Contemporary Mathematics
https://ojs.wiserpub.com/index.php/CM/

Research Article

Hawkes Processes in Finance: An Application in Modeling Transactions in EUA Futures

Michele Bellomo, Francesco Grimaccia*

Department of Energy, Polytechnic University of Milan, Milan, Italy
E-mail: francesco.grimaccia@polimi.it

Received: 6 May 2025; **Revised:** 12 June 2025; **Accepted:** 13 June 2025

Abstract: Hawkes processes are stochastic point processes with self-exciting behavior, where the occurrence of an event increases the probability of subsequent events. In this paper, we provide a concise yet comprehensive introduction to Hawkes processes, focused on applications. We then employ marked Hawkes processes to model the occurrences of transactions in European Union Allowances (EUA) futures. The models are trained using an innovative and efficient Maximum Likelihood Estimation (MLE) implementation, that leverages PyTorch library to parallelize computations and exploit automatic differentiation. We evaluate and compare different excitation functions using Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Quantile-Quantile (Q-Q) plots based on the Random Time Change theorem. Additionally, we analyze the distribution of the calibrated parameters across different trading hours, finding the consistent relationship $\alpha = \sqrt[3]{\beta^2}$, to our knowledge never described before, between the self-exciting intensity parameter α and the decay parameter β. Furthermore, we identify two clusters: one characterized by transactions with a strong self-exciting effect, but that decays very quickly, and another with a less intense effect, that persists over longer time. Using logistic regression, we analyze the characteristics of the clusters and provide possible interpretations. We consider Hawkes processes very promising yet underutilized techniques for studying financial markets, and we believe this work can encourage researchers and practitioners to dive further into these models.

***Keywords*:** mathematical finance, market microstructure, point processes

MSC: 65L05, 91G80

Abbreviation

EUA	European Union Allowances
MLE	Maximum Likelihood Estimation
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
Q-Q	Quantile-Quantile

DOI: https://doi.org/10.37256/cm.6520257184

1. Introduction

Hawkes processes [1, 2] are stochastic point processes with self-exciting behavior, where the occurrence of an event increases the probability of subsequent events. These processes have been successfully applied in seismology [3–5] and, more recently, in epidemiology [6], social sciences [7], and finance [8–11].

In this paper, we provide a concise yet comprehensive introduction to Hawkes processes, focused on application. [10] provides an excellent overview of Hawkes processes, particularly in reference to financial markets, but the topic is presented in a way that might be difficult to understand for practitioners without a mathematical background in stochastic processes. Additionally, the implementation details related to the training and the diagnostic of Hawkes processes are not discussed. [12, 13] provides a more accessible and intuitive introduction. Our proposed treatment is similar, but we also discuss marked Hawkes processes, and focus on the techniques that will be used in this study.

After introducing Hawkes models, we show how these processes can be used to model transactions in the European Union Allowances (EUAs) futures. The EUA market is one of the main instruments of the European Union's climate policy to reduce greenhouse gas emissions. Each company within the European Union Emissions Trading System (EU ETS) receives or purchases emission allowances, known as EUAs, where each allowance grants the right to emit one ton of carbon dioxide equivalent (CO_2e). Companies that emit less than their allocated allowances can sell the excess allowances on the market. Conversely, companies that emit more than allowed must purchase additional allowances to cover their emissions. EUA allowances are traded either on the spot market for immediate delivery or via futures contracts for delivery at a future date. In our work, we focus on the one-year maturity EUA futures.

The paper is structured as follows. In Section 2, we provide a brief introduction to Hawkes processes and all the methodologies employed in this study. This section can be skipped by readers who are already familiar with the field, except for the Subsection 2.3.2, where we present an innovative and efficient training implementation, and Subsection 2.5, where we explain the modeling choices for the EUA futures. In Section 3, we present the results related to the models training. In particular, in Subsection 3.1, we present the descriptive results and statistics for the trading period under study; in Subsection 3.2 we perform a statistically rigorous comparison of the goodness of fit of the trained models; in Subsection 3.3, we analyze the distribution of the model parameters calibrated across the different trading hours, obtaining useful insights into market conditions; in Subsection 3.4, we enter in the details of three specific trading hours where the models show goodness of fit that is respectively better, worse, and in line with the average. Finally, in Section 4, we discuss the obtained results, while in Section 5, we outline possible future work.

2. Methods

2.1 *Preliminary definitions*

2.1.1 *Point and counting processes*

Hawkes processes are part of a broader class of stochastic processes called point processes. For a rigorous definition of point processes, please refer to [14]. Intuitively, we can think of a point process $(T_n)_{n\geq 1}$ as an increasing sequence of random times

$$0 \leq T_1 \leq T_2 \leq \dots \tag{1}$$

These random variables represent the occurrence (or arrival) times of events. If the sequence is strictly increasing, meaning that two or more events cannot occur at the same instant, the point process is said to be simple.

A point process can be associated with a counting process N_t, a stochastic process that counts the number of arrivals up to time t

$$N_t = \sum_{n \geq 1} 1_{\{T_n \leq t\}} \tag{2}$$

Since the counting process and the point process are two representations of the same model, their terminology is often interchangeable. For example, one may generically refer to a "Poisson process" or a "Hawkes process", and the reader must infer from the context whether the associated counting process or point process is being discussed.

An important quantity for a point process is the conditional intensity, which can be heuristically defined as follows:

Definition 1 (Conditional intensity) Let N_t be a counting process. The conditional intensity $\lambda(t \mid \mathscr{F}_t)$ is defined as:

$$\lambda(t \mid \mathscr{F}_t) = \lim_{\Delta t \to 0^+} \frac{\mathbb{E}(N_{t+\Delta t} - N_t \mid \mathscr{F}_t)}{\Delta t} \tag{3}$$

We recall that the filtration $\mathscr{F}_t$ can be interpreted as the information available up to time t. Therefore, the conditional intensity represents the instantaneous rate of occurrence of events at time t, given the history up to time t. In the following, we will use the simplified notation $\lambda(t) := \lambda(t \mid \mathscr{F}_t)$, that omits the explicit dependence on the filtration.

Finally, we introduce the compensator of a point process:

Definition 2 (Compensator) Given a counting process N_t, the non-decreasing function

$$\Lambda(t) = \int_0^t \lambda(s)\, ds \tag{4}$$

is called the compensator of the process.

2.1.2 *Poisson process*

We now present the simplest and most known point processes: the homogeneous Poisson process. This process represents a good benchmark for comparing more complex models, and, as we will see, it also provides an important diagnostic method for Hawkes process's goodness of fit. There are several equivalent definitions of a Poisson process; we provide the one that defines it based on its behavior over intervals.

Definition 3 (Poisson process) The counting process N_t is said to be a Poisson process of rate $\lambda > 0$ if:

- $N(0) = 0$;
- The process has independent increments;
- The number of events in any interval of length t is Poisson distributed with mean λt

$$\mathbb{P}(N_{T+t} - N_T = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}, \quad t > 0,\ k \in \mathbb{N} \tag{5}$$

The following proposition characterizes the distribution of the interarrival times of a Poisson process, that is, the times between the occurrence of one event and the next.

Theorem 1 (Interarrival times of a Poisson process) The interarrival times of a Poisson process of rate λ are independent identically distributed exponential random variables with parameter λ

$$f_{T_{i+1} - T_i}(x) = \lambda e^{-\lambda x} \mathbf{1}_{\{x \geq 0\}} \tag{6}$$

2.2 *Hawkes processes*

2.2.1 *Hawkes processes definition*

Definition 4 (Hawkes process) A point process is said to be a Hawkes process if the conditional intensity function $\lambda(t)$ takes the form:

$$\lambda(t) = \mu(t) + \sum_{T_i<t} \phi(t - T_i) \tag{7}$$

where $\mu(\cdot)$ is a nonnegative function called background or exogeneous intensity, and $\phi(\cdot)$ is a nonnegative function called kernel or excitation function.

From this definition, it is immediately evident that the intensity of a Hawkes process explicitly depends on the occurrence of previous events through the kernel function ϕ. The kernel function quantifies the self-exciting effect of past events and how this effect decays over time. It is common to choose a monotonically decreasing kernel function ϕ, so that the effect of an event occurrence strictly decreases as time passes. Furthermore, to ensure well-behaved models, the effect must decay sufficiently fast to prevent the explosion of the phenomenon, meaning that the number of events diverges to infinity in a finite time interval. In the next subsection, the two most common kernel functions are discussed.

2.2.2 *Kernel functions*

The most popular excitation function for a Hawkes process is the exponential kernel.

Definition 5 (Exponential kernel) The exponential kernel is an excitation function of the form

$$\phi(x) = \alpha e^{-\beta x}, \quad x \geq 0 \tag{8}$$

with $\alpha, \beta > 0$.

Some works use for the exponential kernel the following alternative parameterization:

$$\phi(x) = \alpha\beta e^{-\beta x}, \quad x \geq 0 \tag{9}$$

In both parameterizations, β represents the decay rate of the self-exciting effect related to an event. The interpretation of α depends on the choice of parameterization: in Definition 5, it represents the instantaneous effect of an event on the conditional intensity right after it occurs, whereas in the alternative parameterization it reflects the total cumulative effect over an infinite time. In the following, we will always use the parameterization introduced in Definition 5.

