

# Hawkes processes in energy markets: modelling, estimation and derivatives pricing

Riccardo Brignone<sup>a</sup>, Luca Gonzato<sup>b</sup>, Carlo Sgarra<sup>c,\*</sup>

<sup>a</sup>*Department of Quantitative Finance, Institute for Economic Research, University of Freiburg,  
Rempartstr. 16, 79098 Freiburg i. Br., Germany*

<sup>b</sup>*Department of Statistics and Operations Research, University of Vienna, Kolingasse 14-16, 1090  
Vienna, Austria*

<sup>c</sup>*Department of Mathematics, Politecnico di Milano, Piazza Leonardo da Vinci, 32, I-20133 Milan, Italy*

---

## Abstract

The purpose of the present contribution is to illustrate the extensive use of Hawkes processes in modeling price dynamics in energy markets and to show how they can be applied for derivatives pricing. After a review of the literature devoted to the subject and on the exact simulation of Hawkes processes, we introduce a simple, yet useful, Hawkes-based model for energy spot prices. We present the model under the historical measure and illustrate a structure preserving change of measure, allowing to specify a risk-neutral dynamics. Then, we propose an effective estimation methodology based on particle filtering. Finally, we show how to perform exotic derivatives pricing both through exact simulation and characteristic function inversion techniques.

*Keywords:* Derivatives pricing, Energy Markets, Hawkes Processes, Jump Processes, Simulation

*JEL classification:* C13, C58, C63, G13, Q40

---

## 1. Introduction

Hawkes processes are a particular class of (self-exciting) point processes where the occurrence of any event increases the probability of future events. They became popular modeling tools since their introduction in Hawkes (1971a,b). Hawkes processes are ubiquitous across many fields of science. As a first application, they have been considered in earthquake modeling, where the aftershocks following a main event are described with an Hawkes process; a notable example is given by Ogata (1998). Moreover, self-exciting features proved to be useful to forecast infection diseases (Meyer and Held, 2014), crimes (Mohler *et al.*, 2011) and terrorist attacks (Porter and White, 2012). The popularity of Hawkes processes have also been established in other fields: Xu *et al.* (2014) propose applications in e-marketing, Fox *et al.* (2016) for modeling events and social networks

---

\*Correspondence

*Email addresses:* [riccardo.brignone@finance.uni-freiburg.de](mailto:riccardo.brignone@finance.uni-freiburg.de) (Riccardo Brignone),  
[luca.gonzato@univie.ac.at](mailto:luca.gonzato@univie.ac.at) (Luca Gonzato), [carlo.sgarra@polimi.it](mailto:carlo.sgarra@polimi.it) (Carlo Sgarra)

behaviour, Vinogradov *et al.* (2019) in mechanical twinning, Johnson (1996) finds applications in neuron activity, Nakagawa *et al.* (2019) for modeling population dynamics. Nevertheless, an impressive spread of applications of Hawkes processes has recently taken place in economics and finance. We refer to Hawkes (2018, 2022) for detailed surveys. A first strand of financial applications of Hawkes processes is related to high-frequency trading and limit order book. Some notable examples are given by Chavez-Demoulin *et al.* (2005), Chavez-Demoulin and McGill (2012), Bacry *et al.* (2013) and Cartea *et al.* (2014). For insurance markets, Hainaut (2016, 2017) proposes Hawkes processes to analyse the impact of volatility clustering on the evaluation of equity indexed annuities and the contagion between insurance and financial markets. Further applications in insurance are given by Stabile and Torrisi (2010), Dassios and Zhao (2012), Zhu (2013) and Jang and Dassios (2013). From the point of view of credit risk, we mention the paper by Errais *et al.* (2010), where the authors proposed to model credit default events in a portfolio of securities as correlated point processes, where the events dynamics are described by a mutually exciting Hawkes process. A more detailed review of the literature devoted to Hawkes-based models for commodities will be provided in Section 3. Hawkes processes have been thoroughly investigated from the mathematical viewpoint. In general, they are not affine and not even Markov processes, but in one case they are both, and this is the case in which the memory kernel is of exponential type (the definition will be provided in Section 4). This case is extremely relevant and attractive from the modeling viewpoint since for affine processes it is possible to compute explicitly the characteristic function, allowing for the usage of powerful transform-based pricing methods for derivatives valuation. Moreover, it also allows for more efficient simulation, as we illustrate later in this chapter. As a result, extensive literature is available on derivative pricing. We mention here the contributions by Hainaut and Moraux (2018) on the pricing and hedging of options on stocks, by Brignone and Sgarra (2020) and Brignone *et al.* (2023b) for Asian options, by Kokholm (2016) for exchange options.

In addition, we want to recall that Hawkes processes are not the only class of processes exhibiting self-exciting features: continuous time branching processes with immigration (CBPI) is another relevant class of processes sharing with Hawkes processes this remarkable property. CBPI processes have been used to describe populations dynamics by Pardoux (2016). In finance, Jiao *et al.* (2017) use CBPI to model short rates and illustrate their relationship with Hawkes processes.

Another motivation for the popularity of Hawkes processes in financial markets modeling is their strict relationship with rough volatility models, i.e. models including in the driver one or more fractional Wiener processes with Hurst index  $H < 1/2$ . Jaisson and Rosenbaum (2015) proved that rough volatility models can be obtained as a proper scaled limit of a sequence of Hawkes processes with long memory kernels, and this property turns out to be relevant for computational purposes (El Euch and Rosenbaum, 2019).

The purpose of this work is fourfold: *i*) we provide some more evidence of self-exciting

effects in the spot price of energy commodities; *ii*) we provide a brief literature review on the usage of Hawkes processes in energy finance; *iii*) we provide a literature review on the exact simulation of Hawkes process with exponential kernel, and compare the performances of the various methodologies available in the literature; *iv*) we illustrate a simple but realistic and practically relevant approach to model energy spot prices based on a Hawkes process with exponential kernel, which can be applied for derivatives pricing purposes. Indeed, we first illustrate a change of measure technique to move from the historical to the risk-neutral measure. Second, we propose an effective estimation methodology based on recent econometric literature. Third, we show how to price exactly some derivative instruments such as geometric Asian options under the suggested model. Fourth, we develop an exact simulation scheme for such models and show how to apply it for pricing path-dependent derivatives instruments, such as Bermudan-Asian options, which are popular in commodity markets.

The plan of this chapter is the following: in Section 2 we discuss empirical evidence of jump clustering features in energy commodity price dynamics, in Section 3 we present a review of the literature available on Hawkes models in energy markets, in Section 4 we define Hawkes processes and illustrate their basic properties, in Section 5 we review existing simulation schemes adopted for Hawkes processes and compare their performances. In Section 6 we present a Hawkes-type model for a generic asset, and we discuss the relation between the historical and the risk-neutral dynamics. In Section 7 we propose an estimation methodology for the proposed model and in Section 8 we provide a pricing method for a class of exotic derivatives, the Asian options (with and without early exercise features). Finally, in the last section, we resume the main results and discuss possible developments.

## 2. Empirical evidence of jump clusters

Jump clustering in financial markets is the phenomenon for which, whenever a jump in the price process occurs, there is an increase in the probability of observing a new jump over the next period. This statistical feature is widely observed in financial markets, such as equity (see Fulop *et al.*, 2015; Fulop and Li, 2019, among others), and commodity (see Section 3 for a detailed review). For illustrative purposes, we provide some evidence of jump clusters for two energy commodities. We consider the time series of spot returns of West Texas Intermediate (WTI) and NY Harbor Ultra Low Sulphur Heating Oil (NYULSD) from 30-Jun-2006 to 15-Dec-2022 (4294 days, source: Datastream<sup>®</sup>). These are reported in the top panel of Figure 1. Next, we need to detect jumps in the price processes. To this aim, we suggest applying the iterated re-weighted least squares technique developed in Callegaro *et al.* (2017). This simple algorithm is used to extract jumps from a time series of spot prices and can be summarized as follows. Step 1: the mean and standard deviation of the daily simple returns are computed. Returns which are greater (respectively, smaller) than the mean plus (minus) three standard deviations are consid-

ered as jumps. Step 2: a new time series of simple returns is created by replacing the realizations corresponding to a jump with the sample mean. Steps 1 and 2 are repeated until the ratio between standard deviations of two consecutive iterations is greater than 0.99. By this way, the separation between the jumps and the Gaussian oscillations is done (see Callegaro *et al.*, 2017, Bernis *et al.*, 2021 and Brignone *et al.*, 2023b for more details). The algorithm detects 170 jumps in WTI and 147 jumps in NYULSD. Note that, based on this result, a model for energy commodity prices which omit jumps is likely misspecified. The gray lines in the bottom panel of Figure 1 represent the jumps detected via the aforementioned algorithm. By an inspection of this figure, it is possible to see that jumps appear in clusters. We have periods with high jump activity (e.g. during the global financial crisis in 2008/2009), followed by calm periods where we do not observe any jump. This means that it is unlikely that jumps in the energy commodity spot prices exhibit a constant arrival rate. Following Bernis *et al.* (2021) and Brignone *et al.* (2023b), we assess this fact with a simple statistical test. If the arrival rate is constant, then the jump process is a homogeneous Poisson process and the distribution of the interarrival times is exponential with mean  $1/\lambda$ , where  $\lambda$  is the average arrival rate computed as the ratio between the total number of jumps and the number of observations in the sample (i.e.  $\lambda = 0.0396$  for WTI spot price and  $\lambda = 0.0342$  for NYULSD). We then perform a two sample Kolmogoroff-Smirnov test where the null hypothesis is that the interarrival times are exponentially distributed with mean  $1/\lambda$ . We strongly reject the null hypothesis, with  $p$ -values around  $5.53\text{E-}06$  for WTI spot price and  $2.05\text{E-}04$  for NYULSD. This analysis justifies the usage of stochastic processes for properly modeling the jumps intensity. In this work, we study the usage of Hawkes processes for this purpose.

### 3. Literature review: Hawkes processes in energy markets

A large amount of literature is devoted to model energy prices dynamics by Hawkes processes and in this section we provide a review on the main contributions available on the subject, and we apologize if the list presented will not be exhaustive.

Filimonov *et al.* (2014) investigate endogeneity of jumps in several asset classes, including commodities, and, in particular, crude oil and natural gas. By estimating from real data the branching ratio they provide evidence of endogeneity of jumps in prices, meaning that a large amount of jumps is self-exciting, or, put in a different way, they appear in clusters. Several authors aim at modeling intraday electricity prices by Hawkes processes: Favetto (2008) investigates the European intraday market while Graf von Luckner and Kiesel (2021) investigate the German intraday market, providing evidence of jumps clustering. Blasberg *et al.* (2019) consider the question whether there is a temporal structure prevalent in the parameters of Hawkes processes estimated for adjacent delivery hours. They model a daily seasonality pattern found in the data and provide a suitable economic interpretation. In order to model the other components of the time series, they propose simple (vector) autoregressive models. They conduct a forecasting



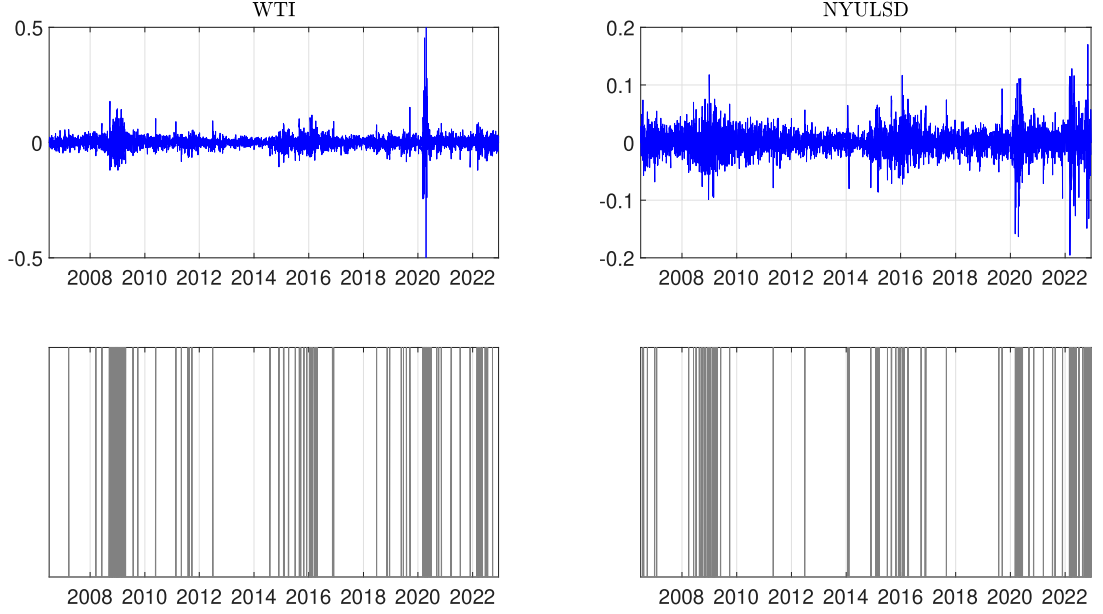


Figure 1: Jump clusters in energy markets. Top subplots contain the time series of the returns of WTI and NYULSD; bottom subplots contain the time series of the jumps (indicated as vertical gray lines) detected using the methodology developed in Callegaro *et al.* (2017).

study and find superior performances of their proposed model against various benchmarks.

