

# A Self-Exciting Switching Jump Diffusion: properties, calibration and hitting time.

Donatien Hainaut\*

*ISBA, Université Catholique de Louvain, Belgium*

Griselda Deelstra†

*Department of Mathematics, Université libre de Bruxelles, Belgium*

May 8, 2018

## Abstract

A way to model the clustering of jumps in asset prices consists in combining a diffusion process with a jump Hawkes process in the dynamics of the asset prices. This article proposes a new alternative model based on regime switching processes, referred to as a self-exciting switching jump diffusion (SESJD) model. In this model, jumps in the asset prices are synchronized with changes of states of a hidden Markov chain. The matrix of transition probabilities of this chain is designed in order to approximate the dynamics of a Hawkes process. This model presents several advantages compared to other jump clustering models. Firstly, the SESJD model is easy to fit to time series since estimation can be performed with an enhanced Hamilton's filter. Secondly, the model explains various forms of option volatility smiles. Thirdly, several properties about the hitting times of the SESJD model can be inferred by using a fluid embedding technique, which leads to closed form expressions for some financial derivatives, like perpetual binary options.

KEYWORDS: self-exciting process, switching process, jump diffusion

JEL Codes: G10, G13.

## 1 Introduction

A careful observation of financial time series reveals that jumps of prices arrive grouped and are usually triggered by a flood of news. This relation is studied by Rangel (2011) who demonstrates the causal link between the economic announcements and the clustering of jumps in stock markets. Ficura (2015) observes the same phenomenon when studying high frequency data of four major foreign exchange rates. The existence of jump clustering has important implications in many areas of finance and insurance. Fulop et al. (2015) emphasize the role of jumps in the valuation of options. Hainaut and Moraux (2018) underline that the efficiency of quadratic hedging strategies is considerably reduced in presence of jump clustering.

A way to deal with the clustering of jumps is to decompose the overall price variability into two components - a continuous Brownian process and a discontinuous self-exciting jump process, called Hawkes process<sup>1</sup>. For instance, Chen and Poon (2013), Boswijk et al. (2015), and Carr and Wu (2016) investigate some self-exciting jump diffusion processes for modeling stock index returns and/or stock index options. Dassios and Zhao (2011) introduce a dynamic contagion process, by extending the Hawkes process with

---

\*Postal address: Voie du Roman Pays 20, 1348 Louvain-la-Neuve (Belgium) . E-mail to: donatien.hainaut(at)uclouvain.be

†Postal address: CP210, boulevard du Triomphe, 1050 Bruxelles (Belgium) . E-mail to: griselda.deelstra(at)ulb.ac.be

<sup>1</sup>The very first process, developed by Hawkes (1971), has been used in seismology to model the frequency of earthquakes and aftershocks.

self-exciting and externally-exciting jumps. Chavez-Demoulin and McGill (2012) or Bacry et al. (2013) study the clustering of financial events with Hawkes processes. In insurance, Dassios and Jang (2003) evaluate catastrophe reinsurance and derivatives in presence of self-excitation. We also refer the interested reader to the literature review of Bacry et al. (2015) for a survey of applications of Hawkes processes in finance. However, these processes have two important drawbacks. Firstly, fitting self-exciting jump diffusions still seems to be a difficult exercise. Parameters may be estimated by a generalized method of moments as in Aït-Sahalia et al. (2015). But this method is not fully reliable since the moments can only be approximated. An alternative method consists in using a peak over threshold procedure, as suggested by Embrechts et al. (2011) and applied by Hainaut (2016 a and b) to interest rates and to a stock index. However, this approach is sensitive to the choice of thresholds. Chen and Poon (2013) use instead a particle Markov Chain Monte-Carlo method to estimate parameters. But this approach is time-consuming and the definition of the parameters prior distribution is challenging. The model proposed in this article does not present this drawback and is estimated with a robust filtering technique.

A second drawback of Hawkes processes is that hitting time properties are difficult to obtain for self-exciting jump diffusions and are still unavailable in explicit form, which implies that American or exotic derivatives only can be evaluated by time-consuming numerical methods in the framework of Hawkes processes. Except in the Brownian case, we do not know the distribution of the hitting time for the stock price process to upper or lower boundaries. This makes the valuation of exotic options difficult. When the stock price is ruled by a diffusion with double-exponential jumps, Kou and Wang (2003) show that the Laplace transform of first passage times to flat boundaries admits a closed form expression. It is then possible to evaluate barrier options by inverting this transform with e.g. the Talbot's method such as detailed in Abate and Whitt (2006). This framework has been extended by Jiang and Pistorius (2008) to phase-type jump diffusions, by Levendorskii (2008) to multi-factor models and by Boyarchenko and Levendorskii (2009) to switching jump diffusion processes. Stochastic volatility models, as the Heston's process (1993), explain better than Lévy processes the behaviour of stock prices because their increments are not independent and identically distributed. However for these models, there is no alternative to Monte-Carlo simulations for pricing barrier options. The impact on option prices of jumps clustering modeled with Hawkes processes is studied in Yong and Wu (2017) or in Hainaut (2016 b), but again path dependent options must be computed by simulations. The model proposed in this article does not present this drawback because the moment generating function of hitting times admits a closed form expression.

This article proposes a new alternative to Hawkes processes, based on regime switching processes. In a pure regime switching model as in Honda (2003) or Guidolin and Timmermann (2008), the parameters are modulated by a hidden Markov chain that represents the economic conditions. Even if this category of processes has an excellent explanatory power, they cannot duplicate the clustering of jumps because they use memory-less exponential random variables for defining the length of the period of staying in a certain regime. To remedy to this issue, we construct a Markov chain with several ordered states. Each of these regimes corresponds to a value for the intensity of a discretized self-exciting counting process. These intensities are involved in the definition of the matrix of transition probabilities. This matrix is designed such that when the chain moves to a higher state, the probability of climbing again in the scale of states increases instantaneously. If the chain does not move up, the probability that it falls in the scale of states, raises also with time. This Markov chain serves in the modelling of the asset price dynamics. Indeed, the asset returns are modeled by the sum of a diffusion and a jump process, where the jumps of the prices are synchronized with the transitions of the Markov chain towards higher states. This model will be called a Self-exciting Switching Jump Diffusion (SESJD) model. Chourdakis (2005) or Hainaut and Colwell (2016) combine jumps synchronized to Markov chain transitions with jumps in a two or three switching regime model but the SESJD model differs from these articles in two directions. Firstly, the state space of our Markov chain has a much larger dimension than 2 or 3 regimes. Secondly, we exclusively consider synchronous jumps in order to model the clustering of jumps. Notice that Dassios and Zhao (2014) proposes

a Markov chain model for contagion. But their framework differs from our approach and is designed for modeling the clustering arrival of loss claims with delayed settlement for an insurance company.

In this paper, we will concentrate upon the case that the jumps of the prices are double-exponential random variables, as in the model of Kou (2002) for option pricing. Compared to Hawkes processes, this SESJD model presents several substantial advantages. First and foremost, this model is easy to fit. Indeed, a slightly modified version of the Hamilton's filter leads to a simple parameter estimation method. Secondly, the model explains various forms of option volatility smiles. Thirdly, fluid embedding techniques of e.g. Rogers (1994), adapted by Jiang and Pistorius (2008) to jump processes, can be applied in order to deduce properties about the hitting times of the SESJD. This leads to closed form expressions for some financial derivatives, like e.g. prices of perpetual binary options.

The paper is organized as follows: Section 2 explains how self-exciting processes can be approximated by using a Markov chain. The SESJD model is then introduced in Section 3. Section 4 is devoted to the SESJD parameter estimation by using a modified Hamilton filter. Section 5 presents a family of measure changes and the dynamics of the SESJD model in a risk-neutral world. European option prices and the related volatility smiles are studied in Section 6. In Section 7, we concentrate upon the SESJD hitting time properties and apply them in order to obtain explicit formula for perpetual binary option prices. Section 8 concludes the paper.

## 2 A Markov chain approximation for self-exciting processes

A self-exciting process has sample paths depending upon its history. One of the most studied self-exciting processes was developed by Hawkes (1971). He proposed a jump process in which the intensity increases as soon as a jump occurs and reverts next to a long term level. A common approach to model stock prices consists in combining a Brownian motion part with a self-exciting jump process. However, the econometric calibration of this type of process is problematic and requires advanced filtering techniques. On the other hand, the moment generating function of prices only admits a semi-closed form expression. Its evaluation usually implies to solve numerically a system of ordinary differential equations. Furthermore, we do not have any information about the hitting times of a certain threshold in the case of Hawkes processes. The present paper proposes an alternative solution to introduce a self-excitation mechanism in the dynamics of stock prices. It consists in approaching a Hawkes process by a continuous Markov chain, with a finite number of states. These states are ordered and the stock price jumps when the Markov chain moves from a low to a higher state. Before introducing this Markov chain, we first recall the definition of a Hawkes process and its main features.

