Ann Inst Stat Math (2018) 70:373–393
https://doi.org/10.1007/s10463-016-0591-8

Self-exciting jump processes with applications to energy markets

Heidar Eyjolfsson[1] · **Dag Tjøstheim**[2]

Received: 4 August 2015 / Revised: 20 October 2016 / Published online: 6 January 2017

Abstract In this paper, we discuss a class of mean-reverting, and self-exciting continuous-time jump processes. We give a short overview, with references, of the development of such processes, discuss maximum likelihood estimation, and put them into context with processes that have been proposed recently. More specifically, we introduce a class of SDE-governed intensity processes with varying jump intensity. We study Markovian aspects of this process, and analyse its stability properties. Finally, we consider parameter estimation of our model class with daily quotes of UK electricity prices over a specific period.

Keywords Self-exciting processes · Jump processes · Markov processes · Energy markets

1 Introduction

In recent years, modelling of financial markets and instruments by means of continuous-time jump processes has become common. In particular the class of Lévy processes has turned into a widely used tool in the literature (see e.g. Cont and Tankov 2004 for general applications, and Benth et al. 2008 for energy market applications). The class of Lévy processes is arguably the most natural extension of the continuous

✉ Heidar Eyjolfsson
heidar@hi.is

Dag Tjøstheim
Dag.Tjostheim@uib.no

1 Science Institute, University of Iceland, Dunhaga 5, 107 Reykjavík, Iceland

2 Department of Mathematics, University of Bergen, Postbox 7803, 5020 Bergen, Norway

Brownian motion paradigm, retaining independence and stationarity of increments, while excluding almost sure path-continuity.

However, one area where Lévy processes are not *directly* applicable is modelling of volatility clustering. Indeed, by the independence and stationarity of Lévy process increments, one rarely observes clustering of jumps. To tackle this problem, time-changed Lévy processes have been introduced. We propose a different approach. In this paper, we study a class of jump processes that are not Lévy processes, namely self-exciting jump processes, with stochastic jump sizes, which feed into the intensity process, and thus determine the future behaviour of the process. Our class of self-exciting processes is special in that its intensity process admits a stochastic differential equation, which makes the intensity process Markovian. We moreover discuss limiting behaviour and likelihood inference in this setting.

Modelling random phenomena by means of point processes, or a combination of continuous stochastic processes together with point processes is common; see e.g. Jacod (1974/1975), Brémaud (1981), Daley and Vere-Jones (1988), and parts of Jacod and Shiryaev (2003) for classical accounts. One of the simplest, and arguably the most important, conceivable point process that has been studied is the Poisson process (it is also a Lévy process), which on a given interval records the number of incidents which occur in an independent fashion. This means that distinct occurrences recorded by the Poisson process have no bearing on each other, that is, the arrival times of incidents are independent of each other. Many authors that use point processes in their models impose such an independent arrival time condition, which is also called a *memoryless property*, on their processes. While this restriction may admittedly serve to simplify models considerably, it is certainly not always warranted. Indeed, often the phenomena one seeks to model by means of point processes seem to arrive in clusters, rather than in an independently arriving fashion, as can be tested by means of statistical hypothesis tests and has been observed by several authors, including Aït-Sahalia et al. (2015).

An alternative to modelling by means of constant intensity processes, is to employ *self-exciting* processes, with stochastic jump-sizes, which essentially are processes that excite their own intensity. In particular this implies that large jumps are likely to be followed by a jump within a short time, thus triggering a potential jump clustering. The class of processes we propose is similar to the so-called Hawkes processes, see Hawkes (1971a, b); Hawkes and Oakes (1974), but with a random jump size. Recently, there has been a renewed interest in processes of this type, see Errais et al. (2010), Bacry and Muzy (2014), Bacry et al. (2013), Jaisson and Rosenbaum (2015) and Embrechts et al. (2011) for studies of self-exciting processes in a financial setting.

Under certain conditions, which will be made clear in Section 3, Markov process theory can be used to study the properties of the process and its intensity process. In particular, one can use the extended generator (as defined by Davis 1993), and Dynkin's formula to infer moments as solutions of ordinary differential equations. The extended generator can furthermore be employed to study stability properties of the process in the sense of Meyn and Tweedie (1993).

Our approach is to model the intensity process as a stochastic differential equation (SDE). This has the advantages that we can also model nonlinear effects via the SDE drift function, and the jump distribution of the SDE can be made to depend on the current intensity level, while still preserving the Markov property, so that e.g. high intensity may trigger large jumps which in turn may lead to increased intensity. We note that our class of self-exciting processes has a discrete time analogue. Indeed, if we discretize the SDE we get a model which resembles recent models for integer time series (see Fokianos et al. 2009) which in turn can be thought of as an analogue to GARCH models.

Having introduced our class of self-exciting processes, and examined its theoretical properties, we demonstrate its applicability with an example. We fit a self-exciting jump process to a dataset which we extract from UK electricity spot price data. To be more specific, we extract a jump process driven Ornstein–Uhlenbeck type trajectory from the UK spot price series, using a method developed by Meyer-Brandis and Tankov (2008), in the same manner as Meyer-Brandis and Morgan (2014). Having extracted the data in that manner, we proceed to fit it to a self-exciting process, and demonstrate the advantages of doing this, compared to fitting a constant intensity process to the data. Indeed we see that the self-exciting process does a good job replicating the periods of jump-clustering which are observed in reality.

The paper is structured as follows. In the next section we introduce a class of self-exciting processes, and discuss some of its basic properties. In Sect. 3, we show that our self-exciting process is a Markov process, and moreover employ Markov theory to study the process. In Sect. 4, we discuss an application of self-exciting processes to model UK electricity spot price data. Finally, we conclude in Sect. 5.

2 Self-exciting processes

In this section we give a short account of the development of self-exciting processes, including the Hawkes process, and introduce the model class which we work with in the current paper. Hawkes and Oakes (1974) give a cluster process interpretation of the self-exciting Hawkes process. They show that the Hawkes process can be thought of as a migrant point process in which there is a standard Poisson process with constant intensity $\lambda_0 > 0$ which records the number of migrants, and that associated to each migrant there is a further inhomogeneous Poisson process, starting at their migration time, with an intensity that tends to zero, which represent the number of descendants of that particular migrant, with each of these descendants in turn generating their own inhomogeneous Poisson descendant process, starting at their time of birth. Given this construction the Hawkes process is the sum of migrants and their descendants (of any generation) at a given time.

This sort of a construction is quite appealing in the sense that it is easy to interpret, and has a whole range of potential application areas, besides the earthquake, and aftershock model they were initially employed to model. For example in population dynamics, the migrants could represent the number of immigrants to a particular country, with the inhomogeneous Poisson processes representing their children. Similarly, one can apply analogies of this type to financial modelling of buy (or sell) orders (as Aït-Sahalia et al. 2015) with migrants representing exogenous orders, and descendants representing orders made as a result of previous orders. Thus, the Hawkes process, initially employed in seismological modeling, has started to find its way into mainstream financial modeling in recent years. In the current paper we propose to use a Markov type self-exciting processes to model jumps in electricity spot price markets.