The exponential kernel has some convenient properties. The most important one is that a Hawkes process with exponential kernel and its corresponding conditional intensity $\lambda(t)$ form a Markov process. This fact can be leveraged to evaluate the conditional intensity and the compensator in an extremely efficient manner. This is particularly useful during training, where the computation of these functions is required many times.

Despite its convenience, empirical studies have shown that in many situations the exponential kernel is not the optimal choice for fitting the data. A popular alternative is the power-law kernel.

Definition 6 (Power-law kernel) The power-law kernel kernel is an excitation function of the form

$$\phi(x) = \frac{k}{(c+x)^p}, \quad x \geq 0 \tag{10}$$

with $k, c > 0$ and $p > 1$.

Some works report the alternative parameterization

$$\phi(x) = \frac{kc}{(1+cx)^{1+p}}, \quad x \geq 0 \tag{11}$$

with $k, c, p > 0$.

The effect of the power-law kernel decays less rapidly compared to the exponential, and it is therefore suitable for capturing longer-term effects. Empirically, the power-law kernel has shown to provide a better fit in the case of seismic data [15] and financial data [16, 17]. However, a Hawkes process with a power-law kernel does not have the Markov property, and therefore its training is generally slower.

2.2.3 *Marked Hawkes process*

A possible extension of Hawkes processes involves assigning to each event a mark, which can contribute to the excitation function. For instance, the mark could represent the intensity of an earthquake, or the volume associated with a financial transaction.

Definition 7 (Marked Hawkes process) A point process $(T_n)_{n\geq 1}$ and a sequence of $(\xi_n)_{n\geq 1}$ random variables form a marked Hawkes process if the point process admits a conditional intensity function of the form:

$$\lambda(t) = \mu(t) + \sum_{T_i<t} \phi(t - T_i, \xi_i) \tag{12}$$

Marks are typically assumed to be independent and identically distributed random variables. Moreover, it is often chosen a kernel with a factorized form that separates the effect of time from that of the mark:

$$\lambda(t) = \mu(t) + \sum_{T_i<t} \phi(t - T_i)\,\chi(\xi_i) \tag{13}$$

Typically, $\chi(\xi_i) = \xi_i^{\nu}$, with ν being a parameter that can take values across the entire real line, but is generally expected to be greater than 0 (one can decide whether to directly impose the positivity constraint or not).

2.3 *Estimation methods*

We now discuss the problem of estimating the parameters of a Hawkes model from data. The most common methods for fitting a Hawkes process are Maximum Likelihood Estimation (MLE), Generalized Method of Moments (GMM), Expectation Maximization (EM) algorithm, and Bayesian inference. MLE consists of selecting parameters that maximize the likelihood of the observed data. It is a very popular and flexible method, but computationally expensive and therefore slow on large datasets. In this article, we propose an innovative MLE implementation that mitigates computational cost by leveraging the optimized tensor computation capabilities of PyTorch. We present the theory regarding MLE for point processes in Subsection 2.3.1, followed by details on our MLE implementation in Subsection 2.3.2.

2.3.1 *Point processes maximum likelihood estimation*

The general form of the likelihood for simple point processes is provided by the following theorem.

Theorem 2 (Point Process Likelihood) Let N_t be a simple point process with conditional intensity $\lambda(\cdot)$ and compensator $\Lambda(\cdot)$. The likelihood function L for N_t on the time period $[0, T]$ is

$$L = \left[\prod_{i=1}^{N_T} \lambda(T_i)\right] e^{-\Lambda(T)}. \tag{14}$$

where $T_1, \ldots, T_{N_T}$ are the arrival times on $[0, T]$.

This formula also applies to a marked Hawkes process, with the only difference that the conditional intensity depends not just on the arrival times of past events, but also on their marks.

Usually the optimization problem for this likelihood does not have a closed-form solution, and therefore it is necessary to proceed numerically. Newton or quasi-Newton methods are popular choices for the numerical likelihood optimization. In quasi-Newton methods like the Broyden-Fletcher-Goldfarb-Shanno method (BFGS) or the Limited-memory BFGS (L-BFGS), the inverse of the Hessian is constructed efficiently and iteratively through successive approximations, rather than being calculated explicitly. Since the parameters to be estimated are defined over intervals (typically the positive real numbers), the constrained versions of these methods for interval-bounded parameters must be used, for example the L-BFGS-B algorithm [18].

For numerical reasons, it is usually preferable to maximize the log-likelihood

$$\ell = \sum_{i=1}^{N_T} \ln(\lambda(T_i)) - \Lambda(T). \tag{15}$$

rather than the likelihood itself, or, alternatively, to minimize the negative log-likelihood (i.e. the log-likelihood with the sign reversed).

2.3.2 *Innovative efficient MLE implementation*

MLE for Hawkes processes presents certain challenges. First, at each iteration of the optimization algorithm, it is necessary to calculate the log-likelihood, which, in turn, requires computing the conditional intensity $\lambda(t) = \mu(t) + \sum_{T_i<t} \phi(t - T_i, \xi_i)$ for each occurrence time $T_i < t$, resulting in an overall numerical complexity of $O({N_T}^2)$ per iteration. Additionally, Newton-style methods require partial derivatives and the Hessian of the function to be optimized. Quasi-Newton methods construct the Hessian approximately, but still need the partial derivatives.

In this work, we present an innovative and efficient implementation of MLE for Hawkes processes, designed to address these two challenges. The implementation is coded in Python and leverages the PyTorch library [19] to achieve two main benefits:

• PyTorch allows for obtaining exact derivatives for any form of the log-likelihood efficiently and without needing to provide the analytical form of the function, thanks to automatic differentiation [20].

• PyTorch has a built-in parallelization of tensor calculations, significantly increasing computational efficiency.

At the time of writing this paper, PyTorch does not provide constrained optimization algorithms. For this reason, in our implementation, all operations for calculating the likelihood and its derivatives are performed using PyTorch tensors, but the optimization leverages SciPy's L-BFGS-B algorithm [21]. The main drawback is that, in this way, the Graphics Processing Unit (GPU) cannot be utilized during optimization. However, the performance is very good even using the Central Processing Unit (CPU).

2.4 *Model diagnostic*

In this section we present tools for assessing the goodness of fit of Hawkes models with respect to data. Using MLE, a Hawkes model is trained by maximizing the log-likelihood l (or minimizing the negative log-likelihood). However, these quantities do not provide an absolute measure of how good a model is, and neither are they reliable metrics for

comparing different models, as they depend on the number of data points and parameters. To compare models trained on datasets with different sizes, the average log-likelihood $\bar{l}$ can be used:

$$\bar{l} = \frac{l}{n} \tag{16}$$

The negative log-likelihood is obtained by dividing the log-likelihood by the number of data points n, thus removing the dependency on sample size. Although it can provide useful insights, this metric should be interpreted with caution, as can have high variability, and it does not account for the number of model parameters.

The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) adjust the value of the log-likelihood l by taking into account the number k of parameters of the model:

$$\mathrm{AIC} = 2k - 2l \qquad \mathrm{BIC} = k\ln(n) - 2l \tag{17}$$

Empirically, it has been observed that the BIC tends to penalize more models with a larger number of parameters. AIC and BIC are tools suited for comparing different models on the same dataset, but they are not appropriate for models trained on different sample sizes, nor do they provide any indication of the model's absolute goodness.

For Hawkes processes, and more generally for point processes, tools to assess goodness of fit in an absolute way are provided by the Random Time Change Theorem, which we present here in the extended formulation given in [14].

Theorem 3 (Random Time Change Theorem) Consider an unbounded, increasing sequence of time points $\{t_1, t_2, \dots\}$ in the half-line $(0, \infty)$, and a monotonic, continuous compensator $\Lambda(\cdot)$ such that $\lim_{t\to\infty} \Lambda(t) = \infty$ almost surely. The transformed sequence $\{t_1^*, t_2^*, \dots\} = \{\Lambda(t_1), \Lambda(t_2), \dots\}$, whose counting process is denoted N_t^*, is a realisation of a unit rate Poisson process if and only if the original sequence $\{t_1, t_2, \dots\}$ is a realisation from the point process defined by $\Lambda(\cdot)$.

We recall that the compensator $\Lambda(\cdot)$ of a point process is the function defined in 2. The Random Time Change Theorem states that a generic point process (and therefore a Hawkes process) fits a sequence of time points $\{t_1, t_2, \dots\}$ if and only if the transformed sequence obtained through the process's compensator is a realization of a Poisson process. Thanks to this theorem, it is possible to reduce the diagnostics of a Hawkes model to that of a simple Poisson model. In particular, from Proposition 1, we know that inter-arrival times of a Poisson process are independent and identically distributed exponential random variables. Therefore, the model fits well if and only if the inter-arrival times $\{\tau_1, \tau_2, \tau_3, \dots\} = \{t_1^*, t_2^* - t_1^*, t_3^* - t_2^*, \dots\}$ of the transformed sequence are exponentially distributed. This can be verified quantitatively through a Kolmogorov-Smirnov (KS) test. However, this test tends to be too restrictive for real word data, as highly sensitive to outliers. A possible alternative is the Quantile-Quantile (Q-Q) plot. In the Q-Q plot, the theoretical quantiles of the exponential distribution are plotted against the empirical quantiles of the interarrival times. If the data follow the distribution, the points should align along the bisector of the first quadrant. This qualitative approach is less sensitive to outliers but requires human evaluation, making it unsuitable for automatic assessments and introducing potential subjective biases.