A Hawkes process is a counting process denoted by  $(N_t)_{t \geq 0}$  with an intensity process  $(\lambda_t)_{t \geq 0}$ , which reverts to a level  $\theta$  at a speed  $\alpha$  and increases of  $\eta$  ( $\alpha, \theta, \eta \in \mathbb{R}^+$ ) when  $N_t$  jumps by one unit:

$$d\lambda_t = \alpha(\theta - \lambda_t)dt + \eta dN_t. \quad (1)$$

This intensity process  $(\lambda_t)_{t \geq 0}$  is a Markov process and by direct integration, it is clear that the influence of past jumps on the current value of the intensity decays exponentially:

$$\lambda_t = \theta + e^{-\alpha t}(\lambda_0 - \theta) + \int_0^t \eta e^{\alpha(u-t)} dN_u. \quad (2)$$

The integrand in this last expression is called the kernel function. The expected intensity is in this case equal to (for a proof see e.g. Hainaut and Moreaux, 2018),

$$\mathbb{E}(\lambda_t | \mathcal{F}_0) = \left( \frac{\alpha\theta}{\eta - \alpha} + \lambda_0 \right) e^{(\eta - \alpha)t} - \frac{\alpha\theta}{\eta - \alpha}. \quad (3)$$

From this last relation, we infer that the process is stable only if  $\eta - \alpha \leq 0$ . In this case, the asymptotic value to which the expectation of  $\lambda_t$  converges when  $t$  tends to infinity is finite and equal to the following ratio

$$\lambda_\infty := \lim_{t \rightarrow \infty} \mathbb{E}(\lambda_t | \mathcal{F}_0) = \frac{\alpha\theta}{\alpha - \eta}.$$

The next proposition recalls that the moment generating function of  $(N_t)_{t \geq 0}$  is an affine function:

**Proposition 2.1.** *The moment generating function (mgf) of  $N_s$  for  $s \geq t$  with  $\omega_1 \in \mathbb{C}_-$ , is given by the relation:*

$$\mathbb{E}(e^{\omega_1 N_s} | \mathcal{F}_t) = \exp(A(\omega_1, t, s) + B(\omega_1, t, s)\lambda_t + \omega_1 N_t),$$

where  $A(\omega_1, t, s)$ ,  $B(\omega_1, t, s)$  are solutions of the system of ordinary differential equations (ODE's):

$$\begin{cases} \frac{\partial}{\partial t} A &= -\alpha\theta B \\ \frac{\partial}{\partial t} B &= \alpha B - [\exp(B\eta + \omega_1) - 1], \end{cases} \quad (4)$$

with the terminal conditions  $A(\omega_1, s, s) = 0$ ,  $B(\omega_1, s, s) = 0$ .

We refer the reader to Errais et al. (2010) for a proof. The next corollary is an immediate consequence of the previous proposition:

**Corollary 2.2.** *The probability generating function (pgf) of  $N_s$  for  $s \geq t$ , is given by the following expression*

$$\mathbb{E}(u^{N_s - N_t} | \mathcal{F}_t) = \exp(A(\ln u, t, s) + B(\ln u, t, s)\lambda_t),$$

where  $A$  and  $B$  are defined by the ODE's given in (4).

We aim to approximate the dynamics of the intensity process  $(\lambda_t)_{t \geq 0}$  by a Markov chain with a finite dimensional state space. Therefore, we consider a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  on which a Markov chain  $(\delta_t)_{t \geq 0}$  is defined which takes values in a set  $E_0 = \{e_0, \dots, e_n\}$  of unit vectors of  $\mathbb{R}^{n+1}$ . The sub-filtration of  $\mathcal{F}_t$  generated by the Markov chain  $(\delta_t)_{t \geq 0}$  is denoted by  $\{\mathcal{G}_t\}_{t \geq 0}$ . Let us define  $\Delta_\lambda := \frac{\eta}{m}$  and assume that the number of states of the Markov chain,  $n$ , is a multiple of  $m$ . To each state of the Markov chain, we associate a value  $\underline{\lambda}_t = \delta_t^\top \underline{\lambda}$  where  $\underline{\lambda}$  is the following vector:

$$\underline{\lambda} := (\lambda_i)_{i=0:n} = \theta + i\Delta_\lambda \text{ for } i = 0, \dots, n.$$

The next step consists in building the matrix of instantaneous probabilities of transition for  $\delta_t$ , such that  $\underline{\lambda}_t$  has a similar behaviour to the intensity of the Hawkes process,  $\lambda_t$ , as defined in equation (1). The generator associated to this matrix is a  $(n+1) \times (n+1)$  matrix denoted by  $Q_0 := [q_{i,j}]_{i,j=0:n}$ , whose elements satisfy the following standard conditions:

$$q_{i,j} \geq 0, \quad \forall i \neq j, \quad \text{and} \quad \sum_{j=0}^n q_{i,j} = 0, \quad \forall i \in \{0, \dots, n\}. \quad (5)$$

The probability of transiting from state  $i$  to  $j$  for  $i \neq j$  over a short period of time  $\Delta_t$  is equal to  $q_{i,j}\Delta_t + \mathcal{O}(\Delta_t)$ . As  $\sum_{j=0}^n q_{i,j} = 0$  then  $q_{i,i} = -\sum_{j \neq i} q_{i,j} < 0$  and the probability of staying in state  $i$  over  $\Delta_t$  is  $1 + q_{i,i}\Delta_t + \mathcal{O}(\Delta_t)$ . The matrix of transition probabilities over the time interval  $[t, s]$  is denoted as  $P(t, s)$  and is the matrix exponential of this generator matrix times the length of the time interval:

$$P(t, s) = \exp(Q_0(s - t)), \quad s \geq t. \quad (6)$$

The elements  $p_{i,j}(t, s)$  of this transition probability matrix represent the probabilities of switching from state  $i$  at time  $t$  to state  $j$  at time  $s$ :

$$p_{i,j}(t, s) = P(\delta_s = e_j \mid \delta_t = e_i), \quad i, j \in \{0, \dots, n\}. \quad (7)$$

The probability of the chain being in state  $i$  at time  $t$ , denoted by  $p_i(t)$ , depends upon the initial probabilities  $p_k(0)$  at time  $t = 0$  and the transition probabilities  $p_{k,i}(0, t)$ , with  $k = 0, \dots, n$ , as follows:

$$p_i(t) = P(\delta_t = e_i) = \sum_{k=0}^n p_k(0) p_{k,i}(0, t), \quad \forall i \in \{0, \dots, n\}. \quad (8)$$

When considering a small time interval denoted by  $\Delta_t$ , then the probability of switching from the state  $i$  to  $j$  is given by  $q_{i,j}\Delta_t$  for  $i \neq j$ . Furthermore, in view of equation (1), the probability of observing two jumps over  $\Delta_t$  is nearly null given that  $N_t \mid \lambda_t$  is an inhomogeneous Poisson process.

So as to build the generator  $Q_0$  of  $\delta_t$  such that the dynamics of  $\underline{\lambda}_t := \delta_t^\top \underline{\lambda}$  approaches the dynamics of  $\lambda_t$  over a short period  $\Delta_t$ , we analyze two scenarios:

1.  $N_t$  does not jump,
2.  $N_t$  jumps.

Each scenario will be associated to a transition of  $\delta_t$  from a state  $i$  to a state  $j$ . In order to construct the matrix  $Q_0$ , we consider the discrete version of the dynamics of  $\lambda_t$  in equation (1):

$$\lambda_t \approx \lambda_{t-\Delta_t} + \alpha(\theta - \lambda_{t-\Delta_t})\Delta_t + \eta(N_t - N_{t-\Delta_t}). \quad (9)$$

**First scenario:** no jump for the process  $(N_u)_u$  between  $t - \Delta_t$  and  $t$ . From relation (9), if the chain is in state  $i$  at time  $t - \Delta_t$ , we infer that  $\underline{\lambda}_t$  must be equal to

$$\begin{aligned} \underline{\lambda}_t &= \theta + i\Delta_\lambda - \left[ \alpha(\theta + i\Delta_\lambda - \theta) \frac{\Delta_t}{\Delta_\lambda} \right] \Delta_\lambda \\ &= \theta + i\Delta_\lambda - [\alpha i \Delta_t] \Delta_\lambda \end{aligned}$$

which corresponds to the regime  $i - [\alpha\Delta_t i]$  of  $\delta_t$ .