Now let us give a mathematical description of self-exciting jump processes. To that end we introduce the following notation, which we shall employ throughout the paper; for a more detailed discussion see Jacod (1974/1975) and Brémaud (1981). Let $(\Omega, \mathcal{F})$ denote a measurable space, and let $\{(T_n)\}_{n\geq 1}$ denote a point process taking values in $\mathbb{R}_+$. The sequence $\{T_n\}_{n\geq 1}$ is assumed non-negative and non-decreasing, i.e. $0 \leq T_1 \leq T_2 \leq \cdots$ holds. We moreover introduce the counting process associated to the point process

$$N(t) := \sum_{n\geq 1} 1_{\{T_n \leq t\}}, \tag{1}$$

where $t \geq 0$, is the counting process which records all the jumps of the point process. We identify a point process with its counting process (1) and let

$$\mathcal{F}_t^N := \sigma\{N(s) : 0 \leq s \leq t\},$$

where $t \geq 0$. Given a point process adapted to some filtration $\{\mathcal{F}_t\}$, with $\mathcal{F}_t^N \subset \mathcal{F}_t$, suppose that $N(t)$ admits a càdlàg $\mathcal{F}_t$-adapted (and thus predictable) intensity $\lambda(t)$ (in the sense of Brémaud 1981), then

$$\mathbb{E}\left[\int_0^\infty f(s)\mathrm{d}N(s)\right] = \mathbb{E}\left[\int_0^\infty f(s)\lambda(s)\mathrm{d}s\right],$$

holds for all predictable $f : \Omega \times \mathbb{R}_+ \to [-\infty, \infty]$.

Given the above notation, the *Hawkes process* is a point process with a stochastic intensity given by

$$\lambda(t) = \lambda_0 + \int_0^t g(t-s)\mathrm{d}N(s),$$

where $\lambda_0 > 0$ is the base intensity, and g is a deterministic non-negative function, with $\|g\|_{L^1(\mathbb{R}_+)} < 1$. In the case when the kernel function of the above Hawkes model has the form $g(u) = \beta \mathrm{e}^{-\alpha u}$, we can show that $\lambda(t)$ is the solution to the following pure-jump Langevin equation

$$\mathrm{d}\lambda(t) = \alpha(\lambda_0 - \lambda(t))\mathrm{d}t + \beta\mathrm{d}N(t),$$

with $\lambda(0) = \lambda_0$. Indeed, if we suppose that the above SDE dynamics hold, and let $f(u,t) = u\mathrm{e}^{\alpha t}$, it follows from the Itô lemma (see Theorem II.33 in Protter 2005) applied to $t \mapsto f(\lambda(t), t)$ that

$$
\begin{aligned}
\lambda(t)\mathrm{e}^{\alpha t}-\lambda_0 &= \alpha\int_0^t \lambda(s)\mathrm{e}^{\alpha s}\mathrm{d}s+\int_0^t \mathrm{e}^{\alpha s}\mathrm{d}\lambda(s)\\
&= \alpha\int_0^t \lambda(s)\mathrm{e}^{\alpha s}\mathrm{d}s+\int_0^t \mathrm{e}^{\alpha s}\{\alpha(\lambda_0-\lambda(s))\mathrm{d}s+\beta\mathrm{d}N(s)\}\\
&= \lambda_0(\mathrm{e}^{\alpha t}-1)+\beta\int_0^t \mathrm{e}^{\alpha s}\mathrm{d}N(s),
\end{aligned}
$$

from which it follows that

$$
\lambda(t)=\lambda_0+\beta\int_0^t \mathrm{e}^{-\alpha(t-s)}\mathrm{d}N(s).
$$

Note in particular that a nice property of the above exponential self-exciting model is that for a fixed $\delta>0$, it is autoregressive, in the sense that

$$
\lambda(t+\delta)=(1-\mathrm{e}^{-\alpha\delta})\lambda_0+\mathrm{e}^{-\alpha\delta}\lambda(t)+\beta\int_t^{t+\delta}\mathrm{e}^{-\alpha(t+\delta-s)}\mathrm{d}N(s).
$$

Thus, as we shall see in the next section, at time t information about $\lambda(t)$ is sufficient to make predictions about a future time point $t+\delta$, which means that, given $\lambda(t)$, further information about the trajectory of the process prior to time t does not improve our future prediction, and that under certain regularity conditions discussed in the following section, in this case, the self-exciting process fulfills the Markov property. On the other hand, it is easy to see that kernel functions which are not on this form generally do not fulfill a Markov property. If we assume that the kernel function g is regular enough and that we observe no jumps in $(t,t+\delta]$, then we obtain by Taylor expansion that

$$
\begin{aligned}
\lambda(t+\delta) &= \lambda(t)+\int_0^{t+\delta}(g(t+\delta-s)-g(t-s))\mathrm{d}N(t)\\
&= \lambda(t)+\delta\int_0^{t+\delta} g'(t-s)\mathrm{d}N(s)+O(\delta^2).
\end{aligned}
$$

So, unless g is exponential, $\lambda(t)$ will not be autoregressive, or Markovian.

Inspired by the above constructions we now introduce a Markovian intensity process which is excited by its own jumps in a stochastic manner, to be more specific, we introduce a point process with SDE intensity dynamics, where the intensity process jumps whenever an associated point process jumps. The stochastic jump process is given by

$$
U(t)=\sum_{k=1}^{N(t)} X_k, \tag{2}
$$

where $N(t)$ is the counting process (1), and $\{X_k\}$ is a family of random variables, X_k has the distribution $\nu(\lambda(T_k-),\cdot)$, for a given family $\{\nu(\lambda,\cdot)\}_{\lambda>0}$ of probability distributions, and $t- := \lim_{s\uparrow t} s$. Thus, we allow the value of the intensity process immediately before the jump to influence the jump size distribution. We introduce the SDE

$$\mathrm{d}\lambda(t) = \mu(\lambda(t))\mathrm{d}t + \beta \mathrm{d}U(t), \tag{3}$$

where $\beta \in \mathbb{R}$ is a constant and we assume throughout the paper that $\mu : \mathbb{R}_+ \to \mathbb{R}$, is Lipschitz continuous. We denote by $\mathcal{F}_t^\lambda = \sigma\{\lambda(s) : s \le t\}$ the filtration generated by (3).

Definition 1 An SDE-driven self-exciting jump process is a point process with a $\mathcal{F}_t^\lambda$-adaptable intensity $\lambda(t)$, given by (3), with jump-sizes, X_k, which follow the distribution $\nu(\lambda(T_k-), \cdot)$, where $\{\nu(\lambda, \cdot)\}_{\lambda>0}$ is a family of probability distributions, and $\nu(\lambda, \cdot)$ is supported on $[\lambda_0 - \lambda, \infty)$.

This equation admits a unique solution, see Theorem V.7 in Protter (2005) for proof of existence and uniqueness. Note that since $\lambda(t)$ is càdlàg and adapted it is predictable. The advantages of specifying the intensity process as an SDE in the above manner are the following. First of all, it extends the exponential Hawkes model, in that the exponential Hawkes process is obtained as a special case. It moreover allows us to record stochastic jumps which can depend on the current value of the intensity process. Finally, as we shall see in the following section, it is Markovian, in the sense that given the state of the process at the current time, the future is independent of the past. The setting is rather general in the sense that one is free to specify quite general drift functions and jump-size distributions. Typically, the jumps of the point process excite the intensity in the sense that $\lambda(t)$ increases at the jump time, whereas the drift function, μ, drives $\lambda(t)$ back towards its mean level. However, if desirable, this relationship can be turned around, with the jumps contributing to a lower (but still positive) intensity and the drift pushing it up. Thus, we allow negative jumps, but the jump distribution support is not allowed to take the process below the basis (initial) intensity level, $\lambda_0 > 0$. The dependence of the jump-size distributions on the current state of the intensity process, and the deterministic motion of the intensity process between jumps, make it a Markov process. Note, furthermore, that the self-exciting jump process is related to marked point processes, without being one. Recall that a marked point process is a double sequence, $\{(T_n, X_n)\}_{n\ge1}$, where $\{T_n\}_{n\ge1}$ is a point process, and $\{X_n\}_{n\ge1}$ is a sequence of "marks" in some measurable space, which is also called the mark space. In the setting introduced in the definition above, one could think of the jumps introduced in the intensity process as marks. However, we refrain from doing so, since it only serves to complicate things without offering any obvious benefits.

To avoid the explosion of the counting process (1) it is necessary to balance the relationship of the drift, μ, and the jump-size distribution, ν. To that end, we ensure that the integrated intensity process is almost surely finite. Let

$$\Lambda(T) = \int_0^T \lambda(t)\mathrm{d}t,$$

for $T > 0$.