2.5 *EUA futures models*

We aim to model the occurrence times $T_1, \dots, T_n$ of transactions of EUA futures in the trading hours from January 2023 to May 2023 (each hour considered separately). This period was chosen simply because it was the most recent at the beginning of the study. The trading hours for the EUA futures range from 8:00 AM to 6:00 PM (Central European Time, CET/CEST), amounting to a total of 10 trading hours per day.

We consider $T_1, \dots, T_n$ as the transaction times measured in seconds elapsed since the start of the trading hour. The data provides the transaction timestamps with millisecond precision. There are transactions associated with the same timestamp, but different trade IDs. These transactions likely involve large market orders that match multiple limit orders in the order book. Since the likelihood formulation provided by Theorem 2 requires the point process to be simple and

therefore each timestamp to be distinct, simultaneous transactions are treated as a single event, and the corresponding volumes are summed together.

Five different point processes are fitted on the data: a Poisson process, a Hawkes model with exponential kernel, a Hawkes model with power-law kernel, a marked Hawkes model with exponential kernel, and a marked Hawkes model with power-law kernel. The models are then compared using the metrics described in Subsection 2.4. Additionally, the Poisson process, which as described in Subsection 2.1.2 assumes independence of occurrences, is used as a baseline model to test whether the hypothesis of self-excitation underlying all the Hawkes models is satisfied.

For all Hawkes models, we assume the background intensity $\mu(t)$ to be constant during the trading hour ($\mu(t) = \mu$). For the marked Hawkes models, we also assume the factorized form for the kernel described in Subsection 2.2.3, which separates the effect of the transaction time from that of the volumes

$$\lambda(t) = \mu + \sum_{T_i<t} \phi(t - T_i)\,\chi(\xi_i) \tag{18}$$

with $\phi(x) = \alpha e^{-\beta x}$ in case of exponential kernel, $\phi(x) = \dfrac{k}{(c+x)^p}$ for the power-law kernel, and $\chi(\xi_i) = \xi_i^{\nu}$. We do not impose the positivity constraint on ν because, even though we expect it to be greater than 0, we want to verify this hypothesis.

Hawkes models are trained using the innovative MLE implementation described in Subsection 2.3.2, whereas for the Poisson models, which have a closed-form solution for the likelihood optimization, the constant intensity parameter λ is set equal to the empirical intensity of transactions during the trading hours.

3. Results

3.1 *Descriptive results*

During the period from January to May 2023 considered in the study, EUA was experiencing a phase of strong price growth, reaching an all-time high in early March 2023. Subsequently, prices started to decline, returning by early June to the levels observed at the beginning of January. The price chart for EUA futures in the period is shown in Figure 1.

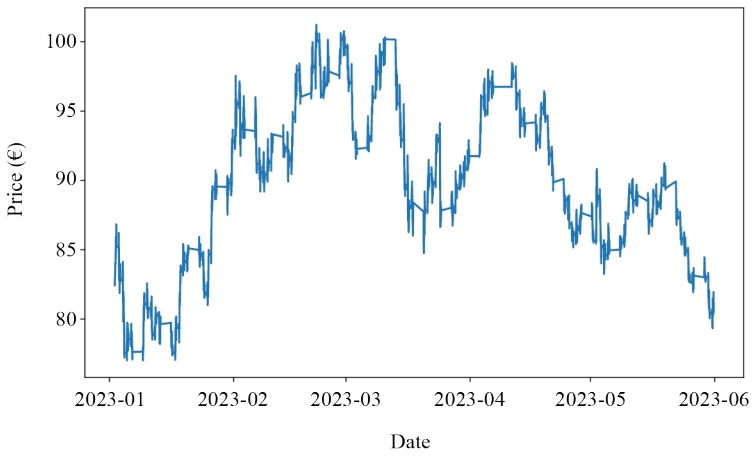

Figure 1. EUA futures price from January to May 2023

Our work does not focus on prices, but rather on transactions and corresponding volumes. These tend to exhibit quite high daily variability, as shown in Figure 2.

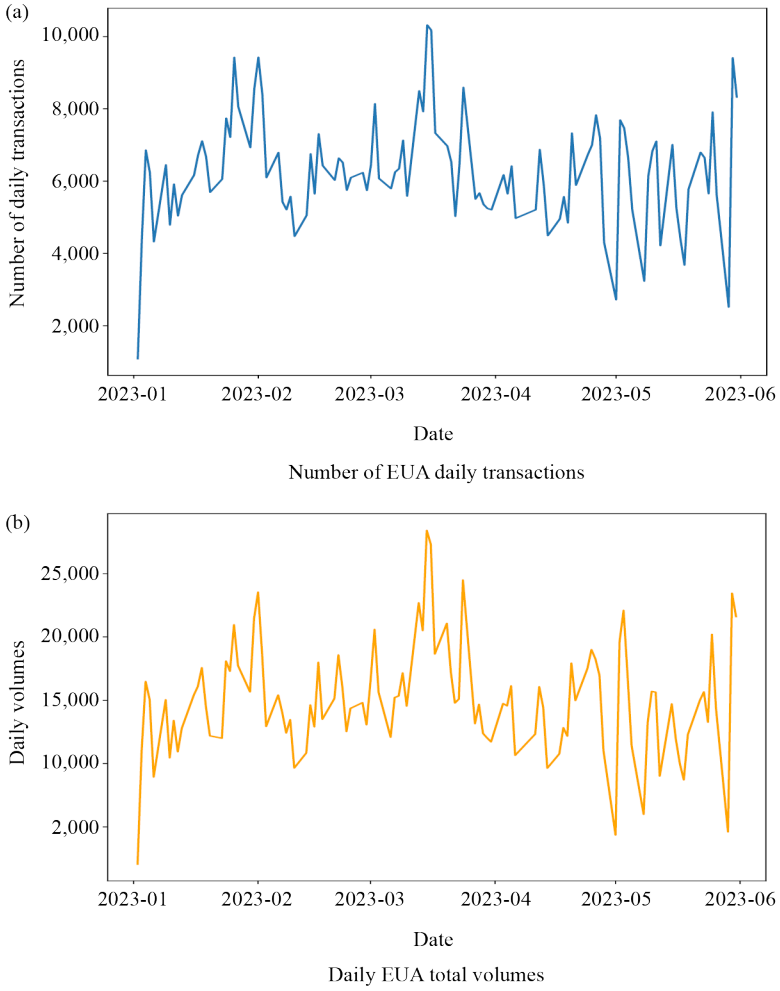

Figure 2. EUA daily transactions and volumes from January to May 2023

The first week of January shows very low numbers of transactions and volumes, likely due to the holiday season, and is therefore excluded from subsequent analyses.

The boxplots in Figure 3 reveal effects on transactions and volumes related to the weekday and the hour.

Weekday and hourly effects are confirmed by multivariate Negative Binomial (NB) regressions on the number of transactions and volumes, the summaries of which are reported in Table 1.

The intercepts in the table represent the average number of transactions and total volumes for 8 a.m. on Monday, taken as reference trading hour. The rate ratios represent the multiplicative effect applied to the intercept related to different weekdays and hours. All these effects are highly statistically significant (marginal p-values < 0.001), except for the effect related to Friday, where no significant differences are observed compared to Monday. Finally, the variable "Weeks from Start" represents the number of trading weeks elapsed since the beginning of the period. The related rate ratios indicate the presence of a minor trend over the considered period (-0.8% per week on the number of transactions and -0.6% per week on volumes).