**Second scenario:** a jump of the process  $(N_u)_u$  between  $t - \Delta_t$  and  $t$ . In a discretized framework,  $N_t$  jumps one unity with a probability  $\lambda_t\Delta_t$ . When  $\delta_{t-\Delta_t} = e_i$  and if we ignore the drift term, the arrival state of  $\delta_t$  corresponds to the following values for  $\underline{\lambda}_t$ :

$$\underline{\lambda}_t = \theta + (i + m) \Delta_\lambda,$$

which corresponds to the regime  $i + m$  of  $\delta_t$ . These points suggest then that a generator compliant with the discretized dynamics of  $\lambda_t$  can be defined as:

$$(q_{i,j})_{i,j=0:n} = \begin{cases} \lambda_i & \text{if } j = i + m \\ \frac{1}{\Delta_t} - \lambda_i & \text{if } j = i - [\alpha\Delta_t i] \\ -\sum_{j \neq i} q_{i,j} & j = i \\ \frac{1}{\Delta_t} - \lambda_i & \text{if } j = 0, i - [\alpha\Delta_t i] < 0, j \neq i \\ \lambda_i & \text{if } j = n, i \in [n - m + 1, n - 1] \\ 0 & \text{else.} \end{cases} \quad (10)$$

For such a generator, the probability of switching from a state  $i$  to  $j$  over a small time-interval  $\Delta_t$  (equal to  $q_{i,j}\Delta_t$ ) is either null, either equal to  $\lambda_i\Delta_t$  or to  $1-\lambda_i\Delta_t$ . By construction, the probability of staying in the same state over  $\Delta_t$  is nearly null. To ensure the positiveness of non-diagonal elements of the generator, we impose that  $\frac{1}{\lambda_n} > \Delta_t$ . Each state of  $\delta_t$  corresponds to a certain value of  $\lambda_i$ . The remainder of this section focuses on properties of this Markov Chain  $\delta_t$  defined by this generator in order to introduce self-excitation in stock prices by a new model, presented in the following section. For this purpose, we define new point processes counting the number of transitions between states. To each pair of distinct states  $(i, j)$  in the state space of the Markov chain  $\delta_t$ , we define a point process  $N_{i,j}(t)$  as follows

$$N_{i,j}(t) := \sum_{0 < s \leq t} \mathbf{1}_{\{\delta_{s-}=e_i\}} \mathbf{1}_{\{\delta_s=e_j\}},$$

where  $\mathbf{1}$  is the indicator function.  $N_{i,j}(t)$  counts the number of transitions from states  $i$  to  $j$  up to time  $t$ . We further define the following intensity process

$$\lambda_{i,j}(t) := q_{i,j} \mathbf{1}_{\{\delta_{t-}=e_i\}}. \quad (11)$$

Compensating the counting process  $N_{i,j}(t)$  by the integral of  $\lambda_{i,j}(\cdot)$ , the resulting process

$$M_{i,j}(t) := N_{i,j}(t) - \int_0^t \lambda_{i,j}(s) ds,$$

is a martingale. The total number of changes of regime over the interval  $[0, t]$  from a state  $i$  to  $j$  with  $i < j$ , is denoted by

$$\tilde{N}_t := \sum_{i=0}^n \sum_{j=0, j \neq i}^n \mathbf{1}_{\{j > i\}} N_{i,j}(t).$$

$\tilde{N}_t$  counts the number of transitions of  $(\delta_t)_t$  towards states with a higher intensity than the current one. The compensator of this point process is  $\int_0^t \tilde{\lambda}_s ds$  where

$$\begin{aligned} \tilde{\lambda}_t &= \sum_{i=0}^n \sum_{j=0, j \neq i}^n \mathbf{1}_{\{j > i\}} \lambda_{i,j}(t) \\ &= \sum_{i=0}^n \mathbf{1}_{\{\delta_{t-}=e_i\}} \sum_{j>i}^n q_{i,j} \cdot \\ &:= \delta_{t-} \bar{\lambda} \end{aligned} \quad (12)$$

where  $\bar{\lambda} = (\bar{\lambda}_i)_{i=0, \dots, n}$  and  $\bar{\lambda}_i = \sum_{j>i}^n q_{i,j}$ . The jump process  $\tilde{N}_t$  is determined by six parameters  $(\alpha, \theta, \eta, \Delta_t, n, m)$  and is self-exciting by construction : when  $\delta_t$  moves from a state  $i$  to  $j > i$ , the probability that  $\delta_t$  switches again, increases as  $\lambda_j > \lambda_i$ . Given that  $\left(\tilde{N}_t\right)_{t \geq 0}$  counts the number of transitions to states with a higher intensity than the current one, the probability that  $\tilde{N}_t$  jumps again, also rises instantaneously. To our knowledge, there is no easy way to prove the convergence of this point process  $\tilde{N}_t$  to the Hawkes process  $N_t$  when  $n \rightarrow \infty$  and  $\Delta_t \rightarrow 0$ . However, we can compare numerically their moment and probability generating functions based upon the following proposition and corollary:

**Proposition 2.3.** *The moment generator function (mgf) of  $\tilde{N}_s$  for  $s \geq t$  with  $\omega_1 \in \mathbb{C}_-$ , is given by the following expression*

$$\mathbb{E} \left( e^{\omega_1 \tilde{N}_s} \mid \mathcal{F}_t \right) = \exp \left( A(\omega_1, t, s, \delta_t) + \omega_1 \tilde{N}_t \right),$$

where

$$\begin{aligned}\tilde{A}(\omega_1, t, s) &= \left[ e^{A(\omega_1, t, s, e_0)}, \dots, e^{A(\omega_1, t, s, e_n)} \right]^\top \\ &= \left[ \tilde{A}(\omega_1, t, s, e_0), \dots, \tilde{A}(\omega_1, t, s, e_n) \right]^\top\end{aligned}$$

is a  $n$ -vector of functions, solution of the system of ODE's:

$$0 = \frac{\partial}{\partial t} \tilde{A}(\omega_1, t, s, e_k) + \sum_{j \neq k}^n q_{k,j} \left( \tilde{A}(\omega_1, t, s, e_j) e^{1_{\{j>k\}} \omega_1} - \tilde{A}(\omega_1, t, s, e_k) \right),$$

under the terminal boundary condition:

$$\tilde{A}(\omega_1, s, s, e_k) = 1 \quad k = 0, \dots, n.$$

**Proof of Proposition 2.3.** If we assume that  $\mathbb{E} \left( e^{\omega_1 \tilde{N}_s} \mid \mathcal{F}_t \right) = f(t, \tilde{N}_t, \delta_t)$  for some function depending only on  $t$ ,  $\tilde{N}_t$  and  $\delta_t$ , then this function  $f$  is by Itô's lemma solution of the following equation with  $\delta_t = e_k$ :

$$\mathcal{A}f(t, \tilde{N}_t, e_k) = 0,$$

with  $\mathcal{A}$  the infinitesimal generator

$$\mathcal{A}f(t, \tilde{N}_t, e_k) = \frac{\partial f}{\partial t} + \sum_{j \neq k}^n q_{k,j} \left( f(t, \tilde{N}_t + 1_{\{j>k\}}, e_j) - f(t, \tilde{N}_t, e_k) \right). \quad (13)$$

Let us assume that the function  $f$  is an exponential affine function of  $\tilde{N}_t$ :

$$f(t, \tilde{N}_t, e_k) = \exp \left( A(\omega_1, t, s, e_k) + B(\omega_1, t, s) \tilde{N}_t \right),$$

where  $A(\omega_1, t, s, e_k)$  for  $k = 0, \dots, n$  and  $B(\omega_1, t, s)$  are time dependent functions with terminal conditions  $A(\omega_1, s, s, e_k) = 0$  and  $B(\omega_1, s, s) = \omega_1$ . The partial derivatives of  $f$  are then given by:

$$\frac{\partial f}{\partial t} = \left( \frac{\partial}{\partial t} A(\omega_1, t, s, e_k) + \frac{\partial}{\partial t} B(\omega_1, t, s) \tilde{N}_t \right) f,$$

whereas the sum in equation (13) is equal to

$$\begin{aligned}& \sum_{j \neq k}^n q_{k,j} \left( f(t, \tilde{N}_t + 1_{\{j>k\}}, e_j) - f(t, \tilde{N}_t, e_k) \right) \\ &= \sum_{j \neq k}^n q_{k,j} \left( e^{A(\omega_1, t, s, e_j) + B(\omega_1, t, s) (\tilde{N}_t + 1_{\{j>k\}})} - e^{A(\omega_1, t, s, e_k) + B(\omega_1, t, s) \tilde{N}_t} \right) \\ &= f \sum_{j \neq k}^n q_{k,j} \left( e^{A(\omega_1, t, s, e_j) - A(\omega_1, t, s, e_k) + 1_{\{j>k\}} B(\omega_1, t, s)} - 1 \right).\end{aligned}$$