Assumption 1 For any $T > 0$, $\Lambda(T) < \infty$ holds almost surely.

According to Theorem II-T8 in Brémaud (1981), an SDE driven self-exciting process which satisfies the above assumption is non-explosive. As we have already indicated the drift function, μ, and the jump-size distribution, ν, need to balance the intensity process to guarantee that $\Lambda(T)$ is almost surely finite. One way of doing this is to compare the intensity SDE (3) with an SDE which is known to be integrable. Thus, for instance in the case when the drift function is decreasing, and the jump-size distribution has positive support, the intensity process can be compared with a zero-drift intensity process which has the same jump-size distribution. If the jump-size distribution does not grow too much, then the resulting zero-drift piecewise deterministic intensity process, can be thought of as the intensity of a birth process, which has an integrable intensity.

The above integrated intensity process determines the jump sequence $\{T_k\}$ in the sense that if $\{\xi_k\}$ is a sequence of independent exponentially distributed random variables with mean 1, then given $\mathcal{F}_{T_k}$,

$$T_{k+1} \stackrel{\mathcal{D}}{=} \inf\{t > T_k : \Lambda(t) - \Lambda(T_k) \geq \xi_k\}, \tag{4}$$

for $k \geq 0$. This means in particular that regions in which the intensity process (3) is high we expect jumps to occur closer to each other than in regions where the intensity (3) is small. It is, furthermore, relevant to note that the random variables $\{T_k\}_{k\geq 2}$ do not fulfill the memoryless property. Recall that a non-negative random variable, T, is called *memoryless* if

$$\mathbb{P}(T > t + s | T > s) = \mathbb{P}(T > t)$$

holds for all $s, t \geq 0$. We remark that it follows that $N(t)$ is not an additive (independent increment) process, and thus in particular not a Lévy process.

Example 1 As an example of a non-linear drift function we might consider a stochastic intensity model on the form

$$\mathrm{d}\lambda(t) = (\alpha + \delta \exp(-\gamma\lambda(t)^2))(\lambda_0 - \lambda(t))\mathrm{d}t + \beta \mathrm{d}U(t),$$

where $\alpha, \beta, \gamma, \delta > 0$. This model extends the linear model in the sense that the speed of mean-reversion varies between $\alpha + \delta \mathrm{e}^{-\gamma\lambda_0^2}$ for $\lambda(t) = \lambda_0$ and tends to α for large values of $\lambda(t)$. Thus, in a neighbourhood where the intensity is high the speed of mean reversion is lower than in neighbourhoods where it is low, which could have the interpretation that in periods of low activity the effects of a jump fade out faster than in periods of high activity. In an energy market context this would in turn mean that in a period where there are few jumps in prices, a jump will typically not excite the market as much as it would in times of high activity. In Fig. 1, we have simulated a trajectory of the non-linear model, using a thinning algorithm, see Ogata (1981), which displays the characteristics we have described. This example is a continuous time analogue of an exponential autoregressive time series model; see e.g. chapter 3 of Teräsvirta et al. (2010).

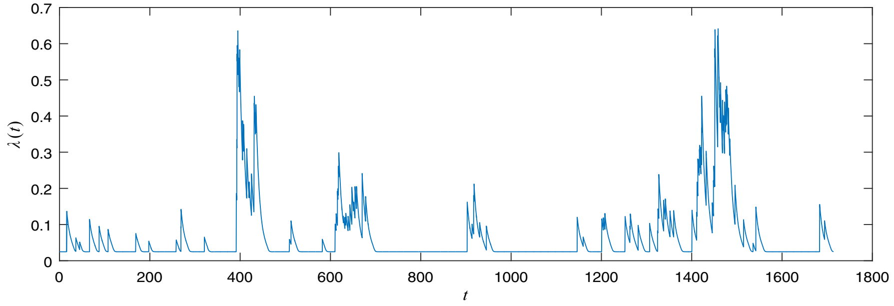

Fig. 1 A simulated trajectory of the non-linear intensity model in Example 1, where $\lambda_0 = 0.0248$, $\alpha = 0.1233$, $\beta = 0.0399$, $\gamma = 1.6259 \times 10^3$ and $\delta = 1.6758$, and the jumps are simulated from an inverse Gaussian distribution with parameters 1.9389 (mean) and 5.4943 (shape). The parameter values are inspired by estimates in Sect. 4

3 Markovian dynamics

In this section, we introduce the concept of a Markov generator, which is a useful tool for analysing the behaviour of the Markovian intensity process (3), in particular we shall use the tools developed here to study the stability of (3). Note that discrete Markov chain techniques have been used to study stability of nonlinear time series models in Tjøstheim (1990) and Fokianos and Tjøstheim (2012).

We denote by $\mathbb{P}_\lambda$ the probability measure induced by the transition function of the Markov process $\lambda(t)$ at time $t \geq 0$, with $\lambda(t) = \lambda$, that is, given $\lambda \geq \lambda_0$, $s \geq 0$ and $B \in \mathcal{B}([\lambda_0, \infty))$, let $\mathbb{P}_\lambda(\lambda(t+s) \in B) := \mathbb{P}(\lambda(t+s) \in B | \lambda(t) = \lambda)$, where $\mathbb{E}_\lambda$, furthermore, denotes the corresponding expected value operator. Associated to the Markov intensity process $\lambda(t)$ is its *strong generator*, which is defined as the derivative of the semigroup $P_t f(\lambda) := \mathbb{E}_\lambda[f(\lambda(t))]$ under the sup-norm, $\|f\| = \sup_{\lambda \geq \lambda_0} |f(\lambda)|$. Moreover, if we denote the set of bounded measurable functions on $[\lambda_0, \infty)$ such that the limit

$$\hat{\mathcal{A}} f := \lim_{t \downarrow 0} \frac{P_t f - f}{t}$$

exists (under the sup-norm) by $\mathcal{D}(\hat{\mathcal{A}})$, then it can be shown that strong continuity of $\{P_t\}$ (continuity of $t \mapsto P_t$ under $\|\cdot\|$) implies that $\mathcal{D}(\hat{\mathcal{A}})$ is dense in the space of bounded measurable functions $f : [\lambda_0, \infty) \to \mathbb{R}$. However, even if this is true, it is usually not the whole space. The following result follows by Proposition 14.13, p.31, in Davis (1993).

Proposition 1 *For $f \in \mathcal{D}(\hat{\mathcal{A}})$, and $t \geq 0$, let*

$$C_f(t) := f(\lambda(t)) - f(\lambda(0)) - \int_0^t \hat{\mathcal{A}} f(\lambda(s)) \mathrm{d}s.$$

Then for any $\lambda \geq \lambda_0$, the process $\{C_f(t)\}_{t \geq 0}$ is a martingale on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P}_\lambda)$, where $\mathcal{F}_t := \sigma\{\lambda(s) : s \leq t\}$.