Callegaro *et al.* (2022) propose and investigate two model classes for forward power price dynamics, based on continuous branching processes with immigration, and on Hawkes processes with exponential kernel, respectively. Both the proposed models exhibit jumps clustering features. They adopt a Heath–Jarrow–Morton approach to capture the whole forward curve evolution. By examining daily data in the French power market, and by performing a goodness-of-fit test, they find evidence of adequacy of Hawkes processes in describing the forward price dynamics.

Clements *et al.* (2015) are the first to develop a multivariate self-exciting point process model for dealing with price spikes across connected regions in the Australian National Electricity Market. They examine the importance of the physical infrastructure connecting the regions on the transmission of spikes, and find that: *i*) spikes are transmitted between the regions; *ii*) the size of spikes is influenced by the available transmission capacity.

Giordano and Morale (2021) propose a new model based on a Hawkes process to model the Italian power market. They find evidence of the presence of jumps in the Italian electricity market, many of which appear clustered over short time periods. The two-factor model they propose is driven by both Hawkes and fractional Brownian processes. They perform a calibration procedure, discussing the seasonality, spikes and estimate the Hurst coefficient. After calibrating and validating the model, they examine its forecasting performance via a class of adequate evaluation metrics.

Eyjolfsson and Tjøstheim (2018) suggest a model for spot prices in power markets based on Hawkes processes. They introduce a class of jump processes with self-exciting jump intensity. In particular, the paper focuses on two main aspects: *i*) stochastic analysis of the proposed processes; *ii*) empirical application using daily quotes of UK electricity prices.

Jiao *et al.* (2019) propose a model for spot power prices based on a CBPI process and, after introducing a suitable relation between the historical and the risk-neutral spot price dynamics, they compute the forward and futures prices and provide a sound description of the sign change in the term structure of risk premium often observed in power markets.

Gonzato and Sgarra (2021) present a Hawkes type model for crude oil prices, including a stochastic dynamics for volatility. Their model specification is completed by a stochastic convenience yield. For model estimation, they apply the two-stage Sequential Monte Carlo (SMC) sampler (Fulop and Li, 2019) to both spot and futures quotations. From the estimation results, they find evidence of self-excitation in the oil market, which leads to an improved fit and a better out of sample forecasting performance with respect to jump-diffusion models with constant intensity.

Brignone *et al.* (2023b) propose a mean-reverting model for commodity prices with Hawkes-type jumps, stochastic volatility and a diffusion-type dynamics for the convenience yield. This model is calibrated on real derivatives data, including energy related products, and is used to price exotic derivatives instruments. The model we are going to propose as a working example in Section 6 is somehow similar to the model proposed in Brignone *et al.* (2023b), where a different estimation procedure has been used. We point out that the simulation scheme used in the last mentioned paper could be successfully applied to price options with early exercise features, and this is the subject of ongoing work.

## 4. Hawkes processes

We give here the definition of Hawkes process (with exponential kernel) and outline some basic properties.

### 4.1. Definition based on the conditional intensity

A Hawkes process is a special kind of point process whose conditional intensity depends on the history of the events. First, let us define the conditional intensity function and the univariate Hawkes process.

**Definition 1.** *Let  $N$  be a point process and let  $\mathcal{F}_t^N$  be the natural filtration generated by  $N$  itself. Then, the left continuous process defined by:*

$$\lambda_t := \lambda(t \mid \mathcal{F}_{t-}^N) = \lim_{h \rightarrow 0^+} \frac{\mathbb{P}[N_{t+h} - N_t > 0 \mid \mathcal{F}_{t-}^N]}{h} \quad (1)$$

*is called the conditional intensity function of the point process.*

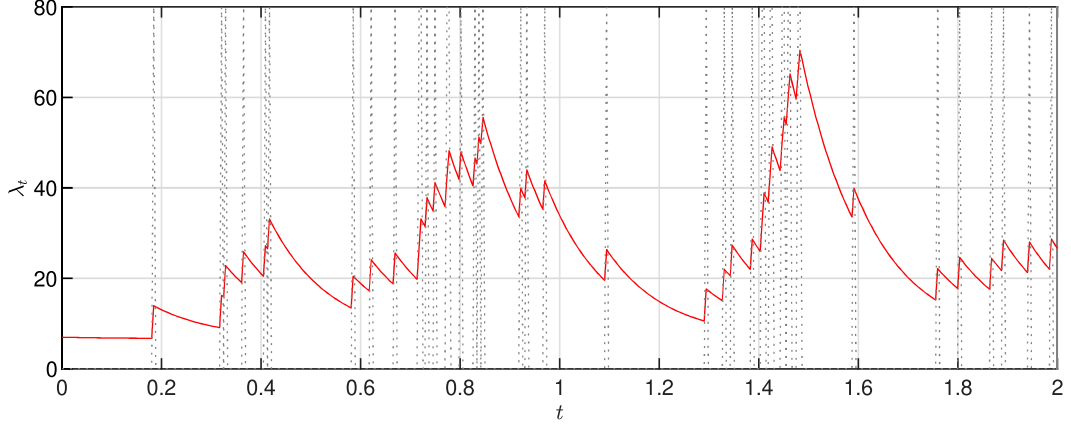


Figure 2: Simulated sample path of the intensity of a Hawkes process.

**Definition 2.** The univariate Hawkes process  $N$  with conditional intensity  $\lambda_t$ , can be defined for all  $t > 0$  and  $h \rightarrow 0^+$  as:

$$\begin{cases} \mathbb{P}[N_{t+h} - N_t = 1 \mid \mathcal{F}_{t-}^N] &= \lambda_t h + o(h) \\ \mathbb{P}[N_{t+h} - N_t > 1 \mid \mathcal{F}_{t-}^N] &= o(h) \\ \mathbb{P}[N_{t+h} - N_t = 0 \mid \mathcal{F}_{t-}^N] &= 1 - \lambda_t h + o(h), \end{cases} \quad (2)$$

where  $o(h)$  denotes infinitesimals of higher order with respect to  $h$ , or, more qualitatively speaking, terms going to 0 faster than  $h$ .

Furthermore, the dynamics of conditional intensity of a Hawkes process with exponentially decay function can be represented by the following stochastic differential equation (SDE):

$$d\lambda_t = k(\theta - \lambda_t)dt + \beta dN_t, \quad (3)$$

where,  $k$  is the *constant rate of decay*,  $\theta$  is the *background intensity*,  $\beta$  is the *magnitude of self-excited jump* and  $N$  is a univariate Hawkes process. For sake of illustration, we report in Figure 2 a simulated trajectory of a Hawkes process. The vertical gray dashed lines represent jump occurrences. Whenever the jump occurs, the intensity increases. As a result, we have a higher probability of observing a new jump. When there are no jumps, the intensity reverts to its long run mean.

Given the initial datum  $\lambda_0$  the solution of (3), can be written as follows:

$$\lambda_t = \theta + (\lambda_0 - \theta)e^{-kt} + \sum_{k=1}^{N_t} \beta e^{-k(T_k - t)}, \quad t \geq 0. \quad (4)$$

Then, (4) can be written again in the following way:

$$\lambda_t = \theta + (\lambda_0 - \theta)e^{-kt} + \beta \int_0^t e^{-k(t-s)} dN_s, \quad (5)$$

which provides the meaning of the expression "exponential memory kernel" sometimes used to denote this class of processes. This definition can be extended to more general memory kernels  $K(s)$ , when  $\lambda_t$  is given by:

$$\lambda_t = \lambda_0 + \beta \int_0^t K(t-s) dN_s.$$

When  $K(s) = S^\gamma$ , the Hawkes process with conditional intensity  $\lambda_t$  is called a Hawkes process with power (memory) kernel. The Markov property of the couple  $(\lambda_t, N_t)$  in this case is lost.

**Definition 3.** *A D-variate Hawkes process is a D-variate counting process  $N$  with intensity vector given by:*

$$\lambda_t^i = \lambda_0^i + \sum_{j=1}^D \int_0^t K^{i,j}(t-s) dN_s^j,$$

where the non-negative and causal (i.e.  $K^{i,j}(s) = 0, s < 0$ )  $L^1$  functions (see Bacry et al., 2015, Section 2)  $K^{i,j}(s)$  denote the matrix kernel.

D-variate Hawkes processes provide a description of mutual contagion or mutual-exciting behavior effects among different assets. A notable example is given by Ait-Sahalia et al. (2015), where the authors propose a mutually-exciting jump-diffusion process to describe the dynamics of market indices in six different countries. Ait-Sahalia and Hurd (2016) investigate the optimal investment problem for a financial market with mutual contagion described by a D-variate Hawkes process, while Kokholm (2016) presents an evaluation algorithm for spread options based on the explicit computation of the characteristic function of a bi-variate Hawkes process with exponential kernel. In the present chapter, we shall focus on the one-dimensional case.

#### 4.2. Moments of $\lambda_T$ and $N_T$

We provide, next, the expressions for the conditional expectation and variance of  $\lambda_t$ , given as in (3), and for the conditional expected number of jumps:

$$\begin{aligned} \mathbb{E}[\lambda_T] &= \frac{\theta k}{k_1} + \left( \lambda_0 - \frac{\theta k}{k_1} \right) e^{-k_1 T}, \\ \text{Var}[\lambda_T] &= \frac{\beta^2}{k_1} \left[ \left( \frac{\theta k}{2k_1} - \lambda_0 \right) e^{-2k_1 T} + \left( \lambda_0 - \frac{\theta k}{k_1} \right) e^{-k_1 T} + \frac{\theta k}{2k_1} \right], \\ \mathbb{E}[N_T] &= \frac{\theta k}{k_1} T + \frac{1}{k_1} \left( \lambda_0 - \frac{\theta k}{k_1} \right) (1 - e^{-k_1 T}), \end{aligned} \quad (6)$$

where,  $k_1 = k - \beta > 0$ . The analytical expression in (6) will be used, following Dassios and Zhao (2013), to evaluate the correctness and accuracy of the simulations in Section 5.

## 5. Simulation of Hawkes processes with exponential kernel

In this section, we illustrate several approaches that have been proposed in the literature to simulate a Hawkes process, which do not require any numerical method for their implementation.<sup>1</sup> Together with the description of the algorithms, we also provide pseudocodes. These are presented in the easiest possible form to allow a straightforward implementation. Matlab<sup>®</sup> codes are available at <https://www.mathworks.com/matlabcentral/fileexchange/74738-hawkes-simulation>. Note that these codes can be slightly different from pseudocodes provided below, since they are more efficient.

### 5.1. Euler scheme

The Euler scheme is used to simulate the exact solution of a discrete time process approximating the system dynamics. In our framework, we can discretize (3) by approximating the jump times  $N_t$  with a Bernoulli random variable, as in Fulop and Li (2019). Therefore,

$$\lambda_{t_i} = \lambda_{t_{i-1}} + k(\theta - \lambda_{t_{i-1}})\Delta t + \beta\Delta N_{t_i}, \quad (7)$$

where,  $\Delta N_{t_i} = N_{t_i} - N_{t_{i-1}} \sim \text{Ber}(\lambda_{t_{i-1}}\Delta t)$ . Despite affected by time discretization bias, this method is widely used in practice due to its simplicity. Moreover, it is expected to be fast in general, with a lower dependence on the expected number of jumps than competing methods. Pseudocode is given in Algorithm 1.