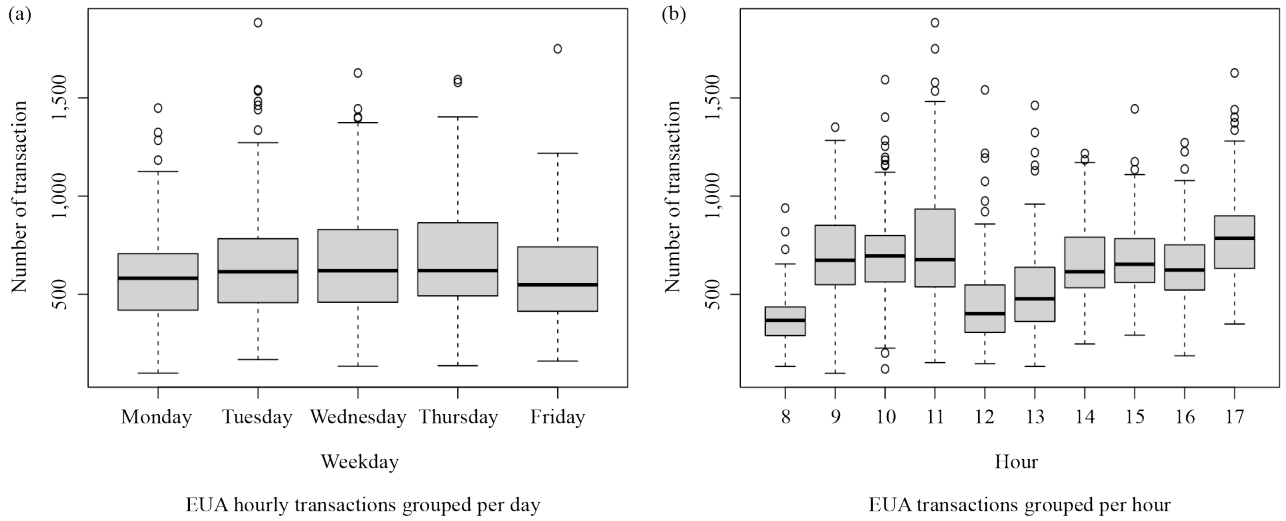

Figure 3. Boxplots of the number of hourly transactions grouped per weekday and hour

Table 1. Summary of NB regression for transactions and volumes

	Transactions model		Volumes model	
Variable	Rate ratio	*p*-value	Rate ratio	*p*-value
Intercept	373.668	< 0.001	785.630	< 0.001
Tuesday	1.123	0.001	1.174	< 0.001
Wednesday	1.137	< 0.001	1.209	< 0.001
Thursday	1.155	< 0.001	1.164	< 0.001
Friday	0.996	0.915	1.006	0.900
Hour 9	1.851	< 0.001	2.197	< 0.001
Hour 10	1.876	< 0.001	2.261	< 0.001
Hour 11	2.042	< 0.001	2.630	< 0.001
Hour 12	1.232	< 0.001	1.392	< 0.001
Hour 13	1.402	< 0.001	1.501	< 0.001
Hour 14	1.775	< 0.001	1.799	< 0.001
Hour 15	1.810	< 0.001	1.850	< 0.001
Hour 16	1.741	< 0.001	1.720	< 0.001
Hour 17	2.140	< 0.001	2.140	< 0.001
Weeks from start	0.992	< 0.001	0.994	0.011

3.2 *Models results and comparison*

Excluding the first week of January, each model discussed in 2.5 is calibrated on 1,010 different trading hours. The calibration over such a large number of hours was made possible thanks to our innovative MLE implementation described in Subsection 2.3.2 and would have been infeasible within a reasonable timeframe using a non-optimized procedure. The MLE reaches convergence for all the models with the exponential kernel (marked and non-marked), while it does

not converge 117 times (11.6%) for the power-law model and 28 times (2.8%) for the marked power-law model. We use logistic regression to investigate the causes of non-convergence (indicated as 1 in the target variable) for models with power-law kernels and we report the Odds Ratios (ORs) and p-values for the related variables. Non-convergence is associated with outliers in inter-arrival times, that is long periods during which no transactions occur (OR 1.033 for every 10-second increase in the maximum inter-arrival time recorded during the trading hour, p-value 0.002). The use of a marked model has a strong protective effect against non-convergence (OR 0.141, p-value < 0.001). By means of an interaction term, we investigate whether outliers in volumes may influence the convergence of marked models. The regression summary indicates a positive association between transactions with high volumes and non-convergence (OR 1.072 for 10-unit increase in the maximum volume exchanged in a single transaction during the trading hour, p-value 0.049).

The convergence issue does not concern the Poisson model since, as explained in Subsection 2.5, it is trained using a closed formula solution. In the following, data related to non-converged models are removed and treated as missing values.

Next, we compare the five models by analyzing the AIC and BIC values obtained on the different trading hours of the period. The boxplots of AIC and BIC for the five models, calculated over the 1,010 trading hours and excluding the non-converged models, are reported in Figure 4.

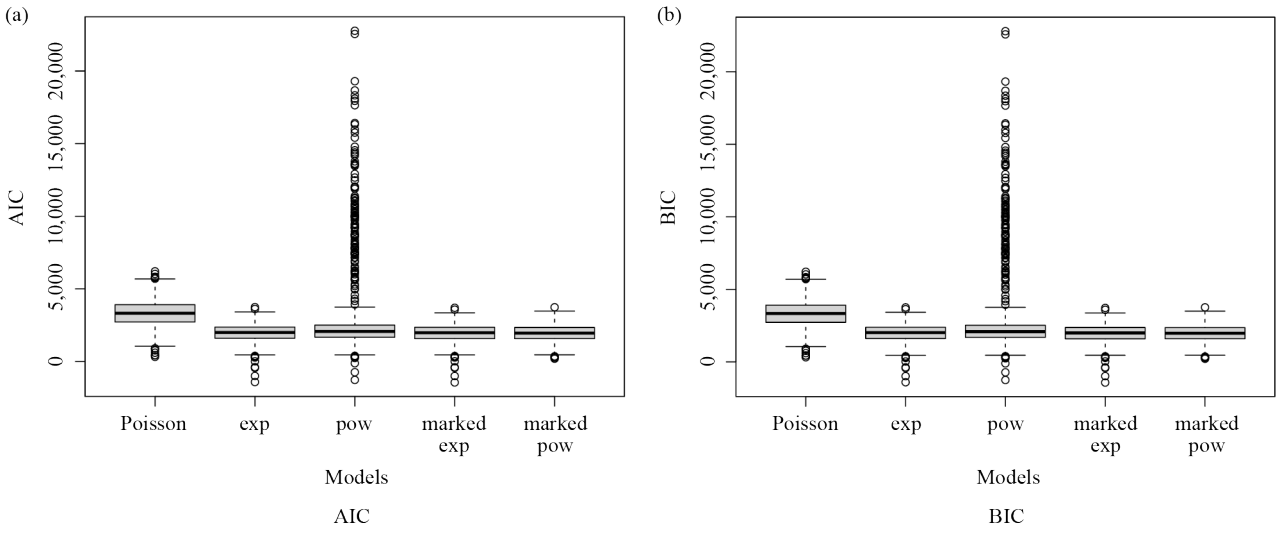

Figure 4. Boxplots of AIC and BIC of the models on the different trading hours

In Figure 4, the large number of outliers for the power-law model stands out, probably associated with solutions that the optimization algorithm considered converged, but are actually far from the true optimum. The AIC and BIC for the Poisson model are significantly larger than those of all the other Hawkes models (Bonferroni-adjusted p-values < 0.001 for both AIC and BIC, tests conducted using the one-sided Wilcoxon test for paired data). This indicates that the Hawkes models perform better and therefore the data exhibit self-exciting characteristics. The marked power-law model has AIC and BIC significantly lower than all the other non-marked models (Bonferroni-adjusted p-values obtained with the same tests < 0.001 for both AIC and BIC), while it doesn't appear to be a significant difference with the marked exponential model (p-value 0.218 for AIC and 0.410 for BIC).

3.3 *Hawkes model parameters distribution*

The two marked models (exponential and power-law) exhibit a similar goodness of fit. Given equal performance, the exponential model is preferable for several reasons:

• It is simpler, has fewer parameters, and these parameters are easily interpretable;
• It possesses the Markov property, which can be exploited during training and simulations;
• It is more robust to non-convergence issues.

In this section, we investigate the distribution of the parameters of the marked exponential Hawkes model and their dependence on exogenous factors. This analysis provides insights into market conditions and opens the possibility of calibrating the parameters of the Hawkes models a priori, which is crucial for practical applications.

3.3.1 *The μ parameter*

With no surprise, the background intensity parameter μ is closely related to the average number of transactions per trading hour, which, as discussed in 3.1, is in turn linked to the hour and the weekday. The summary of the gamma regression between the estimated μ and the weekday hour is reported in Table 2.

Table 2. Summary of the gamma regression for the background intensity μ

Variable	Rate ratio	*p*-value	Variable	Rate ratio	*p*-value
Intercept	0.070	< 0.001	Hour 11	1.879	< 0.001
Tuesday	1.071	0.046	Hour 12	1.129	0.012
Wednesday	1.082	0.022	Hour 13	1.297	< 0.001
Thursday	1.138	< 0.001	Hour 14	1.635	< 0.001
Friday	0.980	0.573	Hour 15	1.687	< 0.001
Hour 9	1.652	< 0.001	Hour 16	1.616	< 0.001
Hour 10	1.674	< 0.001	Hour 17	1.963	< 0.001

3.3.2 *The α and β parameters*

The α and β parameters are the true core of the Hawkes process. In a non-marked Hawkes process, α represents the instantaneous effect that the occurrence of a transaction has on the conditional intensity. In a marked Hawkes process, this instantaneous effect is also mediated by the volume of the transaction and the parameter ν, but the interpretation of α remains similar. On the other hand, β is the time constant, which determines the decay speed of the self-exciting effect.

We analyze these parameters together because, empirically, we have observed that in EUA futures α and β have a strong relationship, as evident from the pairplot of the parameters in Figure 5. Additionally, their distribution form two well-defined clusters, which can be perfectly separated by values lower and higher than $\beta = 200$ and are distinguished in Figure 5 by the red and blue colors.