It follows that

$$\mathbb{E}_\lambda[f(\lambda(t))|\mathcal{F}_s] = f(\lambda(s)) + \mathbb{E}_\lambda\left[\int_s^t \hat{\mathcal{A}} f(\lambda(r))\mathrm{d}r|\mathcal{F}_s\right]$$

holds for any $\lambda \geq \lambda_0$, $0 \leq s \leq t$ and $f \in \mathcal{D}(\hat{\mathcal{A}})$, and in particular, by letting $s = 0$, the process $t \mapsto f(\lambda(t))$ verifies the so-called *Dynkin formula*,

$$\mathbb{E}_\lambda[f(\lambda(t))] = f(\lambda) + \mathbb{E}_\lambda\left[\int_0^t \hat{\mathcal{A}} f(\lambda(r))\mathrm{d}r\right],$$

for any $\lambda \geq \lambda_0$, $t \geq 0$ and $f \in \mathcal{D}(\hat{\mathcal{A}})$. However, as we have already observed, the domain $\mathcal{D}(\hat{\mathcal{A}})$ is usually not the entire space of bounded measurable functions on $[\lambda_0, \infty)$. Thus, following Davis (1993) we, therefore, define $\mathcal{D}(\mathcal{A})$ to be the set of measurable functions f such that there exists a measurable function ψ for which $t \mapsto \psi(\lambda(t))$ is integrable $\mathbb{P}_\lambda$-a.s. for all $\lambda \geq \lambda_0$ and the process

$$C_f(t) = f(\lambda(t)) - f(\lambda(0)) - \int_0^t \psi(\lambda(s))\mathrm{d}s$$

is a local martingale. We, moreover, write $\psi = \mathcal{A} f$, and call $(\mathcal{A}, \mathcal{D}(\mathcal{A}))$ the *extended generator* of the process $\lambda(t)$. Note in particular that $\mathcal{D}(\hat{\mathcal{A}}) \subset \mathcal{D}(\mathcal{A})$.

With these definitions in mind we derive the form of the extended generator, and describe its domain in Proposition 2. Recall that we denote by $\nu(\lambda(T_k-), \cdot)$ the non-negative jump size distribution of X_k, given the value of $\lambda(T_K-)$. The strong generator is given by

$$(\mathcal{A} f)(\lambda) = \mu(\lambda) f'(\lambda) + \lambda \int (f(\lambda + \beta x) - f(\lambda))\, \nu(\lambda, \mathrm{d}x). \tag{5}$$

Proposition 2 *If for any $t > 0$ and a given measurable function, $f : [\lambda_0, \infty) \to \mathbb{R}$, it holds that the map*

$$\lambda \mapsto \int f(\lambda + x)\nu(\lambda, \mathrm{d}x)$$

is measurable, and

$$\mathbb{E}_\lambda\left[\int_0^t \left|\lambda(s) \int (f(\lambda(s) + \beta x) - f(\lambda(s)))\, \nu(\lambda(s), \mathrm{d}x)\right| \mathrm{d}s\right] < \infty, \tag{6}$$

then f is in the domain of the extended generator of $\lambda(t)$, $f \in \mathcal{D}(\mathcal{A})$, where $\mathcal{A}$ is given by (5).

Proof By Itô's lemma (see Theorem II.31 in Protter 2005) it holds that

$$f(\lambda(t)) - f(\lambda(0)) = \int_0^t f'(\lambda(s))\mu(\lambda(s))\mathrm{d}s + \sum_{0<s\le t} \{f(\lambda(s)) - f(\lambda(s-))\}.$$

Now, using that $N(t) - \Lambda(t)$ is a martingale,

$$\begin{aligned}
&\mathbb{E}_\lambda\left[\sum_{0<s\le t}\{f(\lambda(s)) - f(\lambda(s-))\}\right] \\
&= \mathbb{E}_\lambda\left[\int_0^t \int \{f(\lambda(s-) + \beta x) - f(\lambda(s-)\}\nu(\lambda(s-), \mathrm{d}x)\mathrm{d}N(s)\right] \\
&= \mathbb{E}_\lambda\left[\int_0^t \int \{f(\lambda(s-) + \beta x) - f(\lambda(s-)\}\nu(\lambda(s-), \mathrm{d}x)\lambda(s)\mathrm{d}s\right] \\
&= \mathbb{E}_\lambda\left[\int_0^t \lambda(s) \int \{f(\lambda(s) + \beta x) - f(\lambda(s)\}\nu(\lambda(s), \mathrm{d}x)\mathrm{d}s\right],
\end{aligned}$$

where we have exploited the fact that

$$s \mapsto \int \{f(\lambda(s-) + \beta x) - f(\lambda(s-))\}\,\nu(\lambda(s-), \mathrm{d}x),$$

is predictable, and that a stochastic integral of a predictable process with respect to a martingale is a martingale. It follows that for a function f which fulfills (6), the process

$$t \mapsto f(\lambda(t)) - f(\lambda(0)) - \int_0^t \mathcal{A}f(\lambda(s))\mathrm{d}s$$

is a zero-mean martingale, and thus the proof is completed using the definition of the extended generator. □

Having identified the extended generator of our class of processes, we proceed to use it to analyse some of the class properties. To that end, we first of all notice that our class of self-exciting processes is a piecewise deterministic process (PDP) in the sense of Davis (1993). We adopt the following regularity conditions.

Assumption 2 It holds that

(i) $\lambda \mapsto \mu(\lambda)$ is Lipschitz continuous.
(ii) $\nu : [\lambda_0, \infty) \to \mathcal{P}(\mathbb{R})$ (the set of probability measures on $\mathbb{R}$) is a measurable function such that $\nu(\lambda, \{\lambda\}) = 0$ for all $\lambda \ge \lambda_0$.
(iii) The map $\lambda \mapsto \int_{\lambda_0}^\infty f(x)\nu(\lambda, \mathrm{d}x)$ is continuous for continuous and bounded f.

Together with the non-explosion assumption of Sect. 2, the first two of the above assumptions are the so-called "standard conditions" of Davis (1993), which ensure a certain regularity structure on the class of PDP processes. The first one of these concerns the deterministic μ function, which governs the behaviour of the intensity function between jumps. By requiring Lipschitz continuity we exclude explosions and ensure that the process behaves like a deterministic non-explosive Markov process between jumps. The second condition states the measurability of the family of jumps-size distributions, and that we can almost surely detect jumps. The above assumption moreover ensures that our process class is a so-called Borel right process (see Theorem 27.8 in Davis 1993). Finally, the continuity assumption of point three ensures together with the non-explosive property that our process class fulfills the Feller property, i.e. that the map $x \mapsto P_t f(x)$ is bounded and continuous if f is bounded and continuous for $t \geq 0$.

Recall that the Markov process $\lambda(t)$ is said to be *ϕ-irreducible* if ϕ is σ-finite and

$$\mathbb{E}_\lambda \left[\int_0^\infty 1_{\{\lambda(t) \in A\}} \mathrm{d}t \right] > 0$$

whenever $\phi(A) > 0$, for all $\lambda \geq \lambda_0$. The following stability result employs the form of the generator to analyse the stability properties of the intensity process. It turns out that under certain assumptions on the generator the intensity process (3) is asymptotically stable. Given a signed measure on $\mathcal{B}([\lambda_0, \infty))$ and $f \geq 1$, write $\|\mu\|_f := \sup_{|g| \leq f} \left| \int g \mathrm{d}\mu \right|$.

Theorem 1 *Suppose that $\lambda(t)$ is ϕ-irreducible where ϕ is supported on a set with non-empty interior, and there exist constants $c > 0$ and $d \in \mathbb{R}$ such that*

$$\mathcal{A}\lambda = \mu(\lambda) + \lambda \int x \nu(\lambda, \mathrm{d}x) \leq -c\lambda + d \tag{7}$$

for all $\lambda > \lambda_0$, then an essentially unique finite invariant measure, π, exists and $\lambda(t)$ is moreover geometrically ergodic, i.e. there exist $\beta < 1$, $B < \infty$ such that

$$\| P_t(\lambda, \cdot) - \pi \|_f \leq B f(\lambda) \beta^t,$$

where $P_t(\lambda, \cdot) = \mathbb{P}_\lambda(\lambda(t) \in \cdot)$ and $f(\lambda) = 1 + \lambda$.

Proof It follows by the Feller property, the ϕ-irreducibility property (which also holds for the sampled chain) and Theorem 3.4 in Meyn and Tweedie (1992) that all compact subsets of a skeleton chain are petite. So the result follows directly from Theorem 6.1 in Meyn and Tweedie (1993). □

We note that the condition (7) makes statistical inference and maximum likelihood theory possible. See Fokianos et al. (2009) for corresponding conditions in the much simpler context of integer time series modelling with stochastic intensity.

