedure:** As the task of computing the gradient of each parameter of the loglikelihood is a fastidious one because of the recursive definition of the intensity, and also because we have only 5 parameters to fit  $(\mu, \pi, \rho, \alpha, \beta)$ , we choose to use the L-BFGS-B algorithm to perform the minimization of  $\mathcal{L}_T(\lambda)$  which had delivered satisfying performance on fitting simulated data.

**Data simulation:** We use the thinning algorithm to simulate artificial data described by the following algorithm :

---

**Algorithm 2** Simulation by thinning

---

1. Given Bowsher Hawkes process described as above
  2. Set current time  $T = 0$  and event counter  $i = 1$
  3. While  $i \leq N$ 
    - (a) Set the upper bound of Poisson intensity  $\lambda^* = \lambda(T)$ .
    - (b) Sample inter-arrival time: draw  $u \sim U(0, 1)$  and let  $\tau = -\frac{\ln(u)}{\lambda^*}$  (as described in).
    - (c) Update current time:  $T = T + \tau$ .
    - (d) Draw  $s \sim U(0, 1)$ .
    - (e) If  $s \leq \frac{\lambda(T)}{\lambda^*}$ , accept the current sample: let  $T_i = T$  and  $i = i + 1$ . Otherwise reject the sample, return to step (a).
- 

## 5.2 The Day Gaps Multivariate Exponential Hawkes process model

**Definition:** The intensity of the Day Gaps Multivariate Exponential Hawkes process is defined just like the standard multivariate exponential Hawkes process model described in [6] given by :

$$\forall i \in [1 \dots D], \quad \lambda_i(t) = \mu_i + \sum_{j=1}^D \int \phi_{ij}(t-s) dN_j(s)$$

Where :

- $D$  is the number of nodes
- $\mu_i$  are the baseline intensities
- $\phi_{ij}$  are the kernels
- $dN_j$  are the processes differentiates

with an exponential parametrisation of the kernels:

$$\phi_{ij}(t) = \alpha_{ij} \beta_{ij} \exp(-\beta_{ij} t) 1_{t>0}$$

- $\alpha_{ij}$  are the adjacency of the kernel

- $\beta_{ij}$  are the decays of the kernel

With the modification consisting on considering the intensity as a null function between two consecutive days to take the day gaps into consideration :

$$\forall t \text{ between two consecutive days, } \lambda_i(t) = 0$$

**The Loglikelihood:** The minus loglikelihood of a Multivariate Point Process is defined as :

$$\mathcal{L}_t(\boldsymbol{\lambda}) := - \sum_{i=1}^M \left( \int_0^t \log \lambda_i(\tau) dN_i(\tau) - \int_0^t \lambda_i(\tau) d\tau \right).$$

Computing the minus loglikelihood in The Day Gaps Multivariate Exponential Hawkes process model results into the following closed form :

$$\begin{aligned} \mathcal{L}_t(\boldsymbol{\lambda}) := & D * \delta \sum_{m=1}^M \mu_m + \sum_{m=1}^M \sum_{n=1}^M \alpha_{mn} \sum_{d \in \text{days}} \sum_{t_i^n \in \text{day d}} (1 - e^{-\beta_{mn}(\tau_d - t_i^n)}) \\ & - \sum_{m=1}^M \sum_{t_i^m} \log(\mu_m + \sum_{n=1}^M \alpha_{mn} \beta_{mn} R_{mn}(i)) \end{aligned}$$

with  $R_{mn}(i)$  are defined using the Ogoata [11] recursive formula :

$$\begin{aligned} R_{mn}(i) &= \sum_{\{k: t_k^n < t_i^m\}} e^{-\beta_{mn}(t_i^m - t_k^n)} \\ &= e^{-\beta_{mn}(t_i^m - t_{i-1}^m)} R_{mn}(i-1) + \sum_{\{k: t_{i-1}^m \leq t_k^n < t_i^m\}} e^{-\beta_{mn}(t_i^m - t_k^n)} \end{aligned}$$

Where :

- $D$  is the number of days in the data
- $\delta$  is the length of one day
- $\mu_i$  are the baseline intensities
- $\alpha_{ij}$  are the adjacency coefficients
- $\beta_{ij}$  are the decays coefficients
- $\tau_d$  is the time defining the last time of the day number d

**The Fitting procedure:** To fit this model, we use the algorithm described in this paper, using the projected Newton Descent for the convex optimization part and Nelder Mead for the non convex optimization step. The choice of the newton descent is justified by the easy computation of the gradient and hessian matrix which is sparse and because this algorithm has a quadratic convergence rate compared to gradient descent methods.