The first cluster, defined by trading hours with $\beta < 200$, is characterized by smaller instantaneous effects and longer transaction impacts, lasting to a few seconds. The other cluster, with estimated $\beta \geq 200$, exhibits stronger effects, which, however, decay rapidly and dissipate within fractions of a second. The "slower-decaying" cluster ($\beta < 200$) is the largest and is associated with 713 trading hours (71%), compared to 297 hours (29%) for the "fast-decaying" cluster. With logistic regression, we investigate possible factors influencing membership in the "slower-decaying" cluster, taken as 1 in the target variable. The regression summary with the corresponding odds ratios and *p*-values is reported in Table 3.

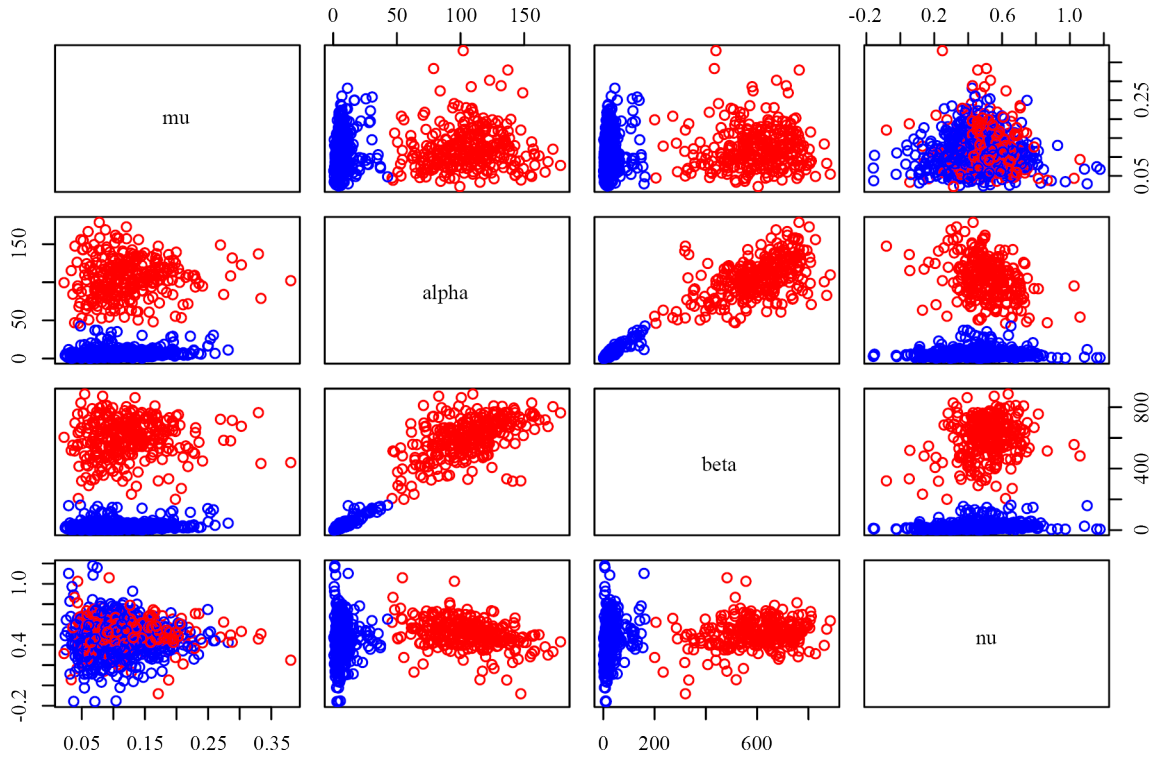

Figure 5. Pairplot of the parameters calibrated on the 1,010 trading hours. The points with $\beta < 200$ are represented in blue, and those with $\beta \geq 200$ are represented in red

Table 3. Summary of the logistic regression for the membership in the "slower-decaying" cluster

Variable	Odds ratio	*p*-value	Variable	Odds ratio	*p*-value
Intercept	1.083	0.748	Hour 11	1.272	0.396
Tuesday	0.925	0.732	Hour 12	1.564	0.118
Wednesday	1.106	0.661	Hour 13	2.035	0.015
Thursday	1.055	0.816	Hour 14	5.877	< 0.001
Friday	0.828	0.414	Hour 15	6.402	< 0.001
Hour 9	1.325	0.322	Hour 16	14.986	< 0.001
Hour 10	1.704	0.064	Hour 17	3.031	< 0.001

We find a strong effect related to the time of day, which, however, differs from that observed for the number of transactions and volumes. In particular, there is a higher probability of belonging to the "slower-decaying" cluster during the afternoon (all hours from 1 PM onward are significant), with a peak at 4 PM, where an odds ratio of 15 is observed with respect to the reference trading hour 8 a.m. on Monday. The different distribution of clusters across the trading hours is also evident from the barplot provided in Figure 6.

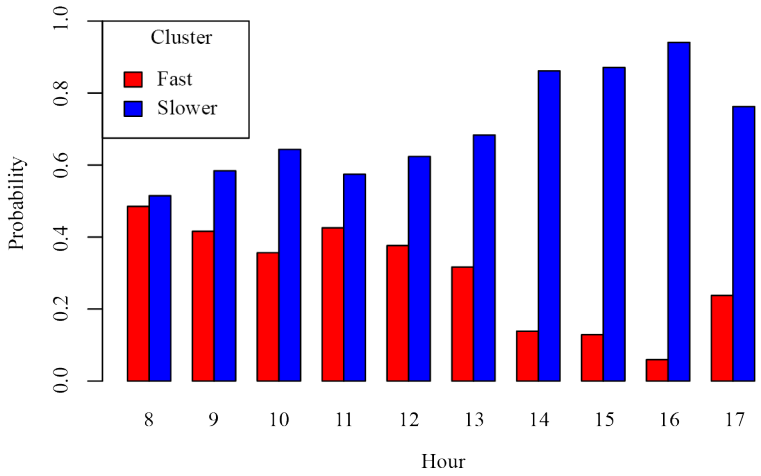

Figure 6. Empirical probability of cluster membership per hour

A simple linear regression between α and β yields an R^2 of 0.95, indicating a strong linear relationship between the two variables. However, from the pairplot in Figure 5, a phenomenon of heteroscedasticity and increasing variance with the growth of the variables is evident. We therefore proceed with a log-log regression, whose summary reports an R^2 of 0.98. Such a high R^2 is rare in real-world data and indicates an almost deterministic relationship between the two variables. The estimated coefficient of the regression is 0.673, which can be approximated to $\frac{2}{3}$ when rounded to the second decimal place. Rewriting α as a function of β, we can therefore obtain the following simplified form of the Hawkes model, reducing the number of parameters without losing precision:

$$\lambda(t) = \mu + \sqrt[3]{\beta^2} \sum_{T_i<t} e^{-\beta(t-T_i)} \xi_i^{\nu} \tag{19}$$

In this model, the ratio between $\alpha = \sqrt[3]{\beta^2}$ and β is equal to $\frac{1}{\sqrt[3]{\beta}}$. In a non-marked Hawkes process with exponential kernel, the ratio α/β represents the branching ratio of the process, i.e., the average percentage of transactions triggered as a consequence of others. In a marked process, the interpretation is slightly different due to the influence of the mark and the parameter ν, but it remains similar.

3.3.3 *The ν parameter*

The histogram of the values of ν calibrated on the trading hours is reported in Figure 7.

As expected, even without enforcing the positivity constraint, ν is greater than 0 in the vast majority of cases. Additionally, it appears to be symmetrically distributed around the mean value of 0.5, which corresponds to an effect of volumes equal to their square root.

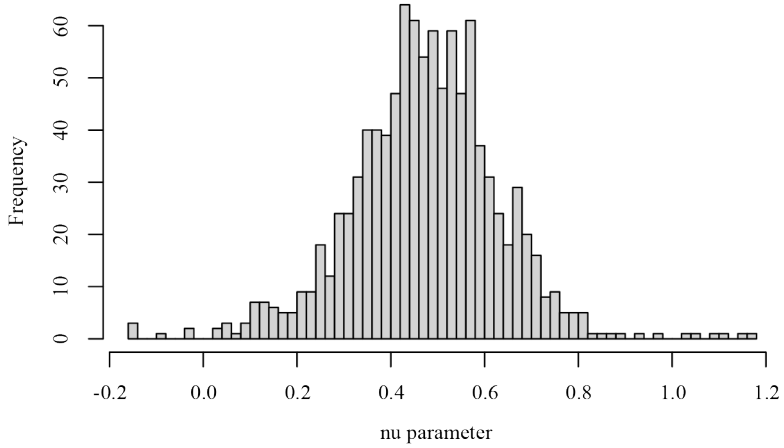

Figure 7. Histogram of the ν parameter distribution

3.4 *In-depth study of selected days*

As discussed in Subsection 2.4, the average log-likelihood is a metric that "normalizes" the log-likelihood by the number of observations. This allows for comparing the performance of a model across different datasets, even when they vary in size. Using the log-likelihood, we rank the marked exponential models by goodness of fit and extract the ones corresponding to the quartiles. In this way, we present three examples of trading hours in which the models perform respectively worse, in line, and better than average. For these hours, we also report the performance of the other models for comparison (Poisson, non-marked exponential, marked and non-marked power-law).