Example 2 In the case of linear drift, $\mu(\lambda) = \alpha(\lambda_0 - \lambda)$, we find that

$$\mathcal{A}\lambda = \left(\beta \int x\nu(\lambda, \mathrm{d}x) - \alpha\right)\lambda + \alpha\lambda_0,$$

so the conditions of Theorem 1 are fulfilled if there exists a constant $K \geq \lambda_0$ such that

$$\alpha > \beta \int x\nu(\lambda, \mathrm{d}x) \tag{8}$$

for all $\lambda > K$. What this means in practice is that in the case of a linear drift, the speed of mean reversion, that is the parameter $\alpha > 0$, must be larger than the expected value of the jump-size distribution, for all values of $\lambda \geq K$. In other words, the rate at which the intensity process is forced towards λ_0, between jumps, is larger than the expected value of the jump-size distribution for all λ outside a bounded interval, which is intuitively reasonable. We finally note that the condition (8) is also sufficient to ensure the stability of the process introduced in Example 1, because $\alpha + \delta\mathrm{e}^{-\gamma\lambda^2}$ tends to α as $\lambda \to \infty$.

4 Modelling jumps of UK power data by means of a self-exciting jump process

Having introduced and analysed a class of self-exciting processes, we apply it to model electricity and energy prices, more precisely we employ a self-exciting process to model the jumps of UK electricity spot price data.

The opening of electricity, and other energy commodity markets, worldwide during the last two decades or so prompted the need for new stochastic models to be developed to model the idiosyncratic features that such markets display. Indeed, electricity and commodity markets are known to display quite distinct features that are rarely observed in more traditional stock markets, and are thus very challenging to model. These features include sudden jumps of many magnitudes, due to a sudden unforeseen shortage of energy supply, followed by an equally steep mean-reversion once the supply has matched the demand, along with the occasional appearance of negative prices due to overproduction. Another interesting feature of such markets is the clustering of jumps, i.e. the apparent observation that jumps are more likely to appear in clusters than as a realization of a Poisson process. In other words, the appearance of jumps seems to be governed by a stochastic intensity process, which is self-exciting.

Due to the relatively frequent number of jumps in the aforementioned markets, a modelling paradigm in which (deseasonalised) spot prices are modelled as the sum of a continuous diffusion part, and a discontinuous jump part has been developed see e.g. Meyer-Brandis and Tankov (2008), Meyer-Brandis and Morgan (2014), and the threshold model of Geman and Roncoroni (2006). This is something which we do in our setting as well. We study a UK (APX) power price series, with a single observation per day, in the period from February 6, 2001 to December 31, 2007 (excluding weekends and holidays). The time series was kindly provided by Montel. The reason we choose

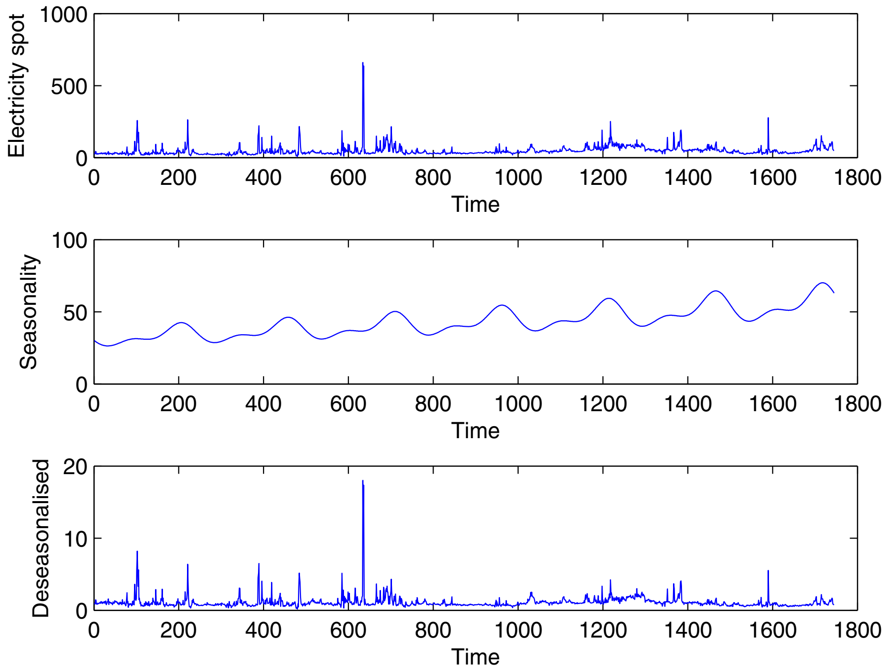

Fig. 2 *Top* daily quotes of UK power data. *Middle* the fitted seasonality function. *Bottom* the deseasonalised UK power data

to analyse this time period of UK power data, is to compare it with an analysis of the same time period performed by Meyer-Brandis and Morgan (2014), in which the authors extract a jump trajectory from the deseasonalised data and fit a compound Poisson driven Ornstein–Uhlenbeck process to the jump trajectory. In particular this means that the inter-arrival times of jumps are memoryless, since they are determined by a Poisson process, which excludes the appearance of clusters. We remedy this problem by introducing a self-exciting process to model the jump part of the model.

We shall employ the similar steps as Meyer-Brandis and Morgan (2014) do with the exception that we fit a self-exciting process to the jump trajectory of the process.

4.1 Removing seasonality and extracting a jump trajectory

The UK spot data is fitted to the seasonality function

$$S(t) = \exp\left(a + bt + \sum_{k=1}^{2}\left(c_{1k}\sin\left(\frac{2k\pi}{252}t\right) + c_{2k}\cos\left(\frac{2k\pi}{252}t\right)\right)\right). \tag{9}$$

The parameter estimates are $\hat{a} = 3.447$, $\hat{b} = 0.0003304$, $\hat{c}_{11} = -0.1535$, $\hat{c}_{21} = 0.00983$, $\hat{c}_{21} = -0.08266$, $\hat{c}_{22} = -0.04492$, and the data, together with the seasonality function, and the deseasonalised process, which is obtained by dividing the

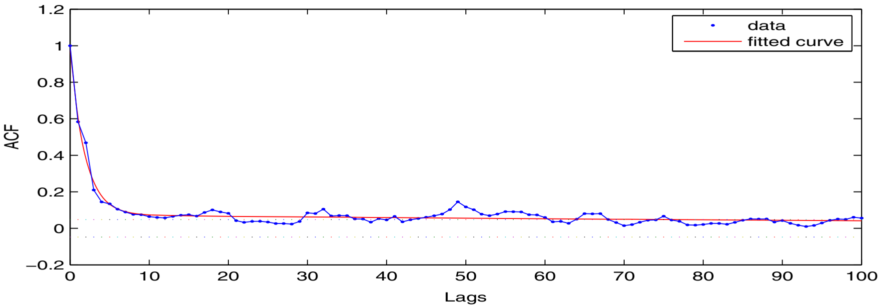

Fig. 3 The empirical, and estimated autocorrelation function of the deseasonalised data

observed data with $S(t)$, is displayed in Fig. 2. The jump trajectory is extracted from the deseasonalised time series using a hard thresholding method described by Meyer-Brandis and Tankov (2008). Note the difference in scale in the upper and lower panels of Fig. 2. The seasonality function in the middle panel clearly displays an upwards trend and seasonal variation independent of jumps. We conjecture that the deseasonalised time series is the sum of two mean reverting Ornstein–Uhlenbeck (OU) processes, the first one driven by a Wiener process, and the second one driven by a self-exciting process. This statement is supported by the empirical autocorrelation function, which is displayed in Fig. 3, to which we have fitted the curve

$$h \mapsto w_1 \mathrm{e}^{-\rho_1 h} + w_2 \mathrm{e}^{-\rho_2 h},$$

where $w_1, w_2, \rho_1, \rho_2 > 0$, with the estimates $w_1 = 0.9267, w_2 = 0.07394, \rho_1 = 0.5459, \rho_2 = 0.005725$. That is, we conjecture that the higher speed of mean reversion corresponds to the jump part of the process, while the smaller speed of mean reversion corresponds to the Wiener process driven OU part of the process. Thus the deseasonalised price is the sum of two OU processes with, the first one driven by a Wiener process with a low speed of mean reversion (0.005725), and the second one driven by a self-exciting jump process with a high speed of mean reversion (0.5459). The rationale behind associating the higher speed of mean reversion to the self-exciting process driven OU process is that a big jump in prices is usually caused by something unusual or unpredicted in the market (e.g. a sudden closure of a power plant), which the market usually corrects for in a reasonably short time (e.g. by increasing power production elsewhere). By contrast the speed of mean reversion associated to the Wiener process driven OU process is much lower as its effects decay much slower.