The gradient of  $\mathcal{L}_t(\boldsymbol{\lambda})$  is thus given by the following close formulas :

$$\frac{\partial \mathcal{L}_t(\boldsymbol{\lambda})}{\partial \mu_m} = D * \delta - \sum_{t_i^m} \frac{1}{\mu_m + \sum_{n=1}^M \alpha_{mn} \beta_{mn} R_{mn}(i)}$$

$$\frac{\partial \mathcal{L}_t(\boldsymbol{\lambda})}{\partial \alpha_{mn}} = \sum_{d \in days} \sum_{t_i^n \in day d} (1 - e^{-\beta_{mn}(\tau_d - t_i^n)}) - \sum_{t_i^m} \frac{\beta_{mn} R_{mn}(i)}{\mu_m + \sum_{l=1}^M \alpha_{ml} \beta_{ml} R_{ml}(i)}$$

The hessian of  $\mathcal{L}_t(\boldsymbol{\lambda})$  is given by the following close formulas :

$$\frac{\partial^2 \mathcal{L}_t(\boldsymbol{\lambda})}{\partial \mu_m \partial \mu_m} = \sum_{t_i^m} \frac{1}{(\mu_m + \sum_{n=1}^M \alpha_{mn} \beta_{mn} R_{mn}(i))^2}$$

$$\frac{\partial^2 \mathcal{L}_t(\boldsymbol{\lambda})}{\partial \mu_m \partial \alpha_{ml}} = \sum_{t_i^m} \frac{\beta_{ml} R_{ml}(i)}{(\mu_m + \sum_{n=1}^M \alpha_{mn} \beta_{mn} R_{mn}(i))^2}$$

$$\frac{\partial^2 \mathcal{L}_t(\boldsymbol{\lambda})}{\partial \alpha_{ml} \partial \mu_m} = \sum_{t_i^m} \frac{\beta_{ml} R_{ml}(i)}{(\mu_m + \sum_{n=1}^M \alpha_{mn} \beta_{mn} R_{mn}(i))^2}$$

$$\frac{\partial^2 \mathcal{L}_t(\boldsymbol{\lambda})}{\partial \alpha_{mk} \partial \alpha_{ml}} = \sum_{t_i^m} \frac{\beta_{mk} R_{mk}(i) \beta_{ml} R_{ml}(i)}{(\mu_m + \sum_{n=1}^M \alpha_{mn} \beta_{mn} R_{mn}(i))^2}$$

and all other coefficients are null. The fitting algorithm is described as follow:

---

**Algorithm 3** Two Stage Hawkes Likelihood Optimization using Projected Newton Descent

---

```

Start from mean of inter arrival times  $\boldsymbol{\beta}_0$ 
for each step of Nelder-Mead method until convergence do
    for each needed computation of  $\mathcal{L}_t^*(\boldsymbol{\beta}_i)$  do
        Start from a well chosen  $(\boldsymbol{\mu}_0, \boldsymbol{\alpha}_0)$ 
        Use Projected Newton Descent to optimize  $\mathcal{L}_t(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta}_i)$  :
        Retrieve resulting  $\mathcal{L}_t^*(\boldsymbol{\beta}_i)$ 
    end for
end for
return last  $(\boldsymbol{\beta}, \boldsymbol{\mu}^*(\boldsymbol{\beta}), \boldsymbol{\alpha}^*(\boldsymbol{\beta}))$ 

```

---

And the projected newton descent is described as follow :