Injecting these expressions into equation (13), leads to the following relation:

$$0 = \left( \frac{\partial A}{\partial t} + \frac{\partial}{\partial t} B \tilde{N}_t \right) e^{A(\omega_1, t, s, e_k)} + \sum_{j \neq k}^n q_{k,j} \left( e^{A(\omega_1, t, s, e_j) + 1_{\{j>k\}} B(\omega_1, t, s)} - e^{A(\omega_1, t, s, e_k)} \right),$$

from which we deduce that  $B(\omega_1, t, s) = \omega_1$ . Using this fact allows to infer that the functions  $A$  are solutions to the following system of ODE's:

$$0 = \frac{\partial A}{\partial t} e^{A(\omega_1, t, s, e_k)} + \sum_{j \neq k}^n q_{k,j} \left( e^{A(\omega_1, t, s, e_j) + 1_{\{j > k\}} \omega_1} - e^{A(\omega_1, t, s, e_k)} \right).$$

If we define  $\tilde{A}(\omega_1, t, s) = (e^{A(\omega_1, t, s, e_k)})_{k=0, \dots, n} = (\tilde{A}(\omega_1, t, s, e_k))_{k=0, \dots, n}$ , this equation becomes:

$$0 = \frac{\partial}{\partial t} \tilde{A}(\omega_1, t, s, e_k) + \sum_{j \neq k}^n q_{k,j} \left( \tilde{A}(\omega_1, t, s, e_j) e^{1_{\{j > k\}} \omega_1} - \tilde{A}(\omega_1, t, s, e_k) \right).$$

**Corollary 2.4.** *The probability generating function (pgf) of  $\tilde{N}_s$  for  $s \geq t$ , is given by*

$$\mathbb{E} \left( u^{\tilde{N}_s - \tilde{N}_t} \mid \mathcal{F}_t \right) = \exp \left( A(\ln u, t, s, \delta(t)) \right),$$

where  $\tilde{A}(\ln u, t, s) = [e^{A(\ln u, t, s, e_0)}, \dots, e^{A(\ln u, t, s, e_n)}]^\top$  is a vector of functions, solution of the ODE system:

$$0 = \frac{\partial}{\partial t} \tilde{A}(\ln u, t, s, e_k) + \sum_{j \neq k}^n q_{k,j} \left( \tilde{A}(\ln u, t, s, e_j) (u)^{1_{\{j > k\}}} - \tilde{A}(\ln u, t, s, e_k) \right)$$

with the terminal boundary  $\tilde{A}(\ln u, s, s, e_k) = 1 \quad k = 0 \dots n$ .

This corollary is an immediate consequence of the last proposition. In appendix A, we illustrate the convergence of the pgf of the point process  $\tilde{N}_t$  towards the pgf of the Hawkes process  $N_t$  in a numerical example.

### 3 The Self-exciting Switching Jump Diffusion (SESJD) Model

In this section, we propose a price process  $S_t$  for a financial asset with jumps induced by  $\tilde{N}_t$ . This process is defined on  $\Omega$  and the filtration generated by the asset prices is denoted by  $\{\mathcal{H}_t\}_{t \geq 0}$ . We recall that the information about the Markov Chain  $(\delta_t)_{t \geq 0}$  is contained in the filtration  $\{\mathcal{G}_t\}_{t \geq 0}$ . The augmented filtration that gathers information about both processes is denoted by  $\mathcal{F}_t = \mathcal{G}_t \vee \mathcal{H}_t$ . Assuming that  $(W_t)_t$  is a Brownian motion under  $\mathbb{P}$ , the instantaneous return of the asset price process is modeled by the following sum of a drift term, a Brownian motion term, and a compensated jumps part:

$$\frac{dS_t}{S_{t-}} = \mu_t dt + \sigma_t dW_t + (e^J - 1) d\tilde{N}(t) - \tilde{\lambda}_t \mathbb{E} (e^J - 1) dt, \quad (14)$$

or formulated in an alternative way:

$$\begin{aligned} \frac{dS_t}{S_{t-}} &= \mu_t dt + \sigma_t dW_t \\ &+ \sum_{i=0}^n \sum_{j=0, j \neq i}^n 1_{\{j > i\}} \left( (e^J - 1) dN_{i,j}(t) - \lambda_{i,j}(t) \mathbb{E} (e^J - 1) dt \right). \end{aligned} \quad (15)$$

The drift rate  $\mu_t$ , and the Brownian volatility  $\sigma_t$  are modulated by the Markov chain  $\delta$ . That is,  $\mu_t = \delta(t)^\top \bar{\mu}$  and  $\sigma_t = \delta(t)^\top \bar{\sigma}$  where  $\bar{\mu} = (\bar{\mu}_0, \dots, \bar{\mu}_n)^\top \in \mathbb{R}^{n+1}$  and  $\bar{\sigma} = (\sigma_0, \dots, \sigma_n)^\top \in \mathbb{R}_+^{n+1}$ .  $\lambda_{i,j}(t)$  and  $\tilde{\lambda}_t$  are defined



in the previous section by equation (11) and (12). We call this model the self-exciting switching jump diffusion (SESJD) model. Applying Itô's lemma to  $\ln S_t$  leads to the following representation:

$$\begin{aligned} d \ln S_t &= \left( \mu_t - \frac{1}{2} \sigma_t^2 \right) dt + \sigma_t dW_t \\ &+ \sum_{i=0}^n \sum_{j=0, j \neq i}^n 1_{\{j>i\}} \left( J dN_{i,j}(t) - \lambda_{i,j}(t) \mathbb{E}(e^J - 1) dt \right) \end{aligned} \quad (16)$$

from which we infer that  $S_t$  is equal to the following exponential process:

$$\begin{aligned} S_t &= S_0 \exp \left( \int_0^t \left( \mu_s - \frac{1}{2} \sigma_s^2 \right) ds + \int_0^t \sigma_s dW_s \right) \\ &\times \exp \left( \sum_{i=0}^n \sum_{j=0, j \neq i}^n 1_{\{j>i\}} \left( \int_0^t J_s dN_{i,j}(s) - \int_0^t \lambda_{i,j}(s) \mathbb{E}(e^J - 1) ds \right) \right), \end{aligned} \quad (17)$$

or

$$\begin{aligned} S_t &= S_0 \exp \left( \int_0^t \left( \mu_s - \frac{1}{2} \sigma_s^2 - \tilde{\lambda}_s \mathbb{E}(e^J - 1) \right) ds \right) \\ &\times \exp \left( \int_0^t \sigma_s dW_s + \int_0^t J_s d\tilde{N}_s \right) \end{aligned} \quad (18)$$

In the remainder of this article, we assume that jumps are i.i.d. copies of a double-exponential distribution. A double-exponential distributed random variable  $J$  may take positive or negative values. Its probability density function (defined on  $\mathbb{R}$ ) is given by

$$\nu(z) = p \rho^+ e^{-\rho^+ z} 1_{\{z \geq 0\}} - (1-p) \rho^- e^{-\rho^- z} 1_{\{z < 0\}}, \quad (19)$$

while the associated cumulative distribution function equals

$$\mathbb{P}[J \leq z] = (1-p) e^{-\rho^- z} 1_{\{z \leq 0\}} + \left[ (1-p) + p \left( 1 - e^{-\rho^+ z} \right) \right] 1_{\{z > 0\}}.$$

This distribution depends on three parameters:  $\rho^+ \in \mathbb{R}^+$ ,  $\rho^- \in \mathbb{R}^-$ , and  $p \in (0, 1)$ , where  $p$  (*resp.*  $(1-p)$ ) denotes the probability of observing an upward (*resp.* downward) exponential jump, and  $\frac{1}{\rho^+}$  (*resp.*  $\frac{1}{\rho^-}$ ) gives the size of an average positive (*resp.* negative) jump. When only unidirectional jumps are considered, all developments remain valid with  $p = 1$  or  $p = 0$ , for positive and negative exponential jumps. The expected value of the size of the jumps ( $J$ ) is the weighted sum of these average sizes;  $\mathbb{E}(J) = p \frac{1}{\rho^+} + (1-p) \frac{1}{\rho^-}$ . The moment generating function of  $J$  is given by

$$\psi(\omega) = \mathbb{E}(e^{\omega J}) = p \frac{\rho^+}{\rho^+ - \omega} + (1-p) \frac{\rho^-}{\rho^- - \omega}. \quad (20)$$

By construction  $(\tilde{N}_t)_{t \geq 0}$  behaves like a self-exciting process: when  $\delta_t$  moves from a regime  $i$  to  $j > i$ ,  $S_t$  jumps by  $J$  and the instantaneous probability of observing a new jump (proportional to  $\tilde{\lambda}_t$ ) increases.