3.4.1 *First example: Below-average model*

As the first quartile, we find 8 AM on March 1, 2023, with an associated average log-likelihood for the marked exponential model of -2.098. The calibrated parameters for the model are: $\mu = 0.044$, $\alpha = 95.230$, $\beta = 556.227$, $\nu = 1.025$. It is a classic start-of-day hour, with few transactions (as we can see from the low background intensity value), and belonging to the "fast-decaying" cluster, which we have seen to be more common in the first trading hour. On the other hand, the parameter ν takes on an extreme value, far from the mean of 0.5 and corresponding to a volume effect that grows linearly with the number of traded units. The AIC and BIC values for the 5 models calibrated on this hour are reported in Table 4.

Table 4. AIC and BIC values for the below-average performing models

	Poi	Exp	Pow	Marked exp	Marked pow
AIC	1,632.47	918.20	1,024.24	901.96	1,116.29
BIC	1,635.83	928.28	1,037.68	915.40	1,133.10

The model with the lowest AIC and BIC values is the marked exponential model, likely because the exponential kernel effectively captures the rapid decay of the self-exciting effect characteristic of this trading hour. The non-marked exponential model performs similarly, possibly because, during this particular hour, many transactions consist of a single unit of volume (this also justifies the unitary value of the parameter ν). The Poisson model stands out for its poorer performance, indicating that transactions tend to be clustered and Hawkes models are more suitable for describing their occurrence. In Figure 8, the Q-Q plots based on the Random Time Change Theorem described in Subsection 2.4 are shown (the Q-Q plot of the power-law model has been omitted for better graphical representation).

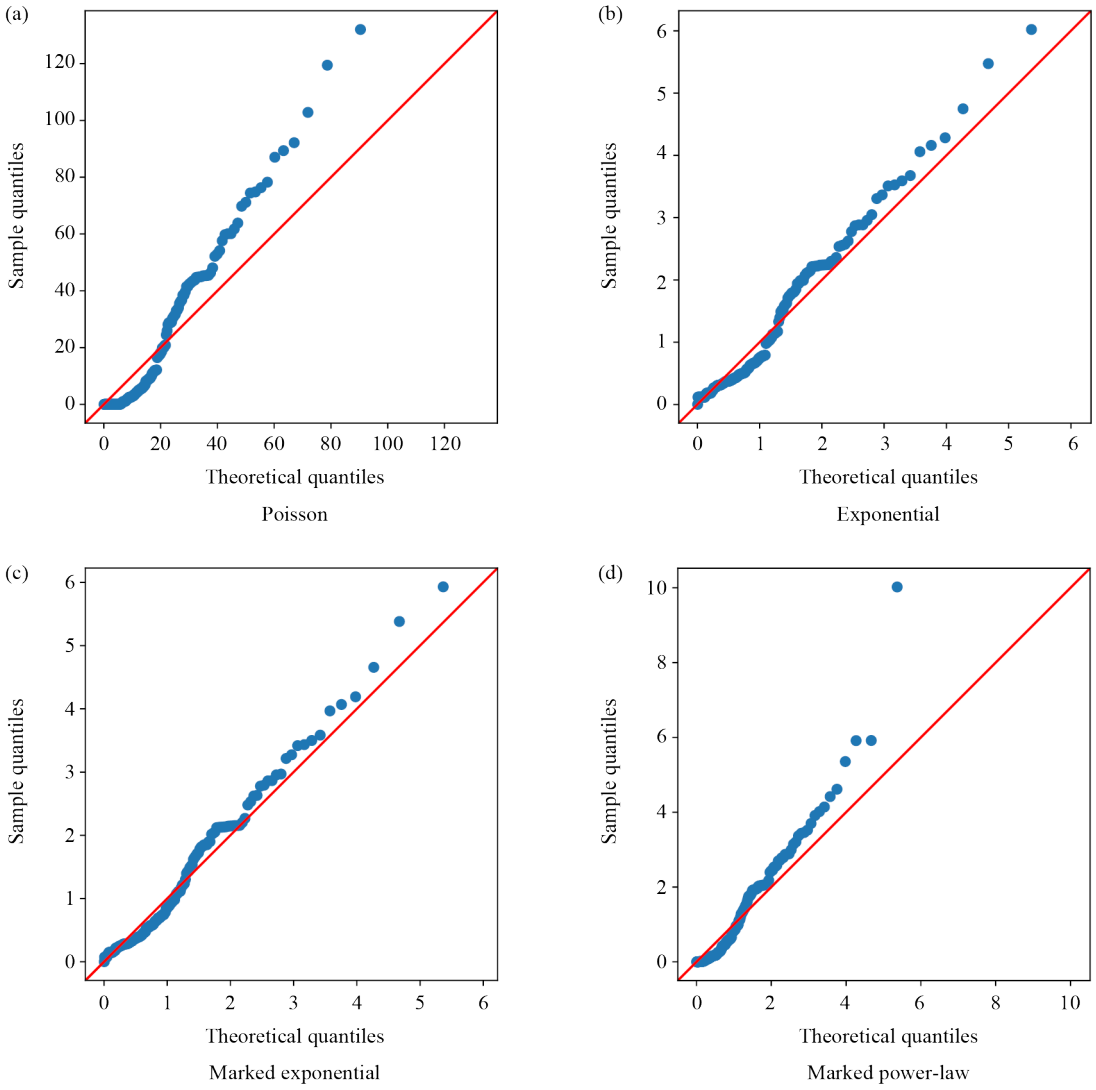

Figure 8. Q-Q plots of the below-average performing models

Despite their below-average performance, these Hawkes models show Q-Q plots that align fairly well along the bisector, especially those with an exponential kernel. The Poisson model, on the other hand, shows a strong deviation from the bisector, indicating that it fails to capture the clustered nature of the transactions.

3.4.2 *Second example: Average-performing model*

As the median, we find 4 PM on January 26, 2023, with an associated average log-likelihood for the marked exponential model of -1.717. The calibrated parameters for the model are: $\mu = 0.148$, $\alpha = 7.504$, $\beta = 31.789$, $\nu = 0.538$. This is a classic 4 PM trading hour, with a medium-high number of transactions, and belonging to the "slower-decaying" cluster, which we have seen to be typical of this trading hour. The parameter ν takes a common value near the mean of 0.5, corresponding to a volume effect that grows with the square root of the number of traded units. The AIC and BIC values for the 5 models calibrated on this hour are reported in Table 5.

Table 5. AIC and BIC values for the average performing models

	Poi	Exp	Pow	Marked exp	Marked pow
AIC	3,923.30	2,699.55	2,696.51	2,658.36	2,648.86
BIC	3,927.95	2,713.50	2,715.10	2,676.95	2,672.10

In this case, the marked power-law model performs a little better than the marked exponential, while the Poisson model is confirmed to be by far the worst.

The Q-Q plots of the models are reported in Figure 9.