We assume that the jump price trajectory has dynamics governed by

$$\mathrm{d}Y(t) = -\rho Y(t)\mathrm{d}t + \mathrm{d}U(t), \tag{10}$$

where $\rho = 0.5459$, the jump component is estimated from the data, and has an intensity which we will specify below. We extract the jump process trajectory using the hard thresholding method described by Meyer-Brandis and Tankov (2008). Taking out the 120 largest jumps from the deseasonalised time series yields the series displayed in

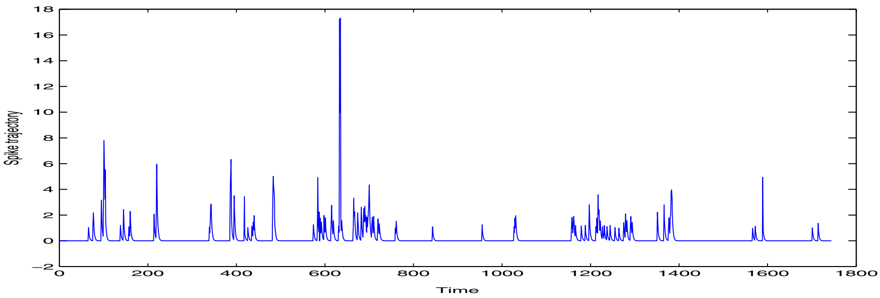

Fig. 4 The jump trajectory with 120 jumps extracted from the data

Fig. 4. Note that the method we use to extract the largest jumps removes the largest jumps according to their size in absolute in value, thus there are a few negative jumps in the series (7 out of 120). Having extracted this jump trajectory one may apply a variety of self-exciting jump processes to it. We proceed to discuss some possibilities in this direction in the next subsection.

4.2 Fitting a self-exciting model to the jump trajectory

We consider simultaneous estimation of the model parameters via maximum likelihood estimation. It is well known (see e.g. Ogata 1978 and Daley and Vere-Jones 1988) that the likelihood of the counting process (1) on the interval $[0, t]$ is given by

$$L_t(\theta) = \prod_{T_k \le t} \lambda(T_k) \exp\left(- \int_0^t \lambda(s)\mathrm{d}s \right),$$

where $\{T_k\}$ is the jump-time sequence, and $\theta \in \mathbb{R}^d$, for some $d \ge 1$ is the parameter vector which specifies the model. The corresponding log-likelihood is given by

$$\log L_t(\theta) = \sum_{T_k \le t} \log \lambda(T_k) - \int_0^t \lambda(s)\mathrm{d}s.$$

The asymptotic properties of the maximum likelihood estimator associated with the above likelihood are described by Ogata (1978). The author gives extensive conditions under which the maximum likelihood estimator, $\hat{\theta}_t$, satisfies

$$\sqrt{t}(\hat{\theta}_t - \theta) \overset{\mathcal{D}}{\to} N(0, I(\theta)^{-1}),$$

where

$$I(\theta) = \left\{ \mathbb{E}_\pi \left[\frac{1}{\lambda(t)} \frac{\partial \lambda(t)}{\partial \theta_i} \frac{\partial \lambda(t)}{\partial \theta_j} \right] \right\}_{i,j=1}^d$$

as $t \to \infty$, and π denotes the invariant measure of the model.

Recall that in our setting the intensity process follows SDE dynamics on the form (3), where the jump process $U(t)$ records stochastic jumps which may depend on the current value of the intensity process. For the purpose of constructing the likelihood function note that the intensity may be written as

$$\lambda(t) = \lambda_0 + \int_0^t \mu(\lambda(s))\mathrm{d}s + \beta \sum_{k=1}^{N(t)} X_k.$$

So for a given equidistant grid $t_0 < t_1 < \cdots < t_N$, with step size $\Delta > 0$, and given the observed jumps $\{X_k\}$, the intensity can be estimated recursively by setting $\lambda(0) = \lambda_0$ and then applying the recursive step

$$\begin{aligned}\lambda(t_n) &= \lambda_0 + \sum_{k=1}^{n} \mu(\lambda(t_{k-1}))\Delta + \beta \sum_{k=1}^{N(t_n)} X_k \\ &= \lambda(t_{n-1}) + \mu(\lambda(t_{n-1}))\Delta + \beta \Delta U(t_n)\end{aligned}$$

for $n = 1, \ldots, N$, where $\Delta U(t_n) = \sum_{k=N(t_{n-1})+1}^{N(t_n)} X_k$ if some jumps occur on $(t_{n-1}, t_n]$, and $\Delta U(t_n) = 0$ otherwise.

Now we proceed to discussing different variants of the intensity model which we will fit to the jump data extracted in the previous subsection. First, we consider a linear model on the form

$$\mathrm{d}\lambda_L(t) = \alpha(\lambda_0 - \lambda_L(t))\mathrm{d}t + \beta \mathrm{d}U(t) \tag{11}$$

to the data, where $\theta = (\lambda_0, \alpha, \beta)$ is the parameter vector to be estimated. Then we augment the linear model and consider

$$\mathrm{d}\lambda_{NL}(t) = (\alpha + \delta \exp(-\gamma \lambda_{NL}(t)^2))(\lambda_0 - \lambda_{NL}(t))\mathrm{d}t + \beta \mathrm{d}U(t), \tag{12}$$

where $\theta = (\lambda_0, \alpha, \beta, \gamma, \delta)$ is the parameter vector to be estimated. For the purpose of maximum likelihood estimation we propose two alternatives. One can first of all feed both positive and negative jumps into the maximum likelihood estimation, and thus inspect the total impact of jumps on the intensity, irrespective of whether they are positive or negative. On the other hand, one can simply disregard the 7 negative jumps and use only the positive jumps.

In Table 1, using both of the approaches, we compare the maximum likelihood estimates of the linear and non-linear models defined in (11) and (12) respectively. Judging from the maximum likelihood estimates it seems that the non-linear model does not perform much better than the linear model. Indeed, if one inserts all of the jumps into the maximum likelihood estimation, then $\hat{\delta} = 5.2062 \times 10^{-8}$ in the non-linear model (which makes $\hat{\gamma} = 0.2264$ redundant) and $\hat{\lambda}_0$, $\hat{\alpha}$ and $\hat{\beta}$ are identical to the corresponding estimates for the linear model, which means that one gains nothing from the non-linear model. While, if one only inserts the positive jumps into the estimation

Table 1 Maximum likelihood estimates of the linear model (11) and the non-linear model (12)

Model and jump input	$\hat{\lambda}_0$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\delta}$
Linear model, all jumps	0.0246	0.1637	0.0641	–	–
Linear model, positive jumps	0.0248	0.1233	0.0399	–	–
Non-linear model, all jumps	0.0246	0.1637	0.0641	0.2264	5.2062×10^{-8}
Non-linear model, positive jumps	0.0249	0.1219	0.0399	0.0880	0.0017