To simulate sample paths of  $S_t$ , we use an Euler discretization of the equation (14) with daily steps of time (255 trading days per year). If on day  $t$  we have that  $\delta_t = e_i$ , we simulate the next state of the Markov chain by drawing a random uniform number on  $[0, 1]$  and inverting the discrete cumulative distribution function built with  $(p_{i,j}(1\text{day}))_{j=0, \dots, n}$ . Next we generate a random number from a normal distribution with mean  $\delta^\top(t+1) \bar{\mu} \frac{1}{255}$  and variance  $(\delta^\top(t+1) \bar{\sigma})^2 \frac{1}{255}$ . If the chain does not switch to a

higher regime, this Gaussian number is the geometric daily return. If the chain moves to a higher state, we add a random double-exponential jump to the Gaussian return.

A consequence of this self-excitation is that jumps arrive grouped as observed in stock markets during periods of financial turmoil. This feature, called “jump clustering”, is visible in Figure 1 which presents a simulated sample path for  $\tilde{\lambda}_t$ ,  $\delta_t$ ,  $S_t$  and  $\frac{dS_t}{S_t}$  over a period of 5 years. The right upper graph presents the different states through which the Markov chain passes. A comparison with the left upper graph clearly shows the relation between  $\tilde{\lambda}_t$  and the regime through which the chain  $\delta_t$  transits. We also observe in the lower graphs that jumps in the price sample paths occur at transition times of  $\delta_t$  and arrive grouped.

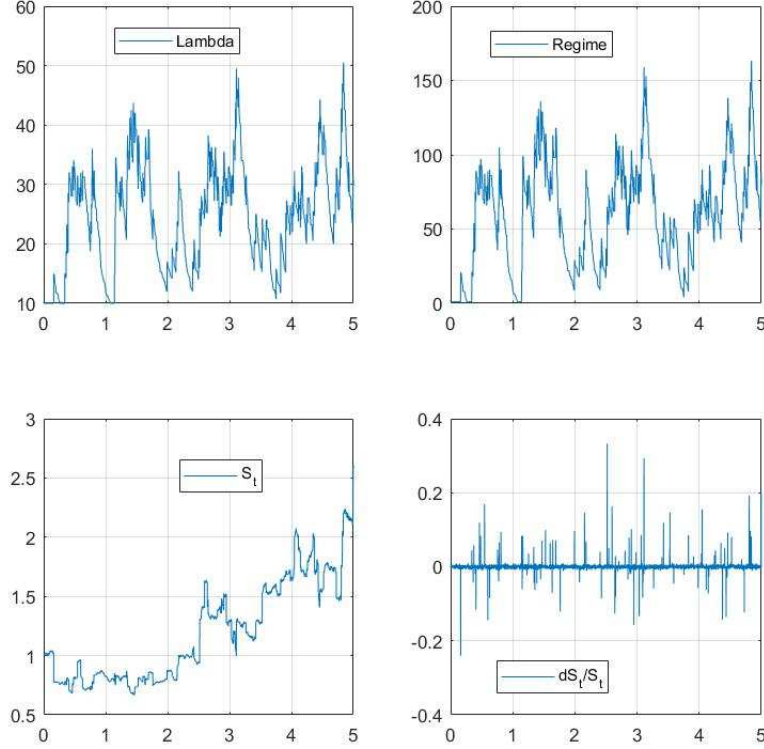


Figure 1: Simulated sample path for  $\tilde{\lambda}_t$ ,  $\delta_t$ ,  $S_t$  and  $\frac{dS_t}{S_t}$ . The parameters defining the Markov chain  $\delta_t$  are:  $\alpha = 8$ ,  $\theta = 10$ ,  $\eta = 5$ ,  $\Delta_t = \frac{1}{200}$ ,  $\lambda_0 = \theta$ , and  $m = 20$ . The drift and the Brownian standard deviation do not depend upon the state of  $\delta_t$ :  $\mu = 0.05$ ,  $\sigma = 0.10$ . The double -exponential distribution of  $J$  has the following parameters:  $p = 0.5$ ,  $\rho^+ = 0.07^{-1}$  and  $\rho^- = -0.07^{-1}$ .

We conclude this section by studying the moment generating function of the log-return of  $S_t$ . Hereto we introduce some new notations. First, the drift of  $d \ln S_t$  is denoted by  $\tilde{\mu}_t$ :

$$\tilde{\mu}_t := \mu_t - \frac{1}{2}\sigma_t^2 - \tilde{\lambda}_t \mathbb{E}(e^J - 1) \quad (21)$$

such that the log-return  $X_t := \ln \frac{S_t}{S_0}$  is given by:

$$X_t := \ln \frac{S_t}{S_0} = \int_0^t \tilde{\mu}_s ds + \int_0^t \sigma_s dW_s + \sum_{i=0}^n \sum_{j=0, j \neq i}^n \int_0^t 1_{\{j > i\}} J dN_{i,j}(s). \quad (22)$$

By construction,  $\tilde{\mu}_t$  takes its value in a  $\mathbb{R}^{n+1}$  vector  $(\tilde{\mu}_0, \dots, \tilde{\mu}_n)$ . According to Itô's lemma for semi-martingales, any function  $f(t, X_t, \delta_t) : \mathbb{R}^+ \times \mathbb{R} \times E_0 \rightarrow \mathbb{R}$  that is  $C^1$  with respect to time and  $C^2$  with respect to  $X_t$  admits the following relation for  $X_t = x$  and  $\delta_t = e_k$ :

$$\begin{aligned} df(t, x, e_k) &= \left( \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x} \tilde{\mu}_t + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} \sigma_t^2 \right) dt + \frac{\partial f}{\partial x} \sigma_t dW_t \\ &\quad + \sum_{j \neq k}^n (f(t, x + 1_{\{j > k\}} J, e_j) - f(t, x, e_k)) dN_t^{kj}. \end{aligned}$$

and its infinitesimal generator  $\mathcal{A}f(t, x, e_k)$  is equal to:

$$\begin{aligned} \mathcal{A}f(t, x, e_k) &= \left( \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x} \tilde{\mu}_t + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} \sigma_t^2 \right) \\ &\quad + \sum_{j \neq k}^n \lambda_{k,j}(t) \int (f(t, x + 1_{\{j > k\}} z, e_j) - f(t, x, e_k)) \nu(z) dz. \end{aligned}$$

We use these results to infer the moment generating function of  $X_s$  which is used in section 6 for European option pricing:

**Proposition 3.1.** *The mgf of  $X_s$  for  $s \geq t$  with  $\omega \in \mathbb{C}_-$ , is given by the following expression*

$$\mathbb{E}(e^{\omega X_s} | \mathcal{F}_t) = \left( \frac{S_t}{S_0} \right)^\omega \exp(A(\omega, t, s, \delta_t)), \quad (23)$$

Where  $\tilde{A}(\omega, t, s) = [e^{A(\omega, t, s, e_0)}, \dots, e^{A(\omega, t, s, e_n)}]^\top$  is a vector of functions, solution of the ODE system:

$$\begin{aligned} 0 &= \frac{\partial}{\partial t} \tilde{A}(\omega, t, s, e_k) + \left( \omega \tilde{\mu}_k + \omega^2 \frac{\sigma_k^2}{2} \right) \tilde{A}(\omega, t, s, e_k) + \\ &\quad + \sum_{j \neq k}^n q_{k,j} \left( \tilde{A}(\omega, t, s, e_j) \psi(1_{\{j > k\}} \omega) - \tilde{A}(\omega, t, s, e_k) \right) \end{aligned}$$

under the terminal boundary condition:

$$\tilde{A}(\omega, s, s, e_k) = 1 \quad k = 0, \dots, n.$$

**Proof of Proposition 3.1.** Let us denote  $f(t, X_t, \delta_t) = \mathbb{E}(e^{\omega X_s} | \mathcal{F}_t)$ . If  $\delta_t = e_k$ , this function is solution of the following equation, implied by the usual argument based on Itô's lemma:

$$\begin{aligned} 0 &= f_t + f_X \tilde{\mu}_k + f_{XX} \frac{\sigma_k^2}{2} + \\ &\quad \sum_{j \neq k}^n q_{k,j}(t) \int (f(t, x + z 1_{\{j > k\}}, e_j) - f(t, x, e_k)) \nu(z) dz \end{aligned} \quad (24)$$