**Algorithm 4** Projected Newton Descent

---

```

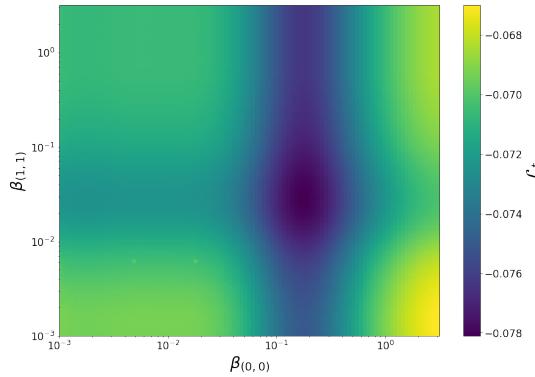
Start from a well chosen  $x_0$ 
for each step until convergence or iter max do
    Set prev_x = x, prev_obj = obj
    Set step = 1
    Set shrink = 0.5 (user choice of a shrinkx in [0, 1])
    while True do
        next_x = project(x - step *  $(\nabla^2 f(x))^{-1} \nabla f(x)$ )
        next_obj = f(next_x)
        if next_obj > prev_obj then
            step *= shrink
        else
            x = next_x
        end if
    end while
    test convergence  $\text{abs}(\text{next\_obj} - \text{prev\_obj}) / \text{abs}(\text{prev\_obj}) < \text{tolerance}$ 
end for
return x

```

---

**5.3 Some empirical results:**

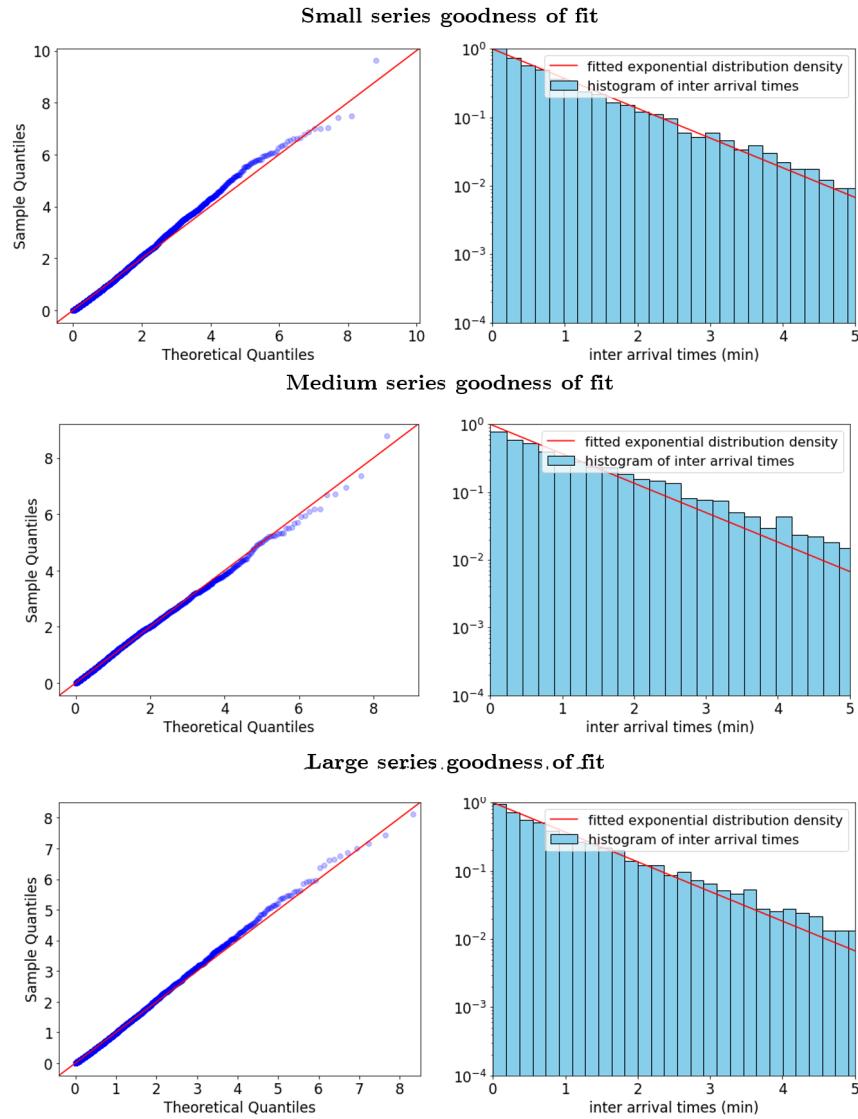
**Loglikelihood landscape:** Visualizing the minus loglikelihood function for 2D exponential Hawkes process with regards to two non convex variables  $(\beta_{0,0}, \beta_{1,1})$  resulted into the following figure, emphasizing the flatness of the function to minimize and the existence of valleys that pose problem to gradient based optimization :



**Fig. 7.**  $\mathcal{L}_t$  plot for a 2D exponential Hawkes process loglikelihood with respect to decays variables  $(\beta_{0,0}, \beta_{1,1})$ , other parameters being fixed:  $(\beta_{0,1}, \beta_{1,0}) = (0.1, 0.1)$ ,  $\mu = (0.02, 0.25)$ ,  $\alpha = ((0.3, 0.15), (0.01, 0.35))$

Bahamou, A., Doumercue, M., Donnat, P.