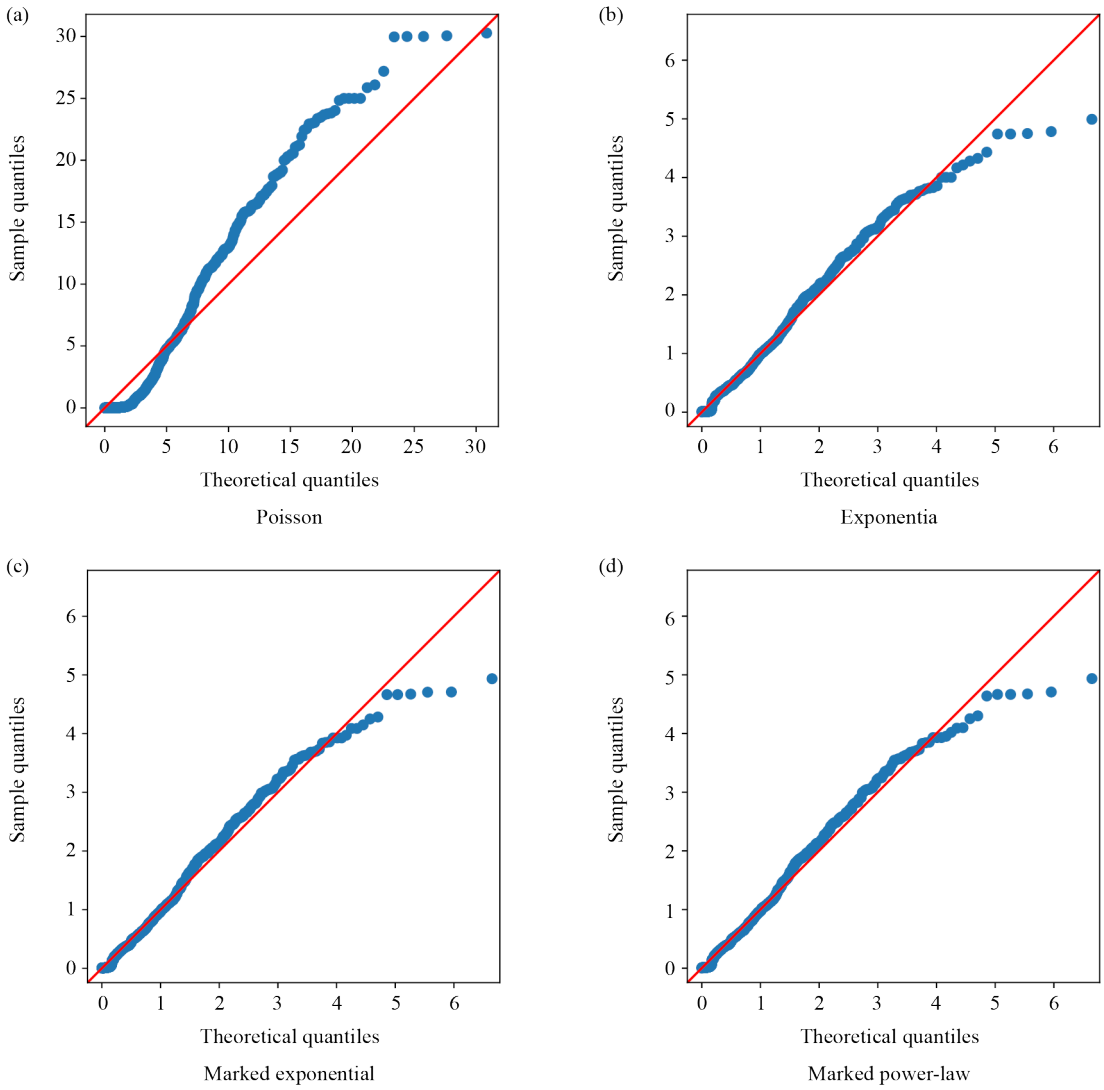

Figure 9. Q-Q plots of the average-performing models

The Poisson model deviates markedly from the bisector, while the Hawkes models tend to align quite well, exhibiting similar plots.

3.4.3 *Third example: Above-average performing model*

As the third quartile there is 5 PM on May 18, 2023, with an associated average log-likelihood for the marked exponential model of -1.367. The calibrated parameters for the model are: $\mu = 0.160$, $\alpha = 5.050$, $\beta = 13.913$, $\nu = 0.329$. The intensity of the transactions is high as the majority of the closing hours and the effects of the transactions last for a relatively long time. The AIC and BIC values of the models are reported in Table 6.

Table 6. AIC and BIC values for the above-average performing models

	Poi	Exp	Pow	Marked exp	Marked pow
AIC	4,581.75	2,779.36	2,707.67	2,760.37	2,047.88
BIC	4,586.67	2,794.10	2,727.33	2,780.03	2,072.45

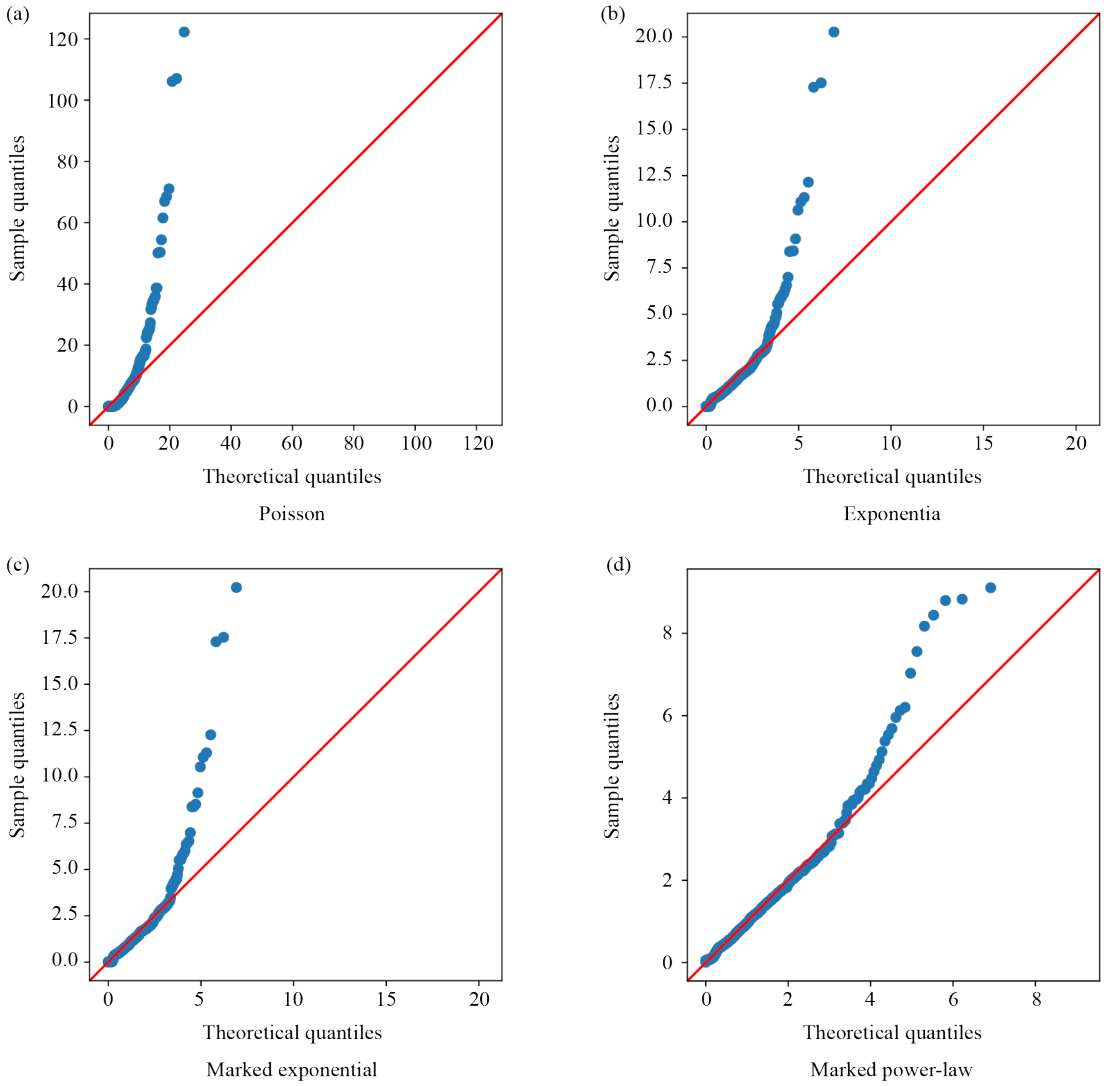

Figure 10. Q-Q plots of the above-average performing models

The marked power-law model performs better than the other models, possibly because it captures the longer-lasting effect of the transactions more effectively. The Q-Q plots of the models are reported in Figure 10.

Despite the average log-likelihood classifying these models as above-average performing, the Q-Q plots appear rather poor. This underscores the absence of a single perfect performance metric and emphasizes the need for different goodness of fit estimation methods to evaluate the models effectively.

4. Discussion

In our work, we provided an introduction to Hawkes processes focused on applications, and presented a new and efficient implementation of MLE for parameter estimation. We then applied Hawkes processes to model the occurrence of transactions in EUA futures. Our optimized MLE implementation allowed us to calibrate the models on a large number of trading hours and to derive interesting and robust statistics on the goodness of fit of the various models. Specifically, Hawkes models consistently outperform the Poisson model, confirming the hypothesis of transaction clustering in EUA futures. Moreover, marked models that account for the influence of volumes provide a better fit than non-marked models. Among marked models, the power-law kernel, which has been shown in many studies to fit empirical data better, does not seem to consistently outperform the exponential kernel over the analyzed period. Therefore, we decided to further investigate the parameter distribution of the marked exponential model, which, while offering comparable performance, has the advantages of being simpler, more interpretable, has better convergence, and benefits from the Markov property. The background intensity parameter μ is closely related to the varying number of transactions characterizing the different trading hours across the days of the week, as discussed in Subsection 3.1. The parameters α and β follow an almost deterministic law, which can be expressed with the surprisingly elegant formula $\alpha = \sqrt[3]{\beta^2}$. Additionally, they are distributed in two main clusters: one with β values greater than 200, and another with β less than 200. Trading hours belonging to the cluster with $\beta \geq 200$ are more common in the early morning hours (particularly the opening hour) and exhibit strong self-exciting effects in transactions, which, however, decay very quickly (on the order of thousandths/hundredths of a second). Modern High-Frequency Trading (HFT) infrastructures have short response times, but we doubt that these clustered transactions result from algorithmic responses to other transactions. Rather, we think that they are the result of single market orders that match hidden limit orders (iceberg orders), which, unlike visible limit orders, may be recorded by the system with a delay of a few milliseconds. In this case, the observed behavior cannot be defined as a "true" self-exciting effect, since the clustered transactions are the result of single trading actions, split into multiple and nearly simultaneous transactions. The cluster with $\beta < 200$ is more interesting and is typical of the afternoon hours, particularly 4 PM. Transactions during this period exhibit a weaker self-exciting effect, but that lasts longer and up to some seconds. This timespan is sufficient for an HFT algorithm to execute transactions in response to prior ones, and we consider probable that a fraction of transactions is indeed triggered by others. The question of why this behavior is characteristic of the afternoon and in particular of 4 PM is interesting and warrants further investigation. Our intuition is that it may be linked to the opening of U.S. markets, which have a strong influence on European markets, including EUA.