Let us further assume that  $f$  is an exponential affine function of  $X_t$ :

$$f(t, X_t, e_k) = \exp(A(\omega, t, s, e_k) + B(\omega, t, s) X_t),$$

where  $A(\omega, t, s, e_k)$  (for  $k = 0, \dots, n$ ) and  $B(\omega, t, s)$  are time dependent functions with terminal conditions  $A(\omega, s, s, e_k) = 0$  and  $B(\omega, s, s) = \omega$ . The partial derivatives of  $f$  with respect to the state variables are given by:

$$f_t = \left( \frac{\partial}{\partial t} A(\omega, t, s, e_k) + \frac{\partial}{\partial t} B(\omega, t, s) X_t \right) f,$$

$$f_X = B(\omega, t, s)f \quad f_{XX} = B(\omega, t, s)^2 f.$$

The last term in equation (24) can be developed as follows

$$\begin{aligned} & \sum_{j \neq k}^n q_{k,j} \int (f(t, x + 1_{\{j>k\}}z, e_j) - f(t, x, e_k)) \nu(z) dz \\ &= e^{B(\omega, t, s)X_t} \sum_{j \neq k}^n q_{k,j} \int \left( e^{A(\omega, t, s, e_j) + 1_{\{j>k\}}B(\omega, t, s)z} - e^{A(\omega, t, s, e_k)} \right) \nu(z) dz \\ &= e^{B(\omega, t, s)X_t} \sum_{j \neq k}^n q_{k,j} \left( e^{A(\omega, t, s, e_j)} \psi(1_{\{j>k\}}B(\omega, t, s)) - e^{A(\omega, t, s, e_k)} \right). \end{aligned}$$

Injecting these expressions into equation (24), leads to the following relation:

$$\begin{aligned} 0 &= \left( \frac{\partial}{\partial t} A + \frac{\partial}{\partial t} B X_t \right) e^{A(\omega, t, s, e_k)} + B \tilde{\mu}_k e^{A(\omega, t, s, e_k)} + \\ &+ B^2 \frac{\sigma_k^2}{2} e^{A(\omega, t, s, e_k)} + \sum_{j \neq k}^n q_{k,j} \left( e^{A(\omega, t, s, e_j)} \psi(1_{\{j>k\}}B(t, s)) - e^{A(\omega, t, s, e_k)} \right), \end{aligned}$$

from which we infer that  $B(\omega, t, s) = \omega$ . Regrouping terms allows to conclude that  $A(\omega, t, s, e_k)$  for  $k = 0, \dots, n$  are solutions of a system of ODE's:

$$\begin{aligned} 0 &= \frac{\partial}{\partial t} A e^{A(\omega, t, s, e_k)} + \omega \tilde{\mu}_k e^{A(\omega, t, s, e_k)} + \\ &+ \omega^2 \frac{\sigma_k^2}{2} e^{A(\omega, t, s, e_k)} + \sum_{j \neq k}^n q_{k,j} \left( e^{A(\omega, t, s, e_j)} \psi(1_{\{j>k\}}\omega) - e^{A(\omega, t, s, e_k)} \right). \end{aligned}$$

If we define  $\tilde{A}(t, s) = (e^{A(\omega, t, s, e_i)})_{i=0, \dots, n}$ , this last equation is rewritten as follows:

$$\begin{aligned} 0 &= \frac{\partial}{\partial t} \tilde{A}(\omega, t, s, e_k) + \left( \omega \tilde{\mu}_k + \omega^2 \frac{\sigma_k^2}{2} \right) \tilde{A}(\omega, t, s, e_k) + \\ &+ \sum_{j \neq k}^n q_{k,j} \left( \tilde{A}(\omega, t, s, e_j) \psi(1_{\{j>k\}}\omega) - \tilde{A}(\omega, t, s, e_k) \right). \end{aligned}$$

The moment generating function of  $X_t$  may be inverted numerically by a discrete Fourier's transform (DFT) to obtain the pdf and to eventually fit the model to a time series of stock returns. However, in the next section we will present a more efficient alternative way to fit the model to a time series of financial data. ■

## 4 SESJD Parameters estimation with a modified Hamilton's filter

As far as we know, there exists no simple way to fit a self-exciting jump diffusion process to a time series. One could use a generalized moment matching method as in Ait-Sahalia (2015) but this approach is based on some approximations of the moments. Further, one could implement a Markov Chain Monte Carlo (MCMC) procedure but this method is time-consuming and the convergence depends upon the choice of the prior distribution of parameters. Contrary to self-exciting jump diffusion processes, the SESJD model does not present this drawback and can be fitted to a time series with an enhanced version of the Hamilton filter (see Hamilton, 1989). This procedure requires the following result:

**Proposition 4.1.** *The probability density function of the sum  $J + \sigma_i W_\Delta$  where  $\Delta$  is a time interval and when  $\delta_t = e_i$  is equal to:*

$$g(z | \delta_t = e_i) = p\rho^+ \exp\left(\frac{1}{2}(\rho^+)^2 \sigma_i^2 \Delta - \rho^+ z\right) \Phi\left(\frac{z - \rho^+ \sigma_i^2 \Delta}{\sqrt{\Delta} \sigma_i}\right) - (1-p)\rho^- \exp\left(\frac{1}{2}(\rho^-)^2 \sigma_i^2 \Delta - \rho^- z\right) \left(1 - \Phi\left(\frac{z - \rho^- \sigma_i^2 \Delta}{\Delta \sigma_i}\right)\right) \quad (25)$$

for  $z \in \mathbb{R}$  and where  $\Phi(\cdot)$  is the cdf of a standard normal random variable  $N(0, 1)$ .

**Proof of Proposition 4.1.**  $g(z | \delta_t = e_i)$  is the convolution of densities of  $J$  and  $\sigma_i W_\Delta$  (denoted by  $\tilde{f}$ ),

$$\begin{aligned} g(z | \delta_t = e_i) &= \int_{-\infty}^{+\infty} \nu(u) \tilde{f}(z - u) du \\ &= p\rho^+ \int_0^{+\infty} e^{-\rho^+ u} \frac{1}{\sqrt{2\pi\Delta}\sigma_i} \exp\left(-\frac{1}{2} \frac{(z - u)^2}{\sigma_i^2 \Delta}\right) du \\ &\quad - (1-p)\rho^- \int_{-\infty}^0 e^{-\rho^- u} \frac{1}{\sqrt{2\pi\Delta}\sigma_i} \exp\left(-\frac{1}{2} \frac{(z - u)^2}{\sigma_i^2 \Delta}\right) du, \end{aligned}$$

which can be rewritten as

$$\begin{aligned} g(z | \delta_t = e_i) &= \frac{p\rho^+}{\sqrt{2\pi\Delta}\sigma_i} \int_0^{+\infty} \exp\left(-\frac{1}{2} \frac{(z - u)^2 + 2\rho^+ \sigma_i^2 \Delta u}{\sigma_i^2 \Delta}\right) du \\ &\quad - \frac{(1-p)\rho^-}{\sqrt{2\pi\Delta}\sigma_i} \int_{-\infty}^0 \exp\left(-\frac{1}{2} \frac{(z - u)^2 + 2\rho^- \sigma_i^2 \Delta u}{\sigma_i^2 \Delta}\right) du. \end{aligned} \quad (26)$$

Since clearly

$$(z - u)^2 + 2\rho^+ \sigma_i^2 \Delta u = ((z - u) - \rho^+ \sigma_i^2 \Delta)^2 - (\rho^+ \sigma_i^2 \Delta)^2 + 2\rho^+ \sigma_i^2 \Delta z,$$

the first integral in equation (26) can be rewritten as

$$\begin{aligned} &\frac{1}{\sqrt{2\pi\Delta}\sigma_i} \int_0^{+\infty} \exp\left(-\frac{1}{2} \frac{(z - u)^2 + 2\rho^+ \sigma_i^2 \Delta u}{\sigma_i^2 \Delta}\right) du \\ &= \frac{\exp\left(\frac{1}{2}(\rho^+)^2 \sigma_i^2 \Delta - \rho^+ z\right)}{\sqrt{2\pi\Delta}\sigma_i} \int_0^\infty \exp\left(-\frac{1}{2} \frac{((z - u) - \rho^+ \sigma_i^2 \Delta)^2}{\sigma_i^2 \Delta}\right) du. \end{aligned} \quad (27)$$