---

#### Algorithm 1 Euler scheme

---

**Input:**  $\lambda_0, k, \theta, \beta, T, n$

**Output:**  $\{\lambda_{t_j}\}_{j=1}^n, \{N_{t_j}\}_{j=1}^n$

---

- 1: Set  $\Delta t = \frac{T}{n}$  and  $N_0 = 0$
  - 2: **for**  $j = 1 : n$  **do**
  - 3:   Set  $\Delta N_{t_j} = 0$  and draw  $u \sim \mathcal{U}(0,1)$  % draw from uniform
  - 4:   **if**  $u < \lambda_{t_{j-1}}\Delta t$  **then**
  - 5:      $\Delta N_{t_j} = 1, N_{t_j} = N_{t_{j-1}} + 1$
  - 6:   **end**
  - 7:   Set  $\lambda_{t_j} = \lambda_{t_{j-1}} + k(\theta - \lambda_{t_{j-1}})\Delta t + \beta\Delta N(t_j)$
  - 8: **end**
- 

### 5.2. Ogata's modified thinning algorithm and a variant

The simulation algorithm for Hawkes processes proposed by Ogata (1981) is the most popular in the literature. It is based on the thinning simulation method for non-homogeneous Poisson processes proposed in Lewis and Shedler (1969). The idea is to

---

<sup>1</sup>In other words, we exclude from our review those methods which are expectantly slow since based on repeated implementation of computationally intensive numerical techniques such as, for example, root finding algorithms (as in Ozaki, 1979), or numerical integration (as in Møller and Rasmussen, 2005).

simulate a homogeneous Poisson process and then remove excess points, such that the remaining points satisfy the conditional intensity  $\lambda_t$ . This algorithm requires the conditional intensity to be upper bounded, i.e.  $\exists \bar{\lambda} < \infty$  such that  $\lambda_t \leq \bar{\lambda}, \forall t$ . A generalization introduced by Ogata (1981) (typically called Ogata's modified thinning) is outlined in Algorithm 2. This generalization requires only the local boundedness of conditional intensity. Indeed, if  $\lambda_t$  is a non-increasing function (i.e.  $k > 0$  in (3)) in the interval between two adjacent occurrences, we have that  $\lambda_t \leq \lambda_{t_i^+}$  for  $t \in (t_i, t_{i+1})$ , where  $t_i^+$  is the time just after  $t_i$ . Therefore, a local bound  $\bar{\lambda}_t$  could be set equal to  $\lambda_{t_i^+}$  in the interval  $(t_i, t_{i+1})$  and it has to be updated after each occurrence. Pseudocode is given in Algorithm 2.

---

**Algorithm 2** Ogata's modified thinning algorithm

---

**Input:**  $\lambda_0, \theta, k, \beta, T$

**Output:**  $N_T, \{T_j\}_{j=1}^{N_T}, \{\lambda_{T_j}\}_{j=1}^{N_T}$

---

```

1: Set  $N_0 = 0, \epsilon = 10^{-6}, j = 0, t = 0, T_0 = 0$ 
2: while  $t < T$  do
3:   Compute  $M = \lambda_{t+\epsilon}$  according to (5)
4:    $u_1 \sim \mathcal{U}(0, 1), u_2 \sim \mathcal{U}(0, 1)$  % draw from two independent uniforms
5:    $R = -\frac{\ln(u_1)}{M}$  % transform in an exponential with mean  $1/M$ 
6:   Compute  $H = \lambda_{t+R}/M$  according to (5)
7:   if  $u_2 > H$  then
8:      $t = t + R$ 
9:   else
10:     $j = j + 1, t = t + R, T_j = t, N_{T_j} = N_{T_{j-1}} + 1$ 
11:  end
12: Compute  $\{\lambda_{T_j}\}_{j=1}^{N_T}$  using formula (5)
```

---

Afterwards, Daley and Vere-Jones (2002) modified Ogata's algorithm by setting  $\bar{\lambda}_t = \lambda_t$ . In this way we are not interested if  $t$  is a point of the process or not; whereas we add a function of time interval of length  $L_t = \bar{k}\lambda_{t+}$ , for arbitrary  $\bar{k}$  (the authors suggest to set  $\bar{k} = 0.5$ ). Then,  $\bar{\lambda}_t$  is updated if a new point of the process occurs or if the time frame  $L_t$  has elapsed. In Algorithm 3 we describe how to implement Ogata's thinning algorithm as in Daley and Vere-Jones (2002).

### 5.3. Dassios and Zhao's exact simulation

Dassios and Zhao (2013) proposed an efficient sampling algorithm for Hawkes process that does not rely on the accept-reject method. Moreover, this algorithm can generate jump times with either stationary or non-stationary intensity. The starting point is the cluster-based representation of a Hawkes process (see Dassios and Zhao, 2013 for more details). The authors decompose the interarrival times into two independent random variables: the first random variable  $S_1$  represents the interarrival time of the next event, if it is coming from the background intensity  $\theta$ . The second random variable  $S_2$  represents the interarrival time of the next event if it comes from either the exponential immigrant kernel  $(\lambda_0 - \theta)e^{-kt}$  or the Hawkes self-exciting kernel from each of the past events  $\sum_{j: t_j < t} e^{-k(t-t_j)}$ . Then, the sampled interarrival time is the minimum of these two

---

**Algorithm 3** Ogata's modified thinning algorithm as in Daley and Vere-Jones (2002)

---

**Input:**  $\lambda_0, \theta, k, \beta, T$ **Output:**  $N_T, \{T_j\}_{j=1}^{N_T}, \{\lambda_{T_j}\}_{j=1}^{N_T}$ 

```
1: Set  $\bar{k} = 0.5, N_0 = 0, \epsilon = 10^{-6}, j = 0, t = 0, T_0 = 0$ 
2: while  $t < T$  do
3:   Compute  $M = \lambda_t$  according to (5) and  $L = \bar{k}\lambda_{t+\epsilon}$ 
4:    $u_1 \sim \mathcal{U}(0,1)$  % draw from uniform
5:    $R = -\frac{\ln(u_1)}{M}$  % transform in an exponential with mean  $1/M$ 
6:   if  $R > L$  then
7:      $t = t + L$ 
8:   else
9:      $u_2 \sim \mathcal{U}(0,1)$  % draw from uniform
10:    Compute  $H = \lambda_{t+R}/M$  according to (5)
11:    if  $u_2 > H$  then
12:       $t = t + R$ 
13:    else
14:       $j = j + 1, t = t + R, T_j = t, N_{t_j} = N_{t_{j-1}} + 1$ 
15:    end
16:  end
17: end
18: Compute  $\{\lambda_{T_j}\}_{j=1}^{N_T}$  using formula (5)
```

---

cases. Moreover, if the second arrival time is not finite, the next event will be taken from the constant rate. The simulation scheme is outlined in Algorithm 4.

---

**Algorithm 4** Dassios and Zhao exact simulation

---

**Input:**  $\lambda_0, \theta, k, \beta, T$ **Output:**  $N_T, \{T_j\}_{j=1}^{N_T}, \{\lambda_{T_j}\}_{j=1}^{N_T}$ 

```
1: Set  $N_0 = 0, j = 0, t = 0, T_0 = 0$ 
2: while  $t < T$  do
3:    $u_1 \sim \mathcal{U}(0,1)$  and  $u_2 \sim \mathcal{U}(0,1)$  % draw from two independent uniforms
4:    $D = 1 + k \frac{\ln u_1}{\lambda_{T_j} - \theta}$ 
5:    $S_1 = -\frac{\ln D}{k}$  and  $S_2 = -\frac{\ln u_2}{\theta}$ 
6:   if  $D > 0$  then
7:      $S = \min(S_1, S_2)$ 
8:   else
9:      $S = S_2$ 
10:  end
11:   $T_{j+1} = T_j + S$  and  $t = T_{j+1}$ 
12:   $\lambda_{j+1} = (\lambda_j - \theta)e^{-k(T_{j+1}-T_j)} + \theta + \beta$ 
13:   $N_{t_{j+1}} = N_{t_j} + 1$ 
14:   $j = j + 1$ 
15: end
```

---

#### 5.4. Comparison of exact methodologies

We start by comparing Algorithms 2, 3 and 4. When evaluating the efficiency of a methodology, three considerations are important: computing time, bias, and variance.

All these methods are unbiased and present the same variance.<sup>2</sup> Hence, we just need to evaluate the running time. To this aim, we implement the following experiment. We generate 100 parameters combinations respecting the following conditions:  $k \in [0.5, 35]$ ,  $\theta \in [0.1, 3]$ ,  $\beta \in [1, 30]$ ,  $\lambda_0 \in [0.1, 5]$ ,  $T \in \{1, 3, 5, 7, 10\}$ ,  $k > \beta$ ,  $\lambda_0 > \theta$ . With these choices we aim to incorporate standard parameter settings used in the literature, in particular Kokholm (2016), Boswijk *et al.* (2016), Gonzato and Sgarra (2021), Fulop and Li (2019) which calibrated the Hawkes model on real financial data. For each parameter set we run the three algorithms with  $10^6$  simulations and register the running time. We choose such a high number of simulations in order to minimize the risk that the comparison between the different methodologies is influenced by the effective number of jumps over all the simulations. Indeed, with only few simulations, we could observe that a method is faster than another one only because the (random) number of jumps is much smaller and not because the method is effectively faster. In Figure 3 we report the running time in seconds and the expected number of jumps for each model parameter set. Results show that Algorithm 4 is the best performing overall. For the same accuracy it results slightly faster than Algorithm 2. This difference is anyway small and basically negligible for small  $E[N_T]$ . The superior performance is due to the fact that random sampling is necessary only when the jump effectively occurs, while in thinning algorithms those steps run also when the jump is discarded, resulting in a higher running time. Moreover, in Algorithm 2, formula (5) must be implemented to compute the upper bound for the intensity  $M$ , the acceptance probability  $H$  and to update the jump intensity, while, in Algorithm 4, this is evaluated only to update the jump intensity (i.e. only when the jump effectively occurs). We compare now Algorithms 2 and 3. Let's start by an inspection of Figure 3. We note the strange results for the running time of the method of Daley and Vere-Jones (2002) for small values of the expected number of jumps. The issue comes exactly from the adjustment they propose. Indeed, a small value for (6) is usually due to small values of the intensity parameters. Therefore, in Algorithm 3, we get stuck at the first `if` statement, since the variable  $L$  will be always smaller than  $R$ , causing a very small update of  $\mathbf{t}=\mathbf{t}+\mathbf{L}$ .<sup>3</sup> To the best of our knowledge, this unsatisfactory behavior of the algorithm was not previously noted in the literature. However, the proposed modification is expected to give a reasonable compromise between setting the bound  $M$  too high, and so generating excessive trial points, and setting it too low, thus requiring too many iterations of steps 5 and 6 in Algorithm 3. In other words, the idea is to avoid running the accept-reject step whenever the possibility of observing a jump is very low, saving the CPU time needed for random sampling. This allows in practice to reduce the number of discarded points, but

---

<sup>2</sup>In unreported tests (which can be made available upon request), we compared the bias and the variance in the computation of (6) through Algorithms 2, 3 and 4, and found that differences in the variance of the Monte Carlo estimators are immaterial.

<sup>3</sup>A partial solution to this problem is given by increasing the value of the parameter  $k$  in Algorithm 3, which controls the function  $L$ .



| Set | $k$   | $\theta$ | $\beta$ | $\lambda_0$ | $E[N_1]$ | Literature                   |
|-----|-------|----------|---------|-------------|----------|------------------------------|
| A   | 20.8  | 1.13     | 13.5    | 1.3         | 2.9569   | Kokholm (2016)               |
| B   | 18.16 | 0.32     | 16.62   | 0.5         | 2.1036   | Boswijk <i>et al.</i> (2016) |
| C   | 30.05 | 1.39     | 23.46   | 1.5         | 5.6051   | Gonzato and Sgarra (2021)    |
| D   | 2.56  | 2.14     | 2       | 2.3         | 4.0532   | Fulop and Li (2019)          |

Table 1: Parameter sets taken from financial literature.

our numerical results show that this benefit is not commensurate with the extra labour involved. Indeed, with respect to Algorithm 2 some extra operations are necessary, i.e. an `if` statement and the computation of  $L$ . As a result, Algorithm 3 results slower than Algorithm 2.

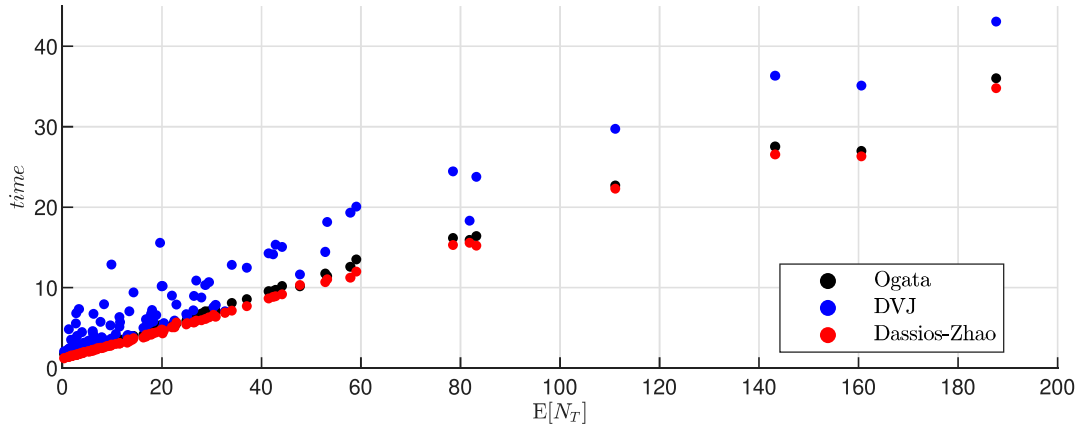


Figure 3: Time in seconds of each simulation scheme with respect to the expected number of jumps.