It is possible that even in this "slower-decaying" cluster, some transactions close in time arise from the breakdown of market orders like in the "fast-decaying" cluster, but this mechanism overlaps and blends with the true self-exciting effect that triggers additional transactions. To distinguish the two dynamics, one could consider a Hawkes process with a kernel composed of two different exponentials, one with rapid and the other with slower decay, or alternatively a kernel formed by an exponential and a power-law function. Such a model could provide an excellent fit for situations like the one described in Subsection 3.4.3, where, despite the average log-likelihood being better than that observed during the majority of trading hours of the period, the Q-Q plot reveals clustering dynamics that are not yet fully captured by a single-kernel model.

Coming back to the formula $\alpha = \sqrt[3]{\beta^2}$, it is interesting that this relationship appears to hold in all cases, regardless of the cluster membership. We don't believe in coincidences and we think that such an elegant and simple formula may have an intrinsic explanation rooted in the microstructure of the market. Further studies are needed to determine the meaning behind this formula and whether this relationship also holds in other contexts.

Finally, let us discuss ν, the parameter related to the effect of volumes on the kernel. During the studied period, ν appears to be symmetrically distributed around the value 0.5, which corresponds to a square root volumes effect. In finance, the square root law for volumes is a well-known formula for price impact of meta-orders [22, 23], and it is a consequence of the v-shaped profile of the latent liquidity. The fact that the square root law for volumes appears to hold also for the self-exciting effect of single transactions on future transactions themselves is interesting and calls for further investigation.

5. Conclusion and future work

Some future work has already been outlined in Section 4, like delving deeper into the significance behind the relationship $\alpha = \sqrt[3]{\beta^2}$ for marked exponential parameters and verifying if it can be extended to other contexts. Another point of interest is to further explore the square root law of volumes in the context of impact on future transaction occurrence, and investigate possible links to the price impact and the latent liquidity theory. Regarding the goodness of fit of the models, it would be interesting to check whether adding a double two-speed decay kernel for the hours of the "slower decaying" cluster, taking into account the distinct dynamics of transactions clustering as discussed in Section 4 could effectively improve the fitting. Further improvements to the model could be distinguishing "buy" from "sell" market orders, introducing a multivariate Hawkes model in which the two types of transactions are considered distinct events, possibly influencing each other. Another significant step forward would be the introduction of a price dynamic, either as an extra mark associated with the transactions or as additional components of the multivariate Hawkes process, where occurrences correspond to upward or downward price changes.

So far, we have focused on training Hawkes models on large datasets and evaluating their goodness of fit. However, Hawkes models also enable real-time short-term predictions through the conditional intensity $\lambda(t)$ and, additionally, the simulation of new data. Predicting and simulating financial data [24] are topics of great interest and in the future it is worth exploring these two applications.

Finally, in order to use Hawkes processes in real-world scenarios, it is essential to calibrate the parameters in advance with respect to the trading hour of interest, rather than retrospectively as done so far. Our studies on parameter distributions have revealed specific patterns, primarily related to the trading hour and the day of the week. These patterns, along with other potential factors like temporally autoregressive effects, could be leveraged to develop statistical and machine learning models for the a priori calibration of parameters. In future work, we aim to develop such models and evaluate the robustness of their predictions.

To conclude, in this paper we have provided a brief practical guide to the use of Hawkes models, along with an example application in modeling the occurrence of transactions in EUA futures. We have demonstrated how these models can accurately describe transaction times and extract informative parameters about market conditions. We think that, despite the exponential growth in the use of Hawkes processes in recent years, their application remains limited compared to their potential, and we believe this work can encourage more researchers and practitioners to dive further into these models.

Disclosure

This study was developed within the MUSA—Multilayered Urban Sustainability Action—project, funded by the European Union—NextGenerationEU, Project code ECS 00000037, under the National Recovery and Resilience Plan (NRRP) Mission 4 Component 2 Investment Line 1.5: Strenghtening of research structures and creation of R&D "innovation ecosystems", set up of "territorial leaders in R&D".

Conflict of interest

The authors declare no conflict of interest.

References

[1] Hawkes AG. Spectra of some self-exciting and mutually exciting point processes. *Biometrics*. 1971; 58(1): 83-90. Available from: https://doi.org/10.1093/biomet/58.1.83.

[2] Hawkes AG. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 1971; 33(3): 438-443. Available from: https://doi.org/10.1111/j.2517-6161.1971.tb01530.x.

[3] Adamopoulos L. Cluster models for earthquakes: Regional comparisons. *Journal of the International Association for Mathematical Geology*. 1976; 8: 463-475. Available from: https://doi.org/10.1007/BF01028982.

[4] Ogata Y. Estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*. 1978; 30: 243-261.

[5] Kwon J, Zheng Y, Jun M. Flexible spatio-temporal Hawkes process models for earthquake occurrences. *Spatial Statistics*. 2023; 54: 100728. Available from: https://doi.org/10.1016/j.spasta.2023.100728.

[6] Rizoiu MA, Mishra S, Kong Q, Carman M, Xie L. SIR-Hawkes: Linking epidemic models and Hawkes processes to model diffusions in finite populations. In: *Proceedings of the 2018 World Wide Web Conference*. Lyon, France; 2018. p.419-428.

[7] Rizoiu MA, Lee Y, Mishra S, Xie L. Hawkes processes for events in social media. In: *Frontiers of Multimedia Research*. Association for Computing Machinery and Morgan & Claypool; 2017. p.191-218.

[8] Brignone R, Sgarra C. Asian options pricing in Hawkes-type jump-diffusion models. *Annals of Finance*. 2020; 16(1): 101-119. Available from: https://doi.org/10.1007/s10436-019-00352-1.

[9] Bormetti G, Calcagnile LM, Treccani M, Corsi F, Marmi S, Lillo F. Modelling systemic price cojumps with Hawkes factor models. *Quantitative Finance*. 2015; 15(7): 1137-1156. Available from: https://doi.org/10.1080/14697688.2014.996586.

[10] Bacry E, Mastromatteo I, Muzy JF. Hawkes processes in finance. *Market Microstructure and Liquidity*. 2015; 1(1): 1550005.

[11] Hawkes AG. Hawkes processes and their applications to finance: a review. *Quantitative Finance*. 2018; 18(2): 193-198. Available from: https://doi.org/10.1080/14697688.2017.1403131.

[12] Laub PJ, Taimre T, Pollett PK. Hawkes processes. *arXiv:1507028222*. 2015. Available from: https://doi.org/10.48550/arXiv.1507.02822.

[13] Laub PJ, Lee Y, Taimre T. *The Elements of Hawkes Processes*. Springer; 2021.

[14] Daley DJ, Vere-Jones D. *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. New York, NY: Springer; 2003.

[15] Ogata Y. Seismicity analysis through point-process modeling: A review. *Seismicity Patterns, Their Statistical Significance and Physical Meaning*. 1999; 155: 471-507. Available from: https://doi.org/10.1007/s000240050275.

[16] Bacry E, Dayri K, Muzy JF. Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data. *The European Physical Journal B*. 2012; 85: 157. Available from: https://doi.org/10.1140/epjb/e2012-21005-8.

[17] Hardiman SJ, Bercot N, Bouchaud JP. Critical reflexivity in financial markets: a Hawkes process analysis. *The European Physical Journal B*. 2013; 86: 442. Available from: https://doi.org/10.1140/epjb/e2013-40107-3.

[18] Byrd RH, Lu P, Nocedal J, Zhu C. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*. 1995; 16(5): 1190-1208. Available from: https://doi.org/10.1137/0916069.

[19] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: An imperative style, high-performance deep learning library. In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc.; 2019. p.8026-8037.

[20] Baydin AG, Pearlmutter BA, Radul AA, Siskind JM. Automatic differentiation in machine learning: a survey. *Journal of Machine Learning Research*. 2018; 18(153): 1-43.

[21] Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*. 2020; 17: 261-272. Available from: https://doi.org/10.1038/s41592-019-0686-2.

[22] Mastromatteo I, Toth B, Bouchaud JP. Agent-based models for latent liquidity and concave price impact. *Physical Review E*. 2014; 89(4): 042805. Available from: https://doi.org/10.1103/PhysRevE.89.042805.

[23] Donier J, Bonart J, Mastromatteo I, Bouchaud JP. A fully consistent, minimal model for non-linear market impact. *Quantitative Finance*. 2015; 15(7): 1109-1121. Available from: https://doi.org/10.1080/14697688.2015.1040056.

[24] Bellomo M, Trimarchi S, Niccolai A, Lorenzo M, Casamatta F, Grimaccia F. A GAN data augmentation approach for trading applications in European carbon emission allowances. In: *2023 IEEE International Conference on Environment and Electrical Engineering and 2023 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*. Madrid, Spain: IEEE; 2023. p.1-5.