Using the substitution  $v = ((z - u) - \rho^+ \sigma_i^2 \Delta)$  implies that  $u = ((z - v) - \rho^+ \sigma_i^2 \Delta)$  and  $du = -dv$ . Moreover, if  $u = 0$  then  $v = z - \rho^+ \sigma_i^2 \Delta$  and when  $u = +\infty$ ,  $v = -\infty$ . As a consequence, the integral in equation (27) becomes:

$$\begin{aligned} &\frac{1}{\sqrt{2\pi\Delta}\sigma_i} \int_0^\infty \exp\left(-\frac{1}{2} \frac{((z - u) - \rho^+ \sigma_i^2 \Delta)^2}{\sigma_i^2 \Delta}\right) du \\ &= \frac{-1}{\sqrt{2\pi\Delta}\sigma_i} \int_{z - \rho^+ \sigma_i^2 \Delta}^{-\infty} \exp\left(-\frac{1}{2} \frac{v^2}{\Delta \sigma_i^2}\right) dv \\ &= \frac{1}{\sqrt{2\pi\Delta}\sigma_i} \int_{-\infty}^{z - \rho^+ \sigma_i^2 \Delta} \exp\left(-\frac{1}{2} \frac{v^2}{\Delta \sigma_i^2}\right) dv \\ &= \Phi\left(\frac{z - \rho^+ \sigma_i^2 \Delta}{\sqrt{\Delta} \sigma_i}\right) \end{aligned}$$

where  $\Phi(\cdot)$  is the cdf of a  $N(0, 1)$  random variable.

On the other hand, the second integral in equation (26) is equal to

$$\begin{aligned} & \frac{1}{\sqrt{2\pi\Delta\sigma_i}} \int_{-\infty}^0 \exp\left(-\frac{1}{2} \frac{(z-u)^2 + 2\rho^-\sigma_i^2\Delta u}{\Delta\sigma_i^2}\right) du \\ &= \frac{\exp\left(\frac{1}{2}(\rho^-)^2\sigma_i^2\Delta - \rho^-z\right)}{\sqrt{2\pi\Delta\sigma_i}} \int_{-\infty}^0 \exp\left(-\frac{1}{2} \frac{((z-u) - \rho^-\sigma_i^2\Delta)^2}{\Delta\sigma_i^2}\right) du \end{aligned} \quad (28)$$

Analogously, using the substitution  $v = ((z-u) - \rho^-\sigma_i^2\Delta)$  implies that  $u = ((z-v) - \rho^-\sigma_i^2\Delta)$  and  $du = -dv$ . Moreover, if  $u = 0$  then  $v = z - \rho^-\sigma_i^2\Delta$  and when  $u = -\infty$ ,  $v = +\infty$ . Therefore, the integral in equation (28) turns out to equal

$$\begin{aligned} & \frac{1}{\sqrt{2\pi\Delta\sigma_i}} \int_{-\infty}^0 \exp\left(-\frac{1}{2} \frac{((z-u) - \rho^-\sigma_i^2\Delta)^2}{\Delta\sigma_i^2}\right) du \\ &= \frac{-1}{\sqrt{2\pi\Delta\sigma_i}} \int_{+\infty}^{z-\rho^-\sigma_i^2\Delta} \exp\left(-\frac{1}{2} \frac{v^2}{\Delta\sigma_i^2}\right) dv \\ &= \frac{1}{\sqrt{2\pi\Delta\sigma_i}} \int_{z-\rho^-\sigma_i^2\Delta}^{+\infty} \exp\left(-\frac{1}{2} \frac{v^2}{\Delta\sigma_i^2}\right) dv \\ &= \left(1 - \Phi\left(\frac{z - \rho^-\sigma_i^2\Delta}{\Delta\sigma_i}\right)\right). \end{aligned}$$

Combining these expressions (26)-(28) leads to the results. ■

In the rest of this section, we denote by  $x_1, x_2, \dots, x_T$ , the time series of log-returns of a financial asset, measured at times  $t_1, \dots, t_T$  equally spaced by  $\Delta$  (which is not necessary equal to  $\Delta_t$  involved in the definition of  $Q_0$  for  $\delta_t$ ):

$$x_i = \ln\left(\frac{S_{t_{i-1}+\Delta}}{S_{t_{i-1}}}\right) \quad i = 1, \dots, T.$$

We assume that the Markov chain  $\delta_t$  only changes of regime at times  $t_i$  for  $i = 1, \dots, T$ . Hence, if the economy stays in the  $j^{\text{th}}$  state over the period of time  $[t_{i-1}, t_i]$ , the log-return is normally distributed  $X_i \sim N(\tilde{\mu}_j\Delta, \sigma_j\sqrt{\Delta})$ . If the system switches from regime  $i$  to  $j$ , the density of the log-return is equal to  $g(z | \delta_t = e_i)$  given by equation (25). We denote by

$$\Theta = (\bar{\mu}, \bar{\sigma}, \alpha, \theta, \eta, p, \rho^+, \rho^-)$$

the set of parameters of the SESJD model.  $\Delta_t$ ,  $m$  and  $n$  are not considered as parameters and are chosen a priori. Using the Bayes' rule, we reformulate the log-likelihood of observed returns as follows:

$$\begin{aligned} \log f(x_1, \dots, x_T | \Theta) &= \log f(x_1 | \Theta) + \log f(x_2 | \Theta, x_1) + \log f(x_3 | \Theta, x_1, x_2) \\ &+ \dots + \log f(x_T | \Theta, x_1, \dots, x_{T-1}) \end{aligned}$$

where  $f(x_k | \Theta, x_1, \dots, x_{k-1})$  is the density function of the return on the  $k^{\text{th}}$  period, for parameters  $\Theta$  and conditionally to previous observations  $x_1, \dots, x_{k-1}$ . The parameters are estimated by maximizing this log-likelihood function.

Therefore, we concentrate upon the terms in the right-hand side of this log-likelihood. Conditioning upon the state of  $\delta_k$  allows us to infer that  $f(x_k | \Theta, x_1, \dots, x_{k-1})$  is equal to:

$$\begin{aligned} f(x_k | \Theta, x_1, \dots, x_{k-1}) &= \sum_{i=0}^n \sum_{j=0}^n p_i(t_{k-1} | \Theta, x_1, \dots, x_{k-1}) p_{i,j}(t_{k-1}, t_k | \Theta) \\ &\times f(x_k | \Theta, \delta_{t_k} = e_j, \delta_{t_{k-1}} = e_i) \end{aligned}$$

where

- $f(x_k|\Theta, \delta_{t_k} = e_j, \delta_{t_{k-1}} = e_i)$  is
  - either the Gaussian density of the return in state  $i$ ,  $N(\tilde{\mu}_i\Delta, \sigma_i\sqrt{\Delta})$  if  $j \leq i$ ,
  - or either  $g(z - \tilde{\mu}_i\Delta | \delta_t = e_i)$  if  $i < j$ , with  $g(z | \delta_t = e_i)$  the probability density function of the sum  $J + \sigma_t W_\Delta$  as given by equation (25).
- $p_{i,j}(t_{k-1}, t_k|\Theta)$  is the probability of transition, as defined by eq. (7), from state  $i$  at time  $t_{k-1}$  to state  $j$  at time  $t_k$  for the set of parameters  $\Theta$ ,
- $p_i(t_{k-1}|\Theta, x_1, \dots, x_{k-1})$  is the probability of presence in state  $i$  at time  $t_{k-1}$ , conditionally to all observations up to  $t_{k-1}$ .