### 5.5. Comparison between Algorithm 4 and the Euler scheme

We compare, next, the performances of Algorithms 1 and 4. To this aim, we need first to estimate the bias of the Euler scheme. We select some parameter sets taken from financial literature (see Table 1). For each parameter set, we compute (6) analytically and through Monte Carlo simulation using Algorithm 1 ( $10^9$  simulations). The absolute difference between the true value and the Monte Carlo estimate is the bias.

Then, we run Algorithms 1 and 4 for different number of simulations  $\mathcal{M}$  and compute the Root Mean Squared Error (RMSE) and register the running time in seconds.<sup>4</sup> For the implementation of the Euler scheme we must also specify the number of time discretization steps (denoted with  $n$  in Algorithm 1). We follow the suggestion of Duffie and Glynn (1995) to select  $n = \sqrt{\mathcal{M}}$ . The numerical results are reported in Table 2. Few comments are in order. The Euler scheme is, as expected, less accurate with a non-negligible average error in computing the expected number of jumps. When we use less than  $\mathcal{M} = 4 \times$

<sup>4</sup>See e.g. Kyriakou *et al.* (2023) and the references therein for more details on the design of this experiment and definition of RMSE.

| Set A             |      |        |        |        |             |       | Set B       |        |        |        |             |       |
|-------------------|------|--------|--------|--------|-------------|-------|-------------|--------|--------|--------|-------------|-------|
| Algorithm 1       |      |        |        |        | Algorithm 4 |       | Algorithm 1 |        |        |        | Algorithm 4 |       |
| $\mathcal{M}$     | $n$  | bias   | RMSE   | time   | RMSE        | time  | $n$         | bias   | RMSE   | time   | RMSE        | time  |
| $4 \times 10^4$   | 200  | 0.0157 | 0.0268 | 0.36   | 0.0223      | 0.48  | 200         | 0.0178 | 0.0425 | 0.25   | 0.0441      | 0.44  |
| $16 \times 10^4$  | 400  | 0.0087 | 0.0140 | 2.27   | 0.0112      | 1.59  | 400         | 0.0083 | 0.0223 | 2.29   | 0.0223      | 1.54  |
| $64 \times 10^4$  | 800  | 0.0047 | 0.0073 | 19.59  | 0.0056      | 6.24  | 800         | 0.0059 | 0.0120 | 19.95  | 0.0109      | 6.19  |
| $256 \times 10^4$ | 1600 | 0.0013 | 0.0031 | 204.11 | 0.0028      | 24.87 | 1600        | 0.0018 | 0.0057 | 206.86 | 0.0054      | 23.17 |

| Set C             |      |        |        |        |             |       | Set D       |        |        |        |             |       |
|-------------------|------|--------|--------|--------|-------------|-------|-------------|--------|--------|--------|-------------|-------|
| Algorithm 1       |      |        |        |        | Algorithm 4 |       | Algorithm 1 |        |        |        | Algorithm 4 |       |
| $\mathcal{M}$     | $n$  | bias   | RMSE   | time   | RMSE        | time  | $n$         | bias   | RMSE   | time   | RMSE        | time  |
| $4 \times 10^4$   | 200  | 0.0321 | 0.0538 | 0.27   | 0.0477      | 0.58  | 200         | 0.0332 | 0.0376 | 0.24   | 0.0183      | 0.51  |
| $16 \times 10^4$  | 400  | 0.0165 | 0.0281 | 2.32   | 0.0239      | 2.06  | 400         | 0.0172 | 0.0193 | 2.33   | 0.0090      | 1.78  |
| $64 \times 10^4$  | 800  | 0.0080 | 0.0141 | 20.43  | 0.0120      | 8.22  | 800         | 0.0083 | 0.0094 | 20.41  | 0.0045      | 7.09  |
| $256 \times 10^4$ | 1600 | 0.0037 | 0.0070 | 224.39 | 0.0060      | 31.75 | 1600        | 0.0043 | 0.0048 | 210.39 | 0.0023      | 27.24 |

Table 2: Speed–accuracy comparisons of Algorithms 1 and 4 for different parameter sets: the case of  $E[N_1]$ . All computing times are in seconds.

$10^4$  simulations, the Euler scheme is faster than Algorithm 4. Indeed, the most time consuming steps in Algorithm 4 (but also Algorithms 2 and 3) are: *i*) random sampling from the uniform distribution; *ii*) updating the vectors of jump times and intensity. Indeed, the total number of jumps is unknown at the beginning of each simulation. This fact precludes more efficient vectorizations and pre-allocations. In other words, it is not possible to sample all the needed random quantities in one time at the beginning of the simulation. Contrarily, the Euler scheme does not suffer of this problem as the number of uniforms to be sampled is  $n \times T$  and does not depend on the effective number of jumps. Hence, for small  $n = \sqrt{\mathcal{M}}$ , the Euler scheme is faster than the competitor. Despite that, in practical applications  $n$  is not a small number since the accuracy of this simulation scheme depends on the number of time discretization points, as it is clear from the evolution of the bias in Table 2. For this reason the Euler scheme is slower than the other methods for higher  $n$  (or, equivalently in our experiment, higher  $\mathcal{M}$ ). Therefore, a good environment for its implementation is when the expected number of jumps is very high, but the simulation horizon is small (e.g.  $T < 1$ ). However, when high accuracy is required, Algorithm 4 is much faster and more accurate than the Euler scheme, with a higher convergence rate of the RMSE, as evident from Figure 4, where we report speed–accuracy portrayals on a log-log scale for the competing methods.

Based on these numerical results, we suggest to use Algorithm 4 to simulate Hawkes processes with an exponential kernel.

## 6. An asset pricing model with self-exciting jumps

In this section we suggest a simple model for spot returns where the jump intensity evolves according to a Hawkes process with exponential kernel. We first define the model under the historical measure, then we propose a structure preserving change of measure which allows the specification of the risk-neutral dynamics.

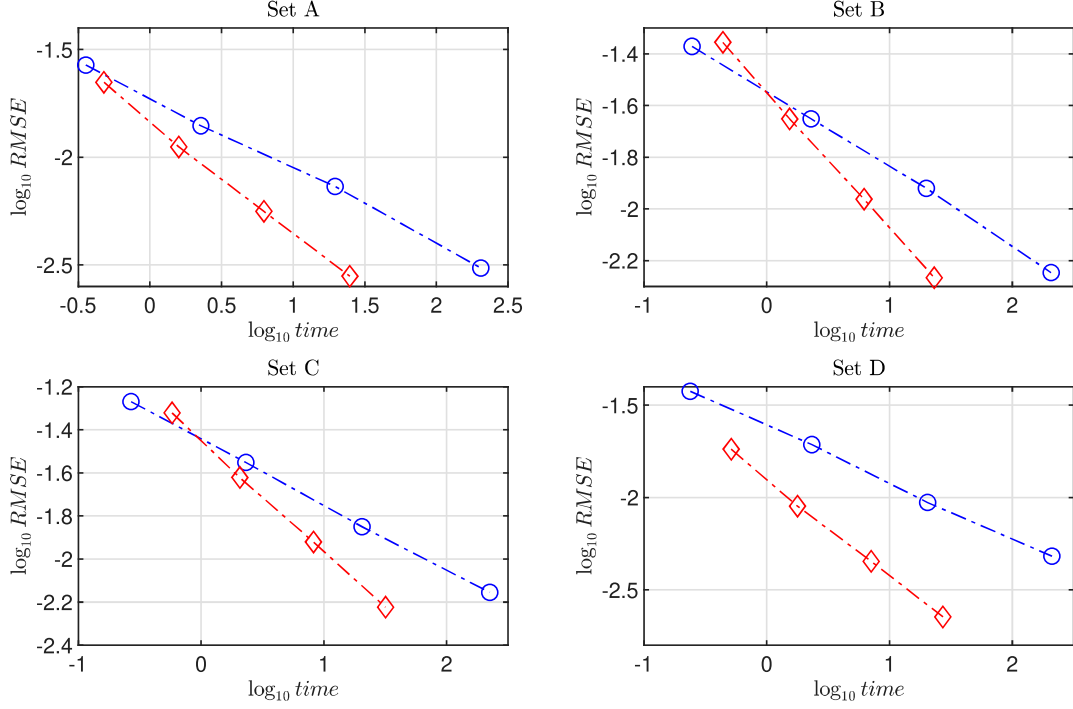


Figure 4: Speed–accuracy comparisons of Algorithm 1 (blue line with circles) and 4 (red line with diamonds) for different parameter sets: the case of  $E[N_1]$ . All computing times are in seconds.

### 6.1. Dynamics under the historical measure

Under the historical measure, asset price and jump intensity are given by:

$$dX_t = d\ln(S_t) = \left( \mu - \delta - \frac{\sigma^2}{2} - \lambda_t E[e^J - 1] \right) dt + \sigma dW_t + J_t dN_t \quad (8)$$

$$d\lambda_t = k(\theta - \lambda_t)dt + \beta dN_t \quad (9)$$

This model can be seen as an extension of Merton (1976).  $J_t$  is the jumps size, which is normally distributed with given mean  $\mu_J$  and standard deviation  $\sigma_J$ .  $\beta$  is constant and is responsible for the self-excitation property of  $\lambda_t$ .  $dW_t$  is a standard Brownian motion independent from the jump component  $JdN_t$ . Further,  $\delta$  represents a constant convenience yield. We further assume that the following non-explosion condition for the Hawkes process  $N_t$  in (9) holds:

$$k > \beta. \quad (10)$$

The following shows that, despite  $\sigma$  is assumed to be constant, the variance of simple returns still changes randomly as function of  $\lambda$ :

$$V_t dt = \text{Var} \left[ \frac{dS_t}{S_t} \right] = \sigma^2 dt + (\mu_J^2 + \sigma_J^2) \lambda_t dt.$$

## 6.2. Dynamics under the risk-neutral measure

In order to deal with derivatives pricing we need to introduce a risk-neutral measure. The purpose of the present subsection is to illustrate briefly the basic technical tools required in order to perform a structure-preserving probability measure change in a Hawkes jump-diffusion framework. To this end, we extend the results in Zhang *et al.* (2009) and Hainaut and Moraux (2018) by introducing an Esscher-type measure change. This is defined as follows. If we denote by  $L_t := JdN_t$  the jump term in (8), and by  $\psi^{\mathbb{P}}(z) := \mathbb{E}[e^{zJ}]$  the moment generating function of the jump size density, we can define the following family of exponential martingales:

$$M_t(\xi, \phi) := \exp \left[ \kappa_1(\xi)\lambda_t + \xi L_t + \kappa_2(\xi)t - \frac{1}{2} \int_0^t \phi^2(u)du - \int_0^t \phi(u)dW_u \right]$$

where  $\kappa_1(\xi), \kappa_2(\xi)$  (with  $\kappa_1(\xi), \kappa_2(\xi)$  functions of  $\xi$ ) denote the risk premium related to the jumps and  $\phi$  (a stochastic process adapted to the reference filtration  $\mathcal{F}_t$ ) denotes the risk premium related to the diffusion component of log-returns. We state the following result. We remark that in the present case, where the logreturn probability density is Gaussian,  $\psi^{\mathbb{P}}(z)$  exists for all values of  $z$ , while for different distribution it will exist only for some values of the parameters characterizing the probability density of the logreturn jump size.