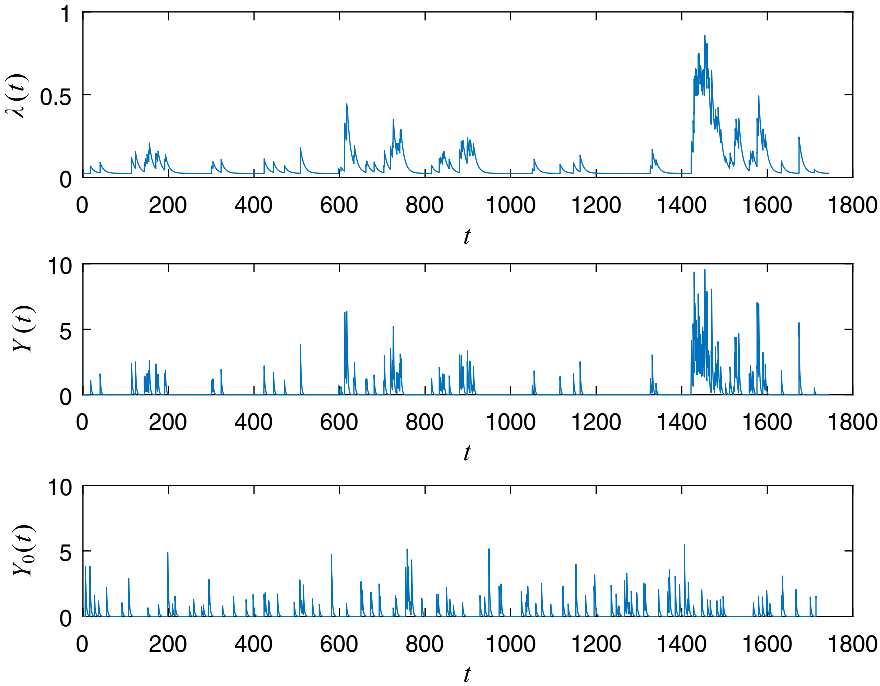

Fig. 5 *Above* a realization of the linear intensity process (11), simulated with $\hat{\lambda}_0 = 0.0232, \hat{\alpha} = 0.1181, \hat{\beta} = 0.0392$, and inverse Gaussian jump size distribution. *Middle* the corresponding jump price trajectory determined by (10), with the same β, and jump-size distribution. *Bottom* a jump price trajectory simulated from a constant intensity estimated from the same data set

it can be argued that the non-linear model explains a little bit more than the linear model, although, we maintain that in this particular case the gain is negligible.

Having observed that the non-linear model does not fit the data better than the linear model for the dataset at hand we proceed to simulate the linear model to display some of its features and compare it to a constant intensity process. To that end we fit an inverse Gaussian distribution to the positive jumps with parameter estimates 1.9389 (mean) and 5.4943 (shape), and simulate the model by means of employing the form of the linear intensity, alternatively one could employ a thinning algorithm, see Ogata (1981), to simulate the process. We maintain that the simulated spike trajectory in the middle panel of Fig. 5 and the extracted jump trajectory displayed in Fig. 4 have comparable characteristics. Indeed, we observe that both figures have clustering of jumps, and relatively calm periods in between, which in turn correspond to the observed changes in the latent intensity process. By contrast, a constant jump intensity process, such as the one Meyer-Brandis and Morgan (2014), cannot be expected to reproduce this clustering behaviour, as is apparent from the simulated trajectory in the bottom panel of Fig. 5. By which we mean that a constant intensity process is liable to underestimate the jump clustering effects which is present in the data, since after all a constant intensity process does not produce clusters. This in turn means that under a constant intensity regime the price dynamics do not include any clustering effects. One might argue that the effects of clustering do not matter, since the presence of jumps will simply smear out the effects of the clusters, but the evidence suggest that they do matter, at least in the short run. To see why let us study the mean behaviour of the self-exciting intensity process. Let $f(\lambda) = \lambda$, then if we apply the extended generator to f it follows that

$$(\mathcal{A}f)(\lambda) = \alpha(\lambda_0 - \lambda) + \lambda\beta\mathbb{E}[Y].$$

So letting $m(t) := \mathbb{E}_\lambda[\lambda(t)]$, where $\lambda \geq \lambda_0$ denotes the current value of the intensity process, an application of Dynkin's formula and Fubini yields

$$m(t) = \lambda + \int_0^t (\alpha\lambda_0 + \rho m(r))\,\mathrm{d}r,$$

where $\rho := \beta\mathbb{E}[Y] - \alpha$. This implies that $m(t)$ solves the ODE

$$m'(t) - \rho m(t) = \alpha\lambda_0,$$

with the initial value $m(0) = \lambda$, hence

$$m(t) = \left(\frac{\alpha\lambda_0}{\rho} + \lambda\right)\mathrm{e}^{\rho t} - \frac{\alpha\lambda_0}{\rho}. \tag{13}$$

Thus notice that if $\rho < 0$ ($\rho = -0.0457$ in our case), it follows that

$$\lim_{t\to\infty} m(t) = -\frac{\alpha\lambda_0}{\rho}, \tag{14}$$

irrespective of the initial condition $\lambda \geq \lambda_0$, whereas $\rho < 0$, and $\lambda = -\alpha\lambda_0/\rho$ guarantees constant $m(t) = -\alpha\lambda_0/\rho$, for all t. We remark that one can do similar calculations to derive the form of higher moments. Now, once we have estimated the parameters of the intensity process $t \mapsto \lambda(t)$ we can essentially observe it, by observing the jumps which occur. Which in turn means that we can make predictions about the short term behaviour of the intensity process. Thus, in particular, for the model which we have estimated in the current section this means that the dynamics of the model in the near future are highly influenced by the current state of the model. If we observe that we are currently in a calm period with $\lambda(t)$ close to or equal to the base intensity, λ_0, then we expect the intensity to mean revert to its stationary mean (14).

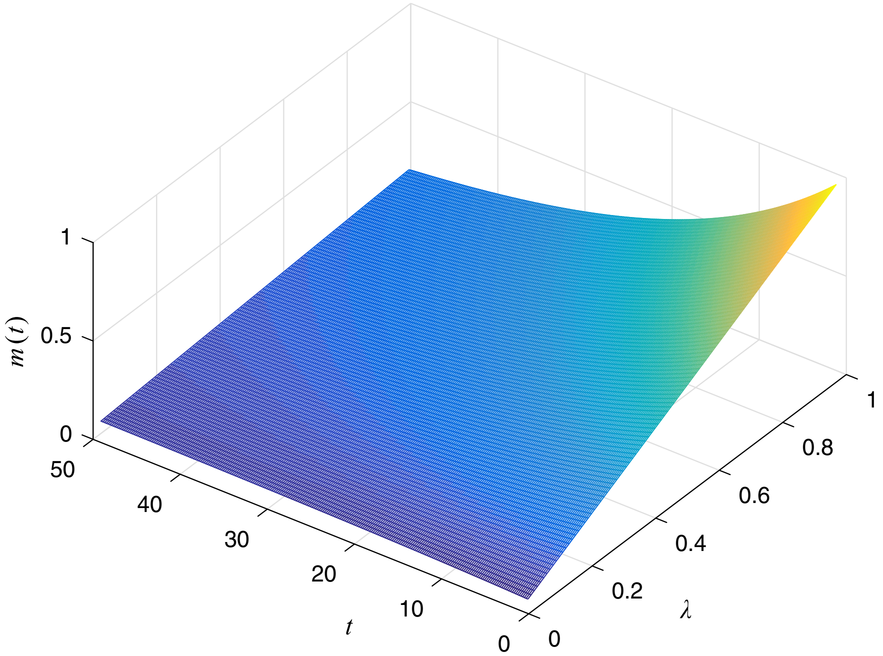

Fig. 6 The mean function (13) with the parameter estimates in the current section and varying initial positions $\lambda \geq \lambda_0$

Likewise, if we observe that we are currently going through a high intensity regime, then we will expect the intensity to mean revert down to its stationary mean (14). For the long run, however, information on the current state of the model is irrelevant, as all we can say is that the model will mean revert to the stationary level (14). The stationary mean is close to being equivalent to a constant intensity model in the sense that if one calculates the value of the stationary mean (14) by replacing the parameters with their corresponding estimates, then the value one obtains is 0.0667, which is close to the constant intensity estimate given by the total number of jumps divided by the length of the period, which is 0.0659. This behaviour is observed in Fig. 6, and it is relevant because it means that predictions based on this model will depend on the current value of the intensity process, which as we have argued fits the data better than a constant intensity process. Thus, we remark that the estimated model can be employed to make forecast, or to price financial derivatives, either by means of Monte Carlo simulations or exact calculations such as Prigent (2001) discusses. We shall, however, not dwell further into that domain in the current paper.