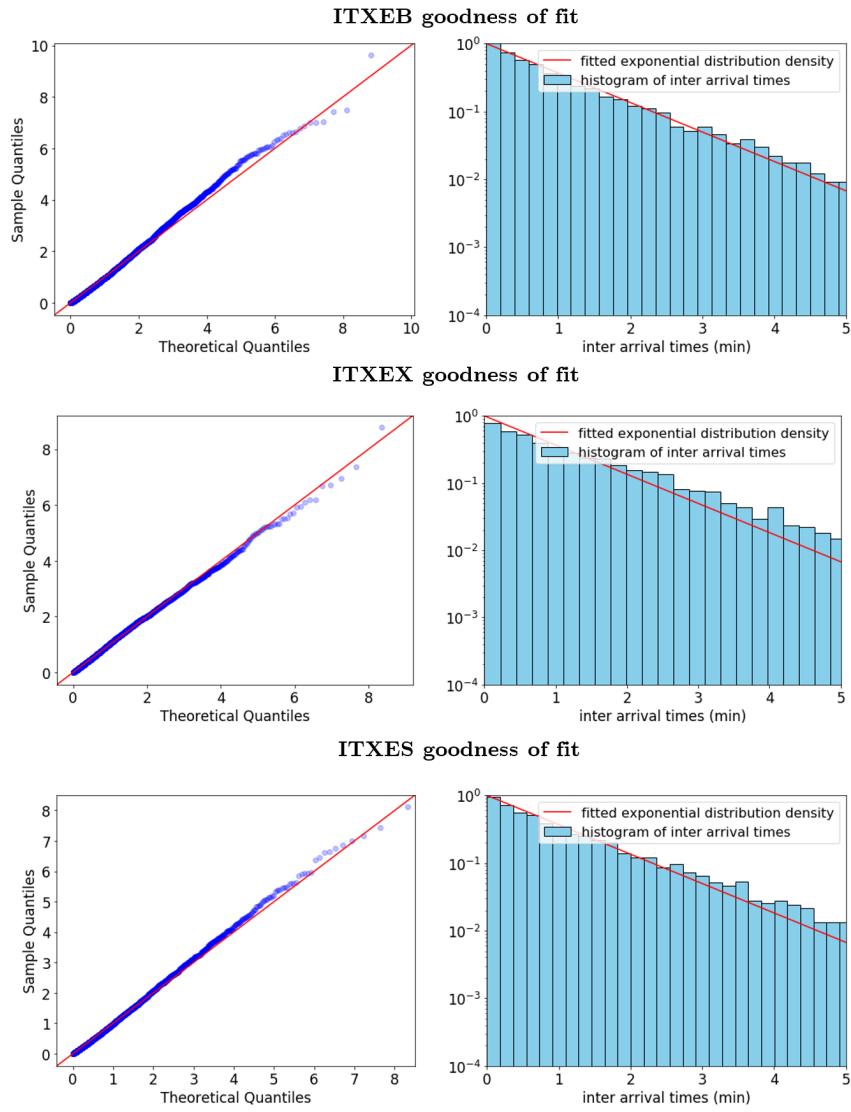
**Goodness of fit size bins study:** Assessing the goodness of fit of the multivariate Hawkes process fitted on 2017 ITXEB reported trades split into three series with regards to the size of trades:



**Fig. 8.** QQPlots of goodness of fit of the multivariate Hawkes process fitted on 2017 ITXEB by size bin reported trades

## Hawkes processes for credit indices time series analysis

**Goodness of fit Multi index study:** Assessing the goodness of fit of the multivariate Hawkes process fitted on 2017 ITXEB, ITXES and ITXEX reported trades



**Fig. 9.** QQPlots of goodness of fit of the multivariate Hawkes process fitted on 2017 ITXEB, ITXES and ITXEX