**Proposition 1.** *If, for any  $\xi$ , there exist functions  $\kappa_1(\xi), \kappa_2(\xi)$  solutions of the following system of algebraic equations:*

$$\begin{cases} \kappa_1 k - [\exp(\kappa_1(\xi)\beta)\psi(\xi) - 1] = 0 \\ \kappa_2 + \kappa_1 k\theta = 0, \end{cases} \quad (11)$$

*then  $M_t(\xi, \phi)$  is a local martingale. If, moreover, the non-explosion condition (10) holds for  $N_t$  and the Novikov condition  $(\mathbb{E}[\exp \int_0^T \phi^2(u)du] < \infty)$  holds for  $\phi$ ,  $M_t$  is a true martingale.*

**Proof.** We extend to the present modelling framework the results in Zhang *et al.* (2009) (in our case the process  $J_x dN_t$  driving the log-returns dynamics is marked). Let's define  $m_t := \ln M_t$ ; then

$$dm_t = \kappa_1(\xi)k(\theta - \lambda_t)dt + \kappa_2(\xi)dt + \kappa_1(\xi)\beta dN_t + \xi JdN_t - \frac{1}{2}\phi^2(t)dt - \phi(t)dW_t$$

By applying Itô's lemma to  $M_t$  we obtain:

$$\begin{aligned} dM_t &= M_t dm_t + \frac{1}{2} M_t d[m, m]_t^c \\ &\quad + M_t \int_{-\infty}^{+\infty} [\exp(\kappa_1(\xi)\beta + \xi z) - 1 - (\kappa_1(\xi)\beta + \xi z)] F(dz) dN_t, \end{aligned}$$

where  $[m, m]_t^c$  denotes the quadratic variation of the continuous component of  $m_t$  and  $F(\cdot)$  denotes the random measure of the jumps size  $J$  and  $\lambda_t \nu(dz)dt$  denotes the predictable compensator of the process  $JdN_t$ . We can write the last equation in the following form:

$$\begin{aligned} dM_t = & M_t \left[ \kappa_1(\xi)k(\theta - \lambda_t)dt + \kappa_2(\xi)dt + \left( \kappa_1(\xi)\beta + \xi \int_{-\infty}^{+\infty} zF(z)dz \right) dN_t \right] \\ & + M_t \left[ -\frac{1}{2}\phi^2(t)dt - \phi(t)dW_t \right] + \frac{1}{2}M_t d[m, m]_t^c \\ & + M_t \int_{-\infty}^{+\infty} [\exp(\kappa_1(\xi)\beta + \xi z) - 1 - (\kappa_1(\xi)\beta + \xi z)]F(dz)dN_t, \end{aligned}$$

which, after the introduction of the compensator of the jump process, can be written as

$$\begin{aligned} dM_t = & M_t \kappa_1(\xi)k\theta dt + M_t \kappa_2(\xi)dt - M_t \lambda_t \kappa_1(\xi)k \\ & - M_t \phi dW_t - M_t \lambda_t \int_{-\infty}^{+\infty} [\exp(\kappa_1(\xi)\beta + \xi z) - 1] \nu(dz)dt + \\ & + M_t \int_{-\infty}^{+\infty} [\exp(\kappa_1(\xi)\beta + \xi z) - 1] [F(dz)dN_t - \lambda_t \nu(dz)dt], \end{aligned}$$

since  $\frac{1}{2}M_t d[m, m]_t^c = \frac{1}{2}\phi^2(t)M_t dt$ . The integral with respect to the compensated jump measure  $F(dz)dN_t - \lambda_t \nu(dz)dt$  is a local martingale as the integrals with respect to the Wiener process  $dW_t$ , so  $M_t$  is a local martingale if and only if (11) hold. In order to prove that the likelihood process  $\frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_t} = \frac{M_t(\xi, \phi)}{M_0(\xi, \phi)}$  is a true martingale with  $\mathbb{E} \left[ \frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_t} \right] = 1$ , we need to apply a uniform integrability criterion separately for the diffusion and the jump part in the dynamics (8) and (9), since they are independent. The Novikov condition grants that the diffusion part of  $M_t$  is a true martingale, while the criterion proposed by Sokol and Hansen (2015), which in this case holds thanks to the non-explosion condition (10), grants that the jump part of  $M_t$  is a true martingale as well. A more detailed proof of the martingale property for the jump part of  $M_t$  can be found in (Brachetta *et al.*, 2023, Prop. 2.6). ■

In strict analogy with Hainaut and Moraux (2018), by defining the moment-generating function under  $\mathbb{Q}$  in the following way:  $\psi^{\mathbb{Q}}(z) = \frac{\exp(\kappa_1(\xi)\beta)\psi(z+\xi)}{\exp(\kappa_1(\xi)\beta)\psi(\xi)}$ , i.e.  $\psi^{\mathbb{Q}}(z) = \frac{\psi(z+\xi)}{\psi(\xi)}$ , it is possible to prove that the measure change just introduced by the likelihood process  $\frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_t} = \frac{M_t(\xi, \phi)}{M_0(\xi, \phi)}$  preserves the model structure. We have the following theoretical result.

**Proposition 2.** *The dynamics under the risk-neutral measure  $\mathbb{Q}$  of  $X_t^{\mathbb{Q}}, \lambda_t^{\mathbb{Q}}$  is given by the following system of stochastic differential equations:*

$$dX_t^{\mathbb{Q}} = \left( r - \frac{1}{2}\sigma^2 - \mathbb{E} \left[ e^{J^{\mathbb{Q}}} - 1 \right] \lambda_t^{\mathbb{Q}} - \delta \right) dt + \sigma dW_t^{\mathbb{Q}} + J^{\mathbb{Q}} dN_t^{\mathbb{Q}}, \quad (12)$$

$$d\lambda_t^{\mathbb{Q}} = k^{\mathbb{Q}}(\theta^{\mathbb{Q}} - \lambda_t^{\mathbb{Q}})dt + \beta^{\mathbb{Q}} dN_t^{\mathbb{Q}}, \quad (13)$$

where  $J^{\mathbb{Q}}N_t^{\mathbb{Q}}$  denotes the jump process with respect to  $\mathbb{Q}$  and  $J^{\mathbb{Q}}$  the jump size with respect

to  $\mathbb{Q}$ . The relations between the relevant parameters under  $\mathbb{P}$  and  $\mathbb{Q}$  are the following:

$$\begin{aligned} dW_t^{\mathbb{Q}} &= dW_t^{\mathbb{P}} + \phi(t)dt, \\ k^{\mathbb{Q}} &= k, \theta^{\mathbb{Q}} = \theta e^{\kappa_1(\xi)\beta} \psi(\xi), \\ \beta^{\mathbb{Q}} &= e^{\kappa_1(\xi)\beta} \psi(\xi) \beta. \end{aligned}$$

Moreover, under  $\mathbb{Q}$  the jump size  $J_x^{\mathbb{Q}}$  still have a normal distribution with mean  $\mu_J^{\mathbb{Q}}$  and volatility  $\sigma_J^{\mathbb{Q}}$ , with moment generating function given by  $\psi^{\mathbb{Q}}(z) = \psi(z + \xi)/\psi(\xi)$ .

**Proof.** The conditional moment generating function of  $\lambda$  under  $\mathbb{Q}$  is

$$\mathbb{E}^{\mathbb{Q}}[\exp(u\lambda_T^{\mathbb{Q}})|\mathcal{F}_t] = \exp(-m_t) \mathbb{E}^{\mathbb{P}}[\exp(m_T + ue^{\kappa_1(\xi)\beta} \psi(\xi) \lambda_T^{\mathbb{P}})|\mathcal{F}_t].$$

If we denote by  $f(t, u, \lambda, m) = \mathbb{E}^{\mathbb{P}}[\exp(m_T + ue^{\kappa_1(\xi)\beta} \psi(\xi) \lambda_T^{\mathbb{P}})|\mathcal{F}_t]$  the conditional expectation, which is a martingale by definition, we can apply Itô's lemma and impose a vanishing condition on the drift. In this way we obtain the following PIDE:

$$\begin{aligned} \frac{\partial f}{\partial t} + [\kappa_1(\xi)k(\theta - \lambda) + \kappa_1(\xi)] \frac{\partial f}{\partial m} + k(\theta - \lambda) \frac{\partial f}{\partial \lambda} - \frac{1}{2} \phi^2 \frac{\partial^2 f}{\partial m^2} \\ + \lambda \int_{-\infty}^{+\infty} [f(t, \lambda + \beta, m + \kappa_1 \beta + \xi z) - f(t, \lambda, m)] \nu(dz) = 0. \end{aligned} \quad (14)$$

If we guess a solution of the following form:

$$f(t, \lambda_t, m_t) = \exp[A(t, T) + e^{\kappa_1(\xi)\beta} \psi(\xi) B(t, T) \lambda_t + C(t, T) m_t],$$

with terminal conditions  $A(T, T) = 0, B(T, T) = u, C(T, T) = 1$ , and denote by  $\epsilon = e^{\kappa_1(\xi)\beta} \psi(\xi)$ , we can write:

$$\frac{\partial f}{\partial t} = \left[ \frac{\partial A}{\partial t} + \epsilon \lambda_t \frac{\partial B}{\partial t} + m_t \frac{\partial C}{\partial t} \right] f, \frac{\partial f}{\partial m} = C f, \frac{\partial^2 f}{\partial m^2} = C^2 f, \frac{\partial f}{\partial \lambda} = \epsilon B f.$$

By inserting these expressions into (14) we obtain the following equation, that must be satisfied for every value of  $\lambda, m$ :

$$\begin{aligned} \frac{\partial A}{\partial t} + [\epsilon k \theta B + k \theta \kappa_1(\xi) C + \kappa_2(\xi) C] + m_t \frac{\partial C}{\partial t} \\ + \lambda_t \left[ \epsilon \frac{\partial B}{\partial t} - \kappa_1(\xi) k C - \epsilon k B + \int_{-\infty}^{+\infty} [e^{C \kappa_1(\xi) \beta + B \epsilon \beta + \xi z} - 1] d\nu(z) \right] - \frac{1}{2} \phi^2 C + \frac{1}{2} \phi^2 C^2 = 0. \end{aligned}$$

We obtain  $C(t, T) = 1$  and

$$\begin{cases} \frac{\partial A}{\partial t} + \epsilon k \theta B + k \theta \kappa_1(\xi) + \kappa_2(\xi) = 0 \\ \epsilon \frac{\partial B}{\partial t} - \kappa_1(\xi) k - \epsilon k B + \int_{-\infty}^{+\infty} [e^{\kappa_1(\xi) \beta + \epsilon B \beta + \xi z} - 1] d\nu(z) = 0. \end{cases}$$

Finally, by using conditions (11), we get:

$$\begin{cases} \frac{\partial A}{\partial t} = -\epsilon k \theta B \\ \epsilon \frac{\partial B}{\partial t} = k \epsilon B - e^{\epsilon \beta B} + 1. \end{cases} \quad (15)$$

We can repeat the same computation of the conditional moment generating function of  $\lambda$  under  $\mathbb{P}$  without the introduction of the terms  $m_t$  and  $m_T$  and we get the following system of ordinary differential equations:

$$\begin{cases} \frac{\partial A}{\partial t} = -k \theta B \\ \frac{\partial B}{\partial t} = k B - e^{\beta B} + 1. \end{cases} \quad (16)$$

By comparing (15) written under  $\mathbb{Q}$  with (16), written under  $\mathbb{P}$ , the result follows, since they are the same, but under  $\mathbb{Q}$  the unknown  $B$  appears multiplied by  $\epsilon = e^{\kappa_1(\xi)\beta}\psi(\xi)$ . ■

**Remark 1.** *We shall assume in the following that the risk premium term  $\phi$  is such that the following equality is satisfied for all  $t \in [0, T]$ :*

$$\mu - r + \lambda_t[(e^{\kappa_1(\xi)\beta}\psi(\xi) - 1)(e^{\mu_J^{\mathbb{P}} + \sigma_J^2/2} - 1)] = \sigma\phi(t), \quad (17)$$

*in such a way that the dynamics of log-returns under  $\mathbb{Q}$  can be written as in (12).*

**Remark 2.** *We remark that, on the basis of the present construction, the relations between the mean and the variance of the jump size with respect to  $\mathbb{Q}$  and  $\mathbb{P}$  are the following:  $\mu_J^{\mathbb{Q}} = \mu_J^{\mathbb{P}} + \xi\sigma_J^2$  and  $\sigma_J^{\mathbb{Q}} = \sigma_J$ .*

The dynamics of the present model under  $\mathbb{Q}$  can then be written as follows:

$$d \ln(S_t) = \left( r - \delta - \frac{\sigma^2}{2} - \lambda_t \mathbb{E}[e^J - 1] \right) dt + \sigma dW_t + J_t dN_t \quad (18)$$

$$d\lambda_t = k(\theta - \lambda_t)dt + \beta dN_t, \quad (19)$$

where we dropped the superscript  $\mathbb{Q}$  on  $\lambda_t, J_t, k, \theta, \beta$  for simplicity of notation.