We close the section with the following remark on the dataset. Note that the jump-trajectory that we have extracted from our data using the above methods lives on a lattice, by which we mean that it can only jump at the beginning of each day, and only once per day, which is admittedly somewhat unrealistic from a point process perspective, since the inter-arrival times of point processes are real numbers. This fact is, however, not a big concern in our setting, as it has a relatively small effect on

Table 2 Maximum likelihood estimates given arrival times, and perturbed arrival times

Arrival times	$\hat{\lambda}_0$	$\hat{\alpha}$	$\hat{\beta}$
$\{T_k\}$	0.0248	0.1233	0.0399
$\{\tilde{T}_k\}$	0.0241	0.1254	0.0408
$\{\check{T}_k\}$	0.0250	0.1516	0.0482

the maximum likelihood estimation of the model, as we observe below. To alleviate potential concerns caused by this, we study the robustness of the maximum likelihood estimator of the linear model under inter-arrival time perturbation. To be more precise we investigate the sensitivity of the maximum likelihood estimators to perturbation of the arrival times $\{T_k\}$. To that end we introduce arrival time perturbations, under which we estimate the model parameters. We consider two types of arrival time perturbations.

1. The arrival times $\{T_k\}$ are perturbed by an iid family, $\{U_k\}$, of random variables which are uniformly distributed on $[-1/2, 1/2]$, yielding the perturbed arrival times $\tilde{T}_k := T_k + U_k$.
2. Initialize the arrival time sequence by first letting it be equal to the original arrival time sequence, $\{\check{T}_k\} := \{T_k\}$, and then for all $k = 1, \ldots, 120$ perturb the remaining $121 - k$ arrival times, i.e. let $\{\check{T}_l\}_{l \geq k} := \{\check{T}_l\}_{l \geq k} + U_k$, where $\{U_k\}$, is an iid family of random variables which are uniformly distributed on $[-1/2, 1/2]$.

The original parameter estimates are displayed with the parameters estimated from the perturbed data in Table 2, they confirm that the perturbed data yield similar parameter estimates.

5 Conclusion

In this paper, we have briefly described self-exciting jump processes, discussed their applicability in different settings, and introduced a general class of self-exciting processes with intensity processes that admit an SDE representation. We have, moreover, applied Markov theory to identify conditions which ensure the stability and asymptotic stationarity of our model class. Finally, we have discussed maximum likelihood estimation with jump trajectory data extracted from UK power markets. The framework which we have presented here can be extended in different directions. First of all by adding time dependence and secondly by extended the framework to multiple dimensions. We plan on exploring these directions along with some applications in a separate article.

References

Aït-Sahalia, Y., Cacho-Diaz, J., Laeven, R. J. (2015). Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, *117*(3), 585–606.

Bacry, E., Muzy, J. F. (2014). Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance*, *14*(7), 1147–1166.

Bacry, E., Delattre, S., Hoffmann, M., Muzy, J. (2013). Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes and their Applications*, *123*(7), 2475–2499 **(a Special Issue on the Occasion of the 2013 International Year of Statistics)**.

Benth, F. E., Šaltyte Benth, J., Koekebakker, S. (2008). *Stochastic modelling of electricity and related markets, Advanced Series on Statistical Science & Applied Probability* (Vol. 11). Hackensack, New Jersey: World Scientific Publishing Co. Pte Ltd.

Brémaud, P. (1981). *Point processes and queues*, Martingale dynamics, Springer Series in Statistics. New York, Berlin: Springer.

Cont, R., Tankov, P. (2004). *Financial modelling with jump processes. Chapman & Hall/CRC Financial Mathematics Series*. Boca Raton, Florida: Chapman & Hall/CRC.

Daley, D. J., Vere-Jones, D. (1988). *An introduction to the theory of point processes*, Springer Series in Statistics. New York: Springer.

Davis, M. H. A. (1993). *Markov models and optimization, Monographs on Statistics and Applied Probability* (Vol. 49). London: Chapman & Hall.

Embrechts, P., Liniger, T., Lin, L. (2011). Multivariate Hawkes processes. An application to financial data. *Journal of Applied Probability*, *48A*, 367–378.

Errais, E., Giesecke, K., Goldberg, L. R. (2010). Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics*, *1*(1), 642–665.

Fokianos, K., Tjøstheim, D. (2012). Nonlinear Poisson autoregression. *Annals of the Institute of Statistical Mathematics*, *64*(6), 1205–1225.

Fokianos, K., Rahbek, A., Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association*, *104*(488), 1430–1439.

Geman, H., Roncoroni, A. (2006). Understanding the fine structure of electricity prices. *The Journal of Business*, *79*(3), 1225–1261.

Hawkes, A. G. (1971a). Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society Series B Methodological*, *33*, 438–443.

Hawkes, A. G. (1971b). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, *58*, 83–90.

Hawkes, A. G., Oakes, D. (1974). A cluster process representation of a self-exciting process. *Journal of Applied Probability*, *11*, 493–503.

Jacod, J. (1974/1975). Multivariate point processes: predictable projection, Radon-Nikodým derivatives, representation of martingales. *Z Wahrscheinlichkeitstheorie und Verw Gebiete*, *31*, 235–253.

Jacod, J., Shiryaev, A. N. (2003). *Limit theorems for stochastic processes, Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*], (2nd ed., Vol. 288). Berlin: Springer.

Jaisson, T., Rosenbaum, M. (2015). Limit theorems for nearly unstable Hawkes processes. *The Annals of Applied Probability*, *25*(2), 600–631.

Meyer-Brandis, T., Morgan, M. (2014). A dynamic Lévy copula model for the spark spread. In F. E. Benth, V. A. Kholodnyi, P. Laurence (Eds.), *Quantitative Energy Finance* (pp. 237–257). New York: Springer.

Meyer-Brandis, T., Tankov, P. (2008). Multi-factor jump-diffusion models of electricity prices. *International Journal of Theoretical and Applied Finance*, *11*(05), 503–528.

Meyn, S. P., Tweedie, R. L. (1992). Stability of Markovian processes I: Criteria for discrete-time chains. *Advances in Applied Probability*, *24*(3), 542–574.

Meyn, S. P., Tweedie, R. L. (1993). Stability of Markovian processes III: Foster–Lyapunov criteria for continuous-time processes. *Advances in Applied Probability*, *25*(3), 518–548.

Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, *30*(1), 243–261.

Ogata, Y. (1981). On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, *27*(1), 23–31.

Prigent, J. L. (2001). Option pricing with a general marked point process. *Mathematics of Operations Research*, *26*(1), 50–66.

Protter, P. E. (2005). *Stochastic integration and differential equations, Stochastic Modelling and Applied Probability* (2nd ed., Vol. 21). Berlin: Springer. Version 2.1, Corrected third printing.

Teräsvirta, T., Tjøstheim, D., Granger, C. W. J. (2010). *Modelling nonlinear economic time series*. Oxford: Oxford University Press.

Tjøstheim, D. (1990). Non-linear time series and Markov chains. *Advances in Applied Probability*, *22*(3), 587–611.