## 7. Model estimation

We perform an empirical estimation of the model (12)–(13) using WTI spot prices ranging from 05 October 2006 to 19 April 2020, for a total of 3531 daily observations. The model under consideration is highly non-linear and non-Gaussian, therefore standard Kalman filtering techniques cannot be applied in this context. Hence, in order to conduct inference, we rely on the Continuous Sequential Importance Resampling (CSIR) algorithm of Malik and Pitt (2011). This technique allows to perform Simulated Maximum

| $\Theta$ | Dist       | Support         | $(\mu_0, \sigma_0)$ | $\Theta$   | Dist       | Support         | $(\mu_0, \sigma_0)$ |
|----------|------------|-----------------|---------------------|------------|------------|-----------------|---------------------|
| $\mu$    | Tr. Normal | $[-0.20, 0.20]$ | $(0.00, 0.10)$      | $\sigma$   | Tr. Normal | $[0.00, 1.00]$  | $(0.20, 2.00)$      |
| $\mu_J$  | Tr. Normal | $[-0.20, 0.20]$ | $(-0.03, 0.20)$     | $\sigma_J$ | Tr. Normal | $[0.00, 0.30]$  | $(0.05, 0.20)$      |
| $k$      | Tr. Normal | $[0.00, 20.0]$  | $(2.50, 4.00)$      | $\theta$   | Tr. Normal | $[0.00, 20.0]$  | $(2.10, 4.00)$      |
| $\beta$  | Tr. Normal | $[0.00, 20.0]$  | $(2.00, 4.00)$      | $\delta$   | Tr. Normal | $[-1.50, 1.50]$ | $(0.00, 1.00)$      |

Table 3: Priors specification.

Likelihood Estimation (SMLE) in non-linear and non-Gaussian state-space models using particle filtering methods. With respect to standard SIR algorithms, the CSIR replaces the resampling step with a linear interpolation. By this way the simulated log-likelihood obtained with the particle filter will be smooth and prone to numerical optimization. To implement the CSIR we need to cast model (12)-(13) in a state-space form. To this end, we consider a first order Euler approximation to the continuous-time system. Denoting by  $\Delta t$  a small time interval, we can define the measurement and transition equations, under the historical measure  $\mathbb{P}$ , as follows

$$X_t = X_{t-1} + \left( \mu - \delta - \frac{\sigma^2}{2} - \lambda_{t-1} \mathbb{E}[e^J - 1] \right) \Delta t + \sigma \sqrt{\Delta t} Z_t + J_t \Delta N_t, \quad (20)$$

$$\lambda_t = \lambda_{t-1} + k(\theta - \lambda_{t-1})\Delta t + \beta \Delta N_t, \quad (21)$$

where  $Z_t \sim \mathcal{N}(0, 1)$ ,  $J_t \sim \mathcal{N}(\mu_J, \sigma_J^2)$  and  $\Delta N_t = N_t - N_{t-1} \sim \text{Bernoulli}(\lambda_{t-1} \Delta t)$ .

Given the state-space model (20)-(21), the SMLE procedure is performed in two steps: first, we sample  $M = 10^4$  random model parameters from the prior distributions in Table 3. The hyperparameters of the prior distributions are chosen to be consistent with those estimated in the recent literature (see e.g. Fulop and Li, 2019; Brignone *et al.*, 2023b). We use a truncated Normal prior to respect the domain of the parameters and to incorporate some upper and lower bounds used later in the optimization step. Given the  $M$  set of parameters, we evaluate the log-likelihood using our CSIR algorithm with  $N = 100$  particles<sup>5</sup> for each of them and take the 8 random model parameter combinations corresponding to the highest log-likelihood. Second, we run 8 optimizations using those model parameters as starting points and take as MLE the parameters with the highest (maximized) log-likelihood. For this second step we use  $N = 3000$  particles which is approximately equal to the number of total observations and ensures that the simulated likelihood estimator is not too noisy (see e.g. Andrieu *et al.*, 2010). During both steps we enforce the non-explosion condition (10):  $k > \beta$ .

In Table 4 we present the parameters estimated through the CSIR algorithm. Few

---

<sup>5</sup>We performed several experiments and we found that, for the log-likelihood computation in step 1,  $N = 100$  and  $M = 10^4$  are enough to fully represent the spectrum of reasonable starting points. A higher number of particles  $N$  is only needed to control the Monte Carlo variance of the likelihood estimator, but not to increase its level. Therefore, we increase  $N$  only for the subsequent optimization in order to stabilize the inferential procedure.



| $\Theta$ | Estimate | Std Error | $\Theta$   | Estimate | Std Error |
|----------|----------|-----------|------------|----------|-----------|
| $\mu$    | -0.0726  | 0.1163    | $\theta$   | 3.8132   | 0.3361    |
| $\delta$ | -0.7693  | 0.1382    | $\beta$    | 10.1608  | 0.3599    |
| $\sigma$ | 0.2809   | 0.0039    | $\mu_J$    | 0.0165   | 0.0037    |
| $k$      | 11.2570  | 0.2966    | $\sigma_J$ | 0.0987   | 0.0048    |

Table 4: Estimated parameters for the model in (20)–(21) on WTI log-returns. The log-likelihood at the estimated parameters is equal to  $8.51E + 03$ .

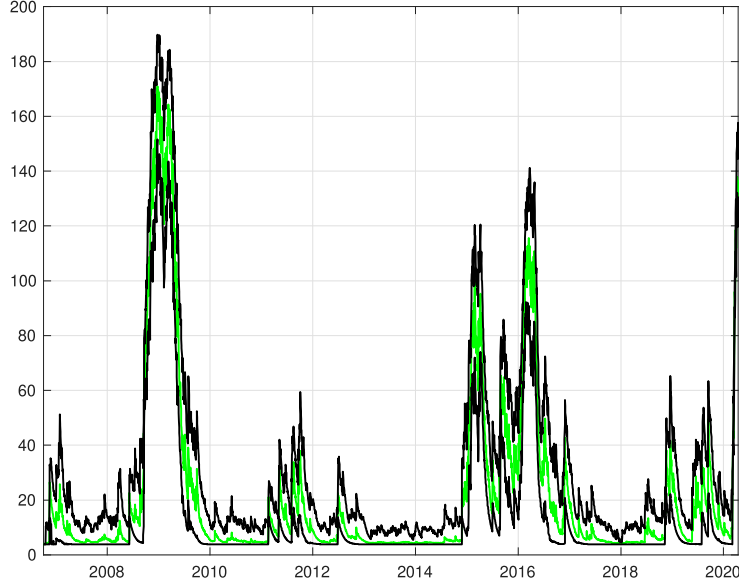


Figure 5: Filtered jump intensity (green line) and 5% and 95% quantiles (black line).

comments are in order. First, all parameters are statistically significant (except  $\mu$ ), which confirms that our estimation strategy can deliver reliable parameter estimates. The result for  $\mu$  is somehow expected since the sample mean of log-returns is almost equal to zero. Nevertheless, the convenience yield  $\delta$  is negative and highly significant. Second, in line with the existing literature, we confirm that jumps are self-exciting in the oil market. Third, the size of jumps is on average positive and the associated parameters are identified with high accuracy. Next, we plot in Figure 5 the filtered jump intensity. From the figure we can appreciate the tightness of confidence intervals, which is another evidence in favor of our estimation approach. In addition, from Figure 6 we note that the filtered jump intensity  $\lambda_t$  is high during periods of market stress with many jumps in close succession and reverts to its long run mean in normal periods.

## 8. Model simulation and exotic derivatives pricing

In this section, we outline some results on affine option pricing models which allows to price quickly and exactly geometric (both with continuous and discrete monitoring) Asian options, which are very popular in commodity markets (see e.g. Kaminski, 1999; Shiraya

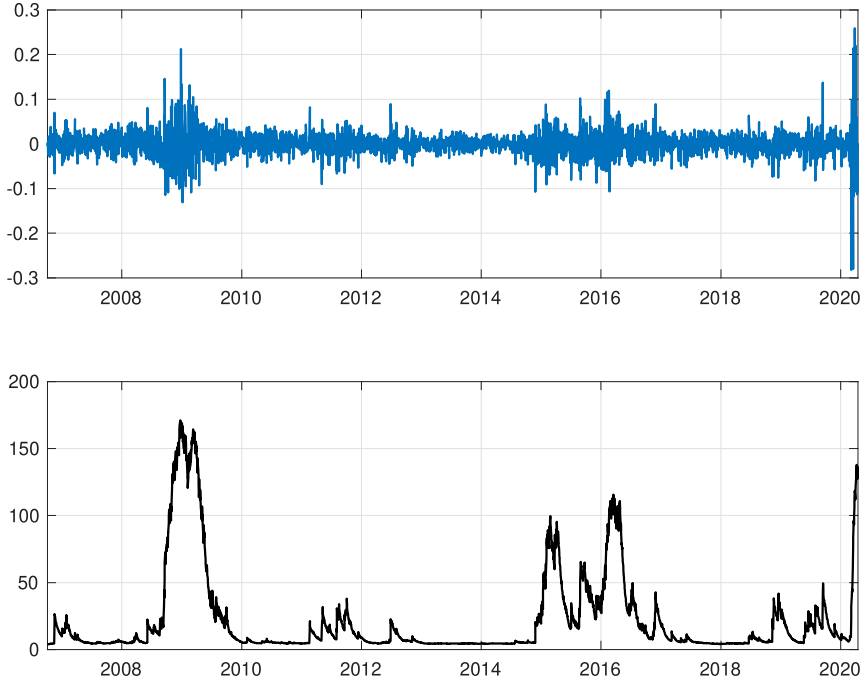


Figure 6: WTI log-returns (top panel) and filtered jump intensity (bottom panel).

and Takahashi, 2011; Brignone *et al.*, 2023b). Then, we develop an exact simulation scheme for the model (18)–(19) which allows to price other derivative instruments with more involved payoff structure, such as arithmetic Asian, Bermudan and Bermudan-Asian options.

### 8.1. Asian options

The following result provides the joint moment generating function of the log-returns and jump intensity.

**Proposition 3.** Define  $X_T := \log\left(\frac{S_T}{S_t}\right)$ , the following holds

$$\mathbb{E}[e^{u_1 X_T + u_2 \lambda_T} \mid \mathcal{F}_t] = e^{u_1 X_t + A(u_1, u_2, \tau) + B(u_1, u_2, \tau) \lambda_t},$$

where  $\tau = T - t$  and

$$\begin{cases} \frac{\partial A(u_1, u_2, \tau)}{\partial \tau} &= u_1 \left( r - \delta - \frac{\sigma^2}{2} \right) + u_1^2 \frac{\sigma^2}{2} + k\theta B(u_1, u_2, \tau), \\ \frac{\partial B(u_1, u_2, \tau)}{\partial \tau} &= -kB(u_1, u_2, \tau) + \left( e^{u_1 \mu_J + u_1^2 \sigma_J^2 / 2 + \beta B(u_1, u_2, \tau)} - 1 \right) - (e^{\mu_J + \sigma_J^2 / 2} - 1)u_1, \end{cases}$$

with initial conditions  $A(u_1, u_2, 0) = 0$ ,  $B(u_1, u_2, 0) = u_2$ .

**Proof.** The proof strictly follows (Brignone *et al.*, 2023b, Proposition 4). ■

Proposition 3 can be further extended.

**Proposition 4.** *The following holds*

$$M(u_1, u_2, u_3, u_4) := \mathbb{E}[e^{u_1 X_T + u_2 \lambda_T + u_3 \int_t^T X_s ds + u_4 \int_t^T \lambda_s ds} \mid \mathcal{F}_t] = e^{(u_1 + u_3 \tau) X_t + A(u_1, u_2, u_3, u_4, \tau) + B(u_1, u_2, u_3, u_4, \tau) \lambda_t},$$

where

$$\begin{cases} \frac{\partial A(u_1, u_2, u_3, u_4, \tau)}{\partial \tau} = (u_1 + u_3 \tau) \left( \mu - \delta - \frac{\sigma^2}{2} \right) + (u_1 + u_3 \tau)^2 \frac{\sigma^2}{2} + k \theta B(u_1, u_2, u_3, u_4, \tau), \\ \frac{\partial B(u_1, u_2, u_3, u_4, \tau)}{\partial \tau} = -k B(u_1, u_2, u_3, u_4, \tau) + \left( e^{(u_1 + u_3 \tau) \mu_J + (u_1 + u_3 \tau)^2 \sigma_J^2 / 2 + \beta B(u_1, u_2, u_3, u_4, \tau)} - 1 \right) \\ - (e^{\mu_J + \sigma_J^2 / 2} - 1)(u_1 + u_3 \tau) + u_4, \end{cases}$$

**Proof.** See Brignone and Sgarra (2020, p. 106). ■

Finally, we state the following theoretical result which follows from Fusai and Kyriakou (2016).

**Proposition 5.** *Suppose  $n \in \mathbb{N}$  and  $\Delta = \frac{T}{n}$  (we are assuming equally-spaced dates for simplicity),  $0 := t_0 < t_1 < \dots < t_n := T$ , the following holds*

$$M_n(u) := \mathbb{E} \left[ e^{u_1 \sum_{j=0}^n X_{t_j}} \mid \mathcal{F}_0 \right] = e^{\sum_{j=1}^n A_j + B_n \lambda_0}$$

where

$$A_0 := 0, \quad A_j := A(ju, B_{j-1}, \Delta), \quad B_0 := 0, \quad B_j := B(ju, B_{j-1}, \Delta).$$

**Proof.** We provide a proof which is simpler than the one proposed in Fusai and Kyriakou (2016). We assume for simplicity that  $X_0 = 0$  and  $t_j - t_{j-1} = \Delta \forall j \in \{1, n\}$ . By induction, for  $n = 2$

$$\begin{aligned} \mathbb{E}[e^{u(X_{t_1} + X_{t_2})} \mid \mathcal{F}_0] &= \mathbb{E}[e^{uX_{t_1}} \mathbb{E}[e^{uX_{t_2}} \mid \mathcal{F}_1] \mid \mathcal{F}_0] = \exp \left( A(u, 0, \Delta) + A(2u, B(u, 0, \Delta), \Delta) + \right. \\ &\quad \left. B(2u, B(u, 0, \Delta), \Delta) \lambda_0 \right) = e^{A_1 + A_2 + B_2 \lambda_0}. \end{aligned}$$

For  $n = 3$ ,

$$\begin{aligned} \mathbb{E}[e^{u(X_{t_1} + X_{t_2} + X_{t_3})} \mid \mathcal{F}_0] &= \mathbb{E}[e^{u(X_{t_1} + X_{t_2})} \mathbb{E}[e^{uX_{t_3}} \mid \mathcal{F}_2] \mid \mathcal{F}_0] = e^{A_1} \mathbb{E}[e^{uX_{t_1}} \mathbb{E}[e^{2uX_{t_2} + B_1 \lambda_{t_2}} \mid \mathcal{F}_1] \mid \mathcal{F}_0] \\ &\quad e^{A_1 + A_2} \mathbb{E}[e^{3X_{t_1} + B_2 \lambda_{t_1}} \mid \mathcal{F}_0] = e^{A_1 + A_2 + A_3 + B_3 \lambda_0}. \end{aligned}$$

Proceeding in the same way for  $n > 3$  one gets the thesis. ■

We will use these results to price geometric Asian options in Section 8.3.

## 8.2. Exact simulation of (18)–(19)

The simulation of the stochastic jump intensity model in (18)–(19) crucially depends on the simulation of the Hawkes process with exponential kernel. Based on our discussion/numerical results in Section 5 we will adopt the exact simulation scheme developed

by Dassios and Zhao (2013) (see Algorithm 4) for this purpose.

Given an initial date  $t_0 = 0$ , a final date  $T$ , an initial value for  $\lambda_0$  and the parameters of the model, the algorithm allows to obtain a sample of the triplet  $(N_T, \{\tau_j\}_{j=1}^{N_T}, \{\lambda_{\tau_j}\}_{j=1}^{N_T})$ , where  $\tau_j$  is the  $j$ -th jump time. Given the triplet, we can compute:

$$\lambda_T = \theta + (\lambda_{\tau_{N_T}} - \theta)e^{-k(T-\tau_{N_T})}, \quad \int_0^T \lambda_s ds = -\frac{\lambda_T - \lambda_0 - k\theta T - \beta N_T}{k}.$$

Then

$$\left(X_T \middle| \lambda_T, \int_0^T \lambda_s ds, \{J_{x,j}\}_{j=1}^{N_T}, N_T\right) \sim \mathcal{N}(m, s^2) \quad (22)$$

where  $\{J_{x,j}\}_{j=1}^{N_T}$  is a sequence of i.i.d. normal (with mean  $\mu_J$  and standard deviation  $\sigma_J$ ) random variables

$$m = \left(r - \delta - \frac{\sigma^2}{2}\right)T - \left(e^{\mu_J + \frac{\sigma_J^2}{2}} - 1\right) \int_0^T \lambda_s ds + \sum_{j=1}^{N_T} J_{x,j}, \quad s^2 = \sigma^2 T. \quad (23)$$

Hence, is possible to simulate  $X_T$  given  $\lambda_0$  exactly and efficiently. Naturally, it is possible to extend the methodology in order to generate paths for log-returns process observed at a discrete time grid, we summarize the procedure in Algorithm 5, which will be used later for pricing exotic derivatives.

---

**Algorithm 5** Exact simulation of the model in (18)–(19)

---

**Input:**  $X_0, \lambda_0, r, \delta, k, \theta, \beta, \sigma, \mu_J, \sigma_J, \{t_0, t_1, \dots, t_n := T\}$

**Output:**  $\{X_{t_j}\}_{j=1}^n$

---

- 1: Set  $\bar{k} = 0$
  - 2: Simulate the triplet  $(N_T, \{\tau_j\}_{j=1}^{N_T}, \{\lambda_{\tau_j}\}_{j=1}^{N_T})$  using Algorithm 4
  - 3: **for**  $j = 1 : n$  **do**
  - 4:   Compute number of jumps in the interval:  $A = \#\{\bar{k} : t_{j-1} < \tau_{\bar{k}} \leq t_j\}$
  - 5:   **if**  $A = 0$  **then**
  - 6:      $L = 0$  and  $\lambda_{t_j} = \theta + (\lambda_{t_{j-1}} - \theta)e^{-k(t_j - t_{j-1})}$
  - 7:   **else**
  - 8:      $L = \sum_{\bar{k}} \beta \mathbb{I}_{t_{j-1} < \tau_{\bar{k}} \leq t_j}$
  - 9:      $\bar{k} = \bar{k} + A$
  - 10:     $\lambda_{t_j} = \theta + (\lambda_{\tau_{\bar{k}}} - \theta)e^{-k(t_j - \tau_{\bar{k}})}$
  - 11:    **end**
  - 12:    Compute  $\int_{t_{j-1}}^{t_j} \lambda_s ds = -\frac{\lambda_{t_j} - \lambda_{t_{j-1}} - k\theta(t_j - t_{j-1}) - L}{k}$
  - 13:    Draw  $\{J_{x,j}\}_{j=1}^A \sim \mathcal{N}(\mu_J, \sigma_J^2)$
  - 14:    Draw  $G \sim \mathcal{N}(0, 1)$
  - 15:     $X_{t_j} = X_{t_{j-1}} + \left(r - \delta - \frac{\sigma^2}{2}\right)(t_{j-1} - t_j) - \left(e^{\mu_J + \frac{\sigma_J^2}{2}} - 1\right) \int_{t_{j-1}}^{t_j} \lambda_s ds + \sum_{j=1}^A J_{x,j} + G\sigma\sqrt{t_{j-1} - t_j}$
  - 16: **end**
-

|                  | Set A ( $\xi = -0.25$ ) | Set B ( $\xi = 0$ ) | Set C ( $\xi = 0.1$ ) | Set D ( $\xi = 0.25$ ) |
|------------------|-------------------------|---------------------|-----------------------|------------------------|
| $r$              | 0.0200                  | 0.0200              | 0.0200                | 0.0200                 |
| $\delta$         | -0.7693                 | -0.7693             | -0.7693               | -0.7693                |
| $\sigma$         | 0.2809                  | 0.2809              | 0.2809                | 0.2809                 |
| $k_\lambda$      | 11.2570                 | 11.2570             | 11.2570               | 11.2570                |
| $\theta_\lambda$ | 3.6859                  | 3.8132              | 3.8869                | 4.0748                 |
| $\beta$          | 9.8216                  | 10.1608             | 10.3573               | 10.8578                |
| $\mu_J$          | 0.0141                  | 0.0165              | 0.0175                | 0.0189                 |
| $\sigma_J$       | 0.0987                  | 0.0987              | 0.0987                | 0.0987                 |

Table 5: Parameters under  $\mathbb{Q}$ .

### 8.3. Pricing exotic derivatives via Monte Carlo simulation

In what follows, we show how to price exotic derivatives under the model (18)–(19) through Algorithm 5. We start by pointing out that model parameters estimated in Section 7 and reported in Table 4 are under the historical measure  $\mathbb{P}$ . However, derivatives pricing must be performed under the risk-neutral measure  $\mathbb{Q}$ . This can be done by specifying a risk-premium parameter  $\xi$ . This should be in principle computed from real option data, for example, by minimizing the distance between model and market implied European call options on futures prices. However, this additional numerical study is beyond the scope of this work. Following Hainaut and Moraux (2018), we will simply assume different possible values of  $\xi$  and show how to compute the parameters under  $\mathbb{Q}$  exploiting the results in Section 6. First, we fix  $r = 2\%$  and  $k_1(\xi) := a\xi$ , where  $a \in \mathbb{R}$ . Second, we compute  $\psi = e^{\xi\mu_J + \xi^2/2\sigma_J^2}$ . Then, we need to recover  $a$  from (11). Finally, from Proposition 2 we have

$$\theta^{\mathbb{Q}} = \theta \exp(k_1\beta^{\mathbb{P}})\psi, \quad \beta^{\mathbb{Q}} = \exp(k_1\beta^{\mathbb{P}})\psi\beta^{\mathbb{P}}, \quad \mu_J^{\mathbb{Q}} = \mu_J + \xi\sigma_J^2.$$

Resulting parameters under  $\mathbb{Q}$  are reported in Table 5. At this point, we can perform option pricing. We start by defining the payoff of arithmetic and geometric Asian put options (the corresponding call option can be obtained through the standard Asian put/call parity):

$$\Pi(T) = \begin{cases} \max\left(0, K - \frac{1}{n+1} \sum_{j=0}^n S_{t_j}\right), & \text{arithmetic Asian} \\ \max\left(0, K - \exp\left(\frac{1}{n+1} \sum_{j=0}^n \log S_{t_j}\right)\right), & \text{discrete geometric Asian} \\ \max\left(0, K - \exp\left(\frac{1}{T} \int_0^T \log S_t dt\right)\right), & \text{continuous geometric Asian} \end{cases} \quad (24)$$

First of all, geometric Asian option prices can be computed exactly exploiting Propositions 3 and 5. Indeed, consider the characteristic functions  $\phi(u) := M(0, 0, \frac{iu}{T})$  and  $\phi_n(iu) = M_n(\frac{iu}{n+1})$  where  $i = \sqrt{-1}$ . We compute the price of discrete and continuously monitored geometric Asian options through the Fourier-cosine (COS) method by Fang and Oosterlee (2008). We refer to the original paper for more details on how option prices

|                    | Set A |       | Set B |       | Set C |       | Set D |       |
|--------------------|-------|-------|-------|-------|-------|-------|-------|-------|
|                    | Price | s.e.  | Price | s.e.  | Price | s.e.  | Price | s.e.  |
| European           | 1.680 | –     | 2.002 | –     | 2.272 | –     | 2.960 | –     |
| Geometric (cts)    | 0.897 | –     | 0.999 | –     | 1.110 | –     | 1.331 | –     |
| Geometric (discr.) | 0.865 | –     | 0.970 | –     | 1.082 | –     | 1.313 | –     |
| Arithmetic         | 0.739 | 0.001 | 0.821 | 0.001 | 0.910 | 0.001 | 1.086 | 0.001 |
| Bermudan           | 4.549 | 0.003 | 4.805 | 0.003 | 5.122 | 0.003 | 5.708 | 0.003 |
| Bermudan-Asian     | 1.761 | 0.001 | 1.826 | 0.001 | 1.931 | 0.002 | 2.066 | 0.002 |
| Time               | 5.49  | –     | 5.25  | –     | 5.80  | –     | 6.49  | –     |

Table 6: Pricing exotic options in self-exciting model. Option contract parameters:  $S_0 = K = 100$ ,  $n = 12$ . Additional model parameters: We set  $\lambda_0 = 4$  for Sets A and B,  $\lambda_0 = 4.5$  for Sets C and D. Arithmetic Asian, Bermudan and Bermudan-Asian options are priced through Monte Carlo simulations with  $10^7$  simulations. Computational time is expressed in minutes.

can be computed given a characteristic function. We point out that the truncation range suggested in Fang and Oosterlee (2008, Eq. 49) can be computed quickly and exactly. Indeed, since the proposed model (18)–(19) is affine, we can compute the cumulants of  $\left(\int_0^T X_s ds | X_0, \lambda_0\right)$  similarly to Brignone and Sgarra (2020, Eq. 19–21) (see also Feunou and Okou, 2018; Brignone *et al.*, 2023a for more details and sample codes). Given risk-neutral cumulants, we compute the truncation range as in Fang and Oosterlee (2008, Eq. 49). Numerical results are reported in Table 6 for At The Money (ATM) put options. Finally, given the model parameter estimates we can also simulate (exactly) risk-neutral asset price paths using Algorithm 5 and price derivatives with more involved payoff. We will focus on options with early exercise features, which are quite popular in energy markets (see e.g. Kaminski, 1999). To this aim, we price Bermudan and Bermudan-Asian put options using the Longstaff and Schwartz (2001) least squares method. Note that in the same way is also possible to price swing options, which are also popular in energy markets. We assume a maturity  $T = 1$  year and that is possible to early exercise at twelve equispaced dates. Results are reported in Table 6.

## 9. Concluding Remarks

In the present chapter, we provided a gentle introduction to Hawkes processes and their use in energy markets modeling. After defining the basic mathematical tools, we discussed the main results available in the literature and presented a specific model as an example in order to illustrate the main methodologies in use for parameters estimation and derivatives pricing. In spite of the huge amount of results obtained, the subject seems still open to further investigations in several directions. Interest is growing for more accurate and efficient estimation methods. New classes have been recently introduced generalizing Hawkes processes, like hybrid Hawkes (Morariu-Patrichi and Pakkanen, 2018) and fractional Hawkes (Ketelbuters and Hainaut, 2022), offering new possible modeling instruments for energy markets.

## References

- Ait-Sahalia, Y., Chaco-Diaz, J. and Laeven, R. J. A. (2015) Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, **117**, 585–606.
- Ait-Sahalia, Y. and Hurd, T. (2016) Portfolio choice in markets with contagion. *Journal of Financial Econometrics*, **14**, 1–28.
- Andrieu, C., Doucet, A. and Holenstein, R. (2010) Particle Markov chain Monte Carlo. *Journal of the Royal Statistical Society. Series B*, **72**, 269–342.
- Bacry, E., Delattre, S., Hoffman, M. and Muzy, J. (2013) Modelling microstructure noise by mutually exciting point processes. *Quantitative Finance*, **13**, 65–77.
- Bacry, E., Mastromatteo, I. and Muzy, J. (2015) Hawkes processes in finance. *Market Microstructure and Liquidity*, **1**, 1550005.
- Bernis, G., Brignone, R., Scotti, S. and Sgarra, C. (2021) A Gamma Ornstein-Uhlenbeck model driven by a Hawkes process. *Mathematics and Financial Economics*, **15**, 747–773.
- Blasberg, A., Graf von Luckner, N. and Kiesel, R. (2019) Modeling the serial structure of the Hawkes process parameters for market order arrivals on the German intraday power market. *IEEE Proceedings, 16th International Conference on the European Energy Market (EEM), Ljubljana, Slovenia, 2019*, 1–6.
- Boswijk, P., Laeven, R. A. and Lalu, A. (2016) Asset returns with self-exciting jumps: Option pricing and estimation with a continuum of moments. *Working Paper*.
- Brachetta, M., Callegaro, G., Ceci, C. and Sgarra, C. (2023) Optimal reinsurance via BSDEs in a partially observable model with jump clusters. *Finance and Stochastics, Forthcoming*, 1–39.
- Brignone, R., Gonzato, L. and Lütkebohmert, E. (2023a) Efficient quasi-Bayesian estimation of affine option pricing models using risk-neutral cumulants. *Journal of Banking and Finance*, **148**, 106745.
- Brignone, R., Gonzato, L. and Sgarra, C. (2023b) Commodity Asian option pricing and simulation in a 4-factor model with jump clusters. *Annals of Operations Research, Forthcoming*.
- Brignone, R. and Sgarra, C. (2020) Asian options pricing in Hawkes-type jump-diffusion models. *Annals of Finance*, **16**, 101–119.

- Callegaro, G., Gaïgi, M., Scotti, S. and Sgarra, C. (2017) Optimal investment in markets with over and under-reaction to information. *Mathematics and Financial Economics*, **11**, 299–322.
- Callegaro, G., Mazzoran, A. and Sgarra, C. (2022) A self-exciting framework for forward dynamics in power markets. *Applied Stochastic Models in Business and Industry*, **38**, 27–48.
- Cartea, Á., Jaimungal, S. and Ricci, J. (2014) Buy low, sell high: A high frequency trading perspective. *SIAM Journal of Financial Mathematics*, **5**, 415–444.
- Chavez-Demoulin, V., Davison, A. and McNeil, A. (2005) Estimating value-at-risk: a point process approach. *Quantitative Finance*, **5**, 227–234.
- Chavez-Demoulin, V. and McGill, J. (2012) High-frequency financial data modeling using Hawkes processes. *Journal of Banking and Finance*, **36**, 3415–3426.
- Clements, A., Herrera, R. and Hurn, A. (2015) Modelling interregional links in electricity price spikes. *Energy Economics*, **51**, 383–393.
- Daley, D. J. and Vere-Jones, D. (2002) *An introduction to the theory of point processes. Volume I: elementary theory and methods*.
- Dassios, A. and Zhao, H. (2012) Ruin by dynamic contagion claims. *Insurance: Mathematics and Economics*, **51**, 93–106.
- Dassios, A. and Zhao, H. (2013) Exact simulation of Hawkes process with exponentially decaying intensity. *Electronic Communications in Probability*, **18**, 1–13.
- Duffie, D. and Glynn, P. (1995) Efficient Monte Carlo estimation of security prices. *Annals of Applied Probability*, **4**, 897–9058.
- El Euch, O. and Rosenbaum, M. (2019) The characteristic function of rough Heston models. *Mathematical Finance*, **29**, 3–38.
- Errais, E., Giesecke, K. and Goldberg, L. (2010) Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics*, **1**, 642–665.
- Eyjolfsson, H. and Tjøstheim, D. (2018) Self-exciting jump processes with applications to energy markets. *Annals of the Institute of Statistical Mathematics*, **70**, 373–393.
- Fang, F. and Oosterlee, C. (2008) A novel pricing method for European options based on Fourier-cosine series expansions. *SIAM Journal on Scientific Computing*, **31**, 826–848.
- Favetto, B. (2008) The European intraday electricity market: a modeling based on the Hawkes process. *Journal of Energy Markets*, **13**, 57–96.



- Feunou, B. and Okou, C. (2018) Risk-neutral moment-based estimation of affine option pricing models. *Journal of Applied Econometrics*, **33**, 1007–1025.
- Filimonov, V., Bicchetti, D., Maystre, N. and Sornette, D. (2014) Quantification of the high level of endogeneity and of structural regime shifts in commodity markets. *Journal of International Money and Finance*, **42**, 174–192.
- Fox, E., Short, M. B., Schoenberg, F. P., Coronges, K. D. and Bertozzi, A. L. (2016) Modeling E-mail networks and inferring leadership using self-exciting point processes. *Journal of the American Statistical Association*, **111**, 564–584.
- Fulop, A. and Li, J. (2019) Bayesian estimation of dynamic asset pricing models with informative observations. *Journal of Econometrics*, **209**, 114–138.
- Fulop, A., Li, J. and Yu, J. (2015) Self-exciting jumps, learning, and asset pricing implications. *The Review of Financial Studies*, **28**, 876–912.
- Fusai, G. and Kyriakou, I. (2016) General optimized lower and upper bounds for discrete and continuous arithmetic Asian options. *Mathematics of Operations Research*, **41**, 531–559.
- Giordano, L. and Morale, D. (2021) A fractional Brownian-Hawkes model for the Italian electricity spot market: estimation and forecasting. *Journal of Energy Markets*, **14**, 65–109.
- Gonzato, L. and Sgarra, C. (2021) Self-exciting jumps in the oil market: Bayesian estimation and dynamic hedging. *Energy Economics*, **99**, 105279.
- Graf von Luckner, N. and Kiesel, R. (2021) Modeling market order arrivals on the German intraday electricity market with the Hawkes process. *Journal of Risk and Financial Management*, **14**, 1–31.
- Hainaut, D. (2016) Impact of volatility clustering on equity indexed annuities. *Insurance: Mathematics and Economics*, **71**, 367–381.
- Hainaut, D. (2017) Contagion modeling between the financial and insurance markets with time changed processes. *Insurance: Mathematics and Economics*, **74**, 63–77.
- Hainaut, D. and Moraux, F. (2018) Hedging of options in the presence of jump clustering. *Journal of Computational Finance*, **22**, 1–35.
- Hawkes, A. G. (1971a) Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 438–443.
- Hawkes, A. G. (1971b) Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, **58**, 83–90.

- Hawkes, A. G. (2018) Hawkes processes and their applications to finance: a review. *Quantitative Finance*, **18**, 193–198.
- Hawkes, A. G. (2022) Hawkes jump-diffusions and finance: a brief history and review. *The European Journal of Finance*, **28**, 627–641.
- Jaisson, T. and Rosenbaum, M. (2015) Limit theorems for nearly unstable Hawkes processes. *Annals of Applied Probability*, **25**, 600–631.
- Jang, J. and Dassios, A. (2013) A bivariate shot noise self-exciting process for insurance. *Insurance: Mathematics and Economics*, **53**, 524–532.
- Jiao, Y., Ma, C. and Scotti, S. (2017) Alpha-CIR model in sovereign interest rate modelling. *Finance and Stochastics*, **21**, 789–813.
- Jiao, Y., Ma, C., Scotti, S. and Sgarra, C. (2019) A branching process approach to power markets. *Energy Economics*, **79**, 144–156.
- Johnson, D. H. (1996) Point process models of single-neuron discharges. *Journal of Computational Neuroscience*, **3**, 275–299.
- Kaminski, V. (1999) *Managing Energy Price Risk*. London, Risk Books.
- Ketelbuters, J. and Hainaut, D. (2022) CDS pricing with fractional Hawkes processes. *European Journal of Operational Research*, **297**, 1139–1150.
- Kokholm, T. (2016) Pricing and hedging of derivatives in contagious markets. *Journal of Banking and Finance*, **66**, 19–34.
- Kyriakou, I., Brignone, R. and Fusai, G. (2023) Unified moment-based modelling of integrated stochastic processes. *Operations Research*, *Forthcoming*.
- Lewis, P. A. and Shedler, G. S. (1969) Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistic Quarterly*, **26**, 403–413.
- Longstaff, F. and Schwartz, E. (2001) Valuing American options by simulation: a simple least-squares approach. *Review of Financial Studies*, **14**, 113–147.
- Malik, S. and Pitt, M. K. (2011) Particle filters for continuous likelihood evaluation and maximisation. *Journal of Econometrics*, **165**, 190–209.
- Merton, R. (1976) Option pricing when underlying stock returns are discontinuous. *The Journal of Financial Economics*, **3**, 125–144.
- Meyer, S. and Held, L. (2014) Power-law models for infectious disease spread. *The Annals of Applied Statistics*, **8**, 1612–1639.

- Mohler, G., Short, M. B., Brantingham, P. J., Schoenberg, F. P. and Tita, G. E. (2011) Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, **106**, 100–108.
- Møller, J. and Rasmussen, J. G. (2005) Perfect simulation of Hawkes processes. *Advances in Applied Probability*, **37**, 629–646.
- Morariu-Patrichi, Y. and Pakkanen, M. (2018) Hybrid marked point processes: Characterization, existence and uniqueness. *Market Microstructure and Liquidity*, **4**, 1950007.
- Nakagawa, T., Subbey, S. and Solvang, H. K. (2019) Integrating Hawkes process and bio mass models to capture impulsive population dynamics. *Dynamics of Continuous, Discrete and Impulsive Systems Series B: Applications & Algorithms*, **26**, 153–170.
- Ogata, Y. (1981) On Lewis simulation method for point processes. *IEEE Transactions on Information Theory*, **27**, 23–31.
- Ogata, Y. (1998) Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, **50**, 379–402.
- Ozaki, T. (1979) Maximum likelihood estimation of Hawkes self-exciting point processes. *Annals of the Institute of Statistical Mathematics*, **31**, 145–155.
- Pardoux, E. (2016) *Probabilistic Models of Population Evolution*. Springer.
- Porter, M. D. and White, G. (2012) Self-exciting hurdle models for terrorist activity. *Annals of Applied Statistics*, **6**, 106–124.
- Shiraya, K. and Takahashi, A. (2011) Pricing average options on commodities. *The Journal of Futures Markets*, **31**, 407–439.
- Sokol, A. and Hansen, N. (2015) Exponential martingales and changes of measure for counting processes. *Stochastic Analysis and Applications*, **33**, 823–843.
- Stabile, G. and Torrisi, G. (2010) Risk processes with non-stationary Hawkes claims arrivals. *Methodology and Computing in Applied Probability*, **12**, 415–429.
- Vinogradov, A., Agletdinov, E. and Merson, D. (2019) Mechanical twinning is a correlated dynamic process. *Scientific Reports*, **9**.
- Xu, L., Duan, J. A. and Whinston, A. (2014) Path to purchase: A mutually exciting point process model for online advertising and conversion. *Management Science*, **60**, 1392–1412.
- Zhang, X., Glynn, P., Giesecke, K. and Blanchet, J. (2009) Rare event simulation for a generalized Hawkes process. *IEEE Proceedings of the 2009 Winter Simulation Conference*, 1291–1298.

Zhu, L. (2013) Ruin probabilities for risk processes with non-stationary arrivals and subexponential claims. *Insurance: Mathematics and Economics*, **53**, 544–550.