Using again the Bayes' rule, the probability  $p_i(t_{k-1}|\Theta, x_1, \dots, x_{k-1})$  may be inferred recursively from  $f(x_{k-1}|\Theta, x_1, \dots, x_{k-2})$  as follows:

$$\begin{aligned} p_i(t_{k-1}|\Theta, x_1, \dots, x_{k-1}) & \quad (29) \\ &= \frac{\sum_{j=0}^n p_j(t_{k-2}|\Theta, x_1, \dots, x_{k-2}) p_{j,i}(t_{k-2}, t_{k-1}|\Theta) f(x_{k-1}|\Theta, \delta_{t_{k-1}} = e_i, \delta_{t_{k-2}} = e_j)}{f(x_{k-1}|\Theta, x_1, \dots, x_{k-2})} \end{aligned}$$

In order to initiate the recursion, we need to determine  $f(x_1|\Theta)$ . If the Markov chain has been running for a sufficiently long enough period of time, we assume that the probability of presence in a given state is equal to its stationary probability, denoted  $p_i(\Theta)$  for  $i = 1, \dots, n$ . Then, we infer that

$$f(x_1|\Theta) = \sum_{i=0}^n \sum_{j=0}^n p_i(\Theta) p_{i,j}(t_0, t_1|\Theta) f(x_1|\Theta, \delta_{t_1} = e_j, \delta_{t_0} = e_i).$$

Therefore, the log-likelihood is evaluated by recursion and maximized numerically to estimate parameters. After this calibration, we filter the states through which the Markov chain transits by using the relation:

$$\mathbb{E} \left( \delta_{t_k}^\top \begin{pmatrix} 0 \\ \vdots \\ n \end{pmatrix} \middle| \mathcal{H}_{t_k} \right) = \sum_{i=0}^n p_i(t_k|\Theta, x_1, \dots, x_k) i.$$

To illustrate this modified Hamilton filter, we fit the SESJD model to a time series of the S&P 500 stock index, containing daily returns in the period between 19/6/2007 and 22/5/2017 (2500 observations). In order to limit the number of parameters to estimate, the drifts  $(\tilde{\mu}_0, \dots, \tilde{\mu}_n)$  and standard deviations  $(\tilde{\sigma}_0, \dots, \tilde{\sigma}_n)$  are assumed to be constant and respectively equal to  $\mu$  and  $\sigma$ . The parameter of discretization  $m$  ranges from 3 to 16.  $n$  is assumed to be equal to  $7m$  whereas  $\Delta_t = \frac{1}{200}$  and  $\Delta$  is chosen to be equal to one trading day ( $\Delta = \frac{1}{252}$ ). Table 1 reports the log-likelihood of several tested models. As we could expect, the goodness of fit is better for the SESJD model than for a pure diffusion process. Moreover, increasing  $m$  and therefore the number of states of  $\delta_t$ , improves the log-likelihood. Table 2 presents the parameter estimates for  $m = 5$  and  $n + 1 = 36$ . On average, we observe at least 2.83 jumps per year ( $\theta$  is the lowest value allowed for the intensity). A jump is negative with a probability of 60%. Whereas the mean sizes of positive and negative jumps are respectively equal to 2.11% and to -2.13%. The parameter estimate of  $\eta$  turns out to be 17.01. As the speed of mean reversion  $\alpha$  is higher than  $\eta$ , the chain easily returns to states in which the intensity is close to  $\theta$ . The standard deviation of the diffusion is around 11%.

We also compare our model to a jump-diffusion process with Hawkes jumps. In the latter model, the stock price is ruled by the SDE (14) but the intensity of jumps is driven by equation (1). We filter the

intensity of Hawkes jump with an enhanced version of the “peaks over threshold” (POT) of Embrechts et al. (2011) and detailed in Hainaut (2016 a) and Hainaut and Moraux (2018). This procedure, reminded in appendix B, is less accurate than MCMC approaches but is much easier to implement. Parameters of the Hawkes jump diffusion model are reported in the fourth column of Table 2 and are clearly consistent with SESJD estimates. However, jumps in the SESJD are smaller on average because the POT method ignores the diffusion part on days for which a jump is detected and then overestimate jumps. The last column of Table 2 reports results of Boswijk et al. (2015) who fit a jump-diffusion with self-excitation to the S&P 500. Their model differs from our approach on the following points. Firstly, we assume that jumps are distributed according to a double-exponential law instead of being normal random variables. Secondly, we do not consider stochastic volatility. However, a comparison of parameter estimates with these reported in Boswijk et al. (2015) allows us to understand the origin of the volatility in each model. The speeds of mean reversion  $\alpha$  are comparable (20.83 in our model versus 18.16) but the mean reversion level is much higher in our model. In Boswijk et al. (2015), self-exciting jumps correspond to rare and violent economic shocks whereas other variations are explained by the stochastic volatility. As we do not include this feature in the dynamics of stock prices, jumps are more frequent in our setting and explain a larger variety of events.

Model	Loglikelihoods
Brownian Motion	7279
$m = 3$	7813
$m = 5$	7820
$m = 9$	7823
$m = 16$	7825

Table 1: The first line presents the log-likelihood for a diffusion with a drift fitted to S&P 500. The other lines present the log-likelihood for the SESJD model, with different levels of discretization.

Parameters	Estimates	Standard Deviations	POT	Boswijk et al.
$\alpha$	20.83	0.9410	18.85	18.16
$\theta$	2.83	0.2723	5.42	0.32
$\eta$	17.01	0.8908	15.30	16.62
$p$	0.40	0.0259	0.37	
$\rho^+$	47.41	2.6270	30.31	
$\rho^-$	-46.87	1.6908	-33.59	
$\mu$	0.1620	0.0237	0.2110	
$\sigma$	0.1084	0.0027	0.1197	

Table 2: First and second columns: Parameter estimates of the SESJD and standard errors for  $m = 5$ ,  $n + 1 = 36$  and  $\Delta_t = \frac{1}{200}$ . Third column: parameter estimates of the Hawkes jump diffusion model with the POT method. Last column: parameter estimates in Boswijk et al. (2015).

The first plot of figure 2 presents the filtered states of the SESJD Markov chain to daily S&P 500



log-returns.  $\delta_t$  climbs in the scale of states during the periods of high volatility: from September 2008 to the end 2009 (the US credit crunch period), from September 2011 to February 2012 (the second period of the double-dip recession) or the first months of 2016 (the fear of deflation). The second plot of figure 2 compares filtered intensities  $\tilde{\lambda}_t$  of the SESJD process and  $\lambda_t$  of the Hawkes diffusion process.  $\tilde{\lambda}_t$  exhibits the same behaviour as  $\lambda_t$ , excepted that  $\tilde{\lambda}_t$  is upper bounded because the number of regimes is limited to 36 in our illustration. We will use the set of parameters reported in Table 2 for further numerical illustrations in the next sections.

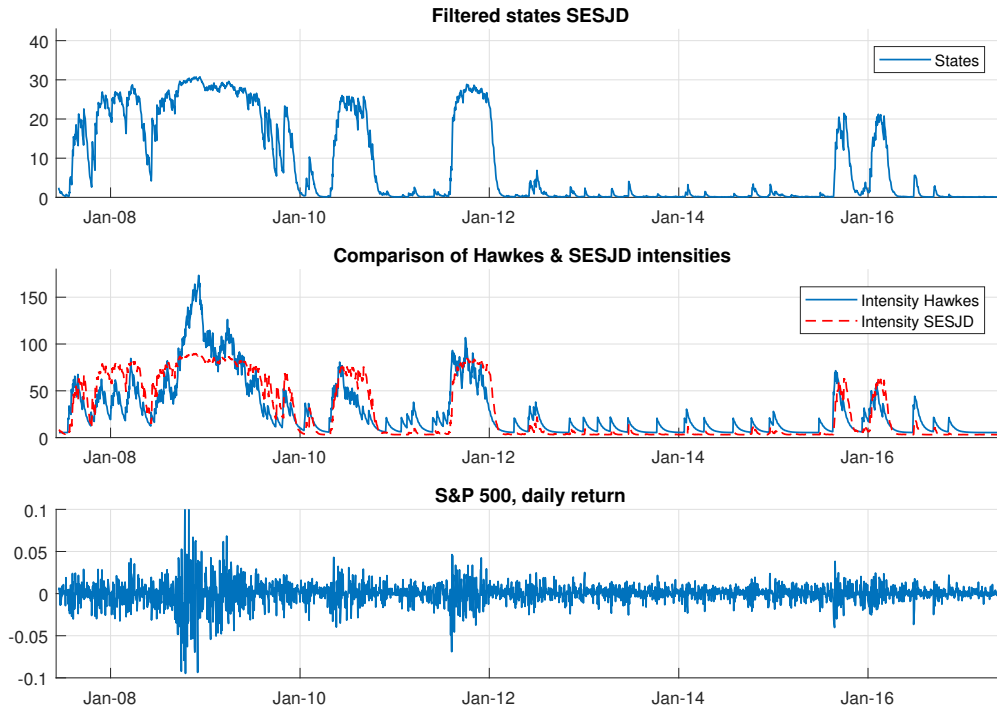


Figure 2: The upper graph shows the filtered sample path of the Markov chain  $\delta_t$ . The second graph compares filtered intensities  $\tilde{\lambda}_t$  of the SESJD process and  $\lambda_t$  of the Hawkes diffusion process. The last plot presents daily log-returns of the S&P 500 from 2007 to 2017.

## 5 Changes of measure and risk neutral world

In this section, we first present a family of change of measures for point processes  $Z_{i,j}(t) := \sum_{k=1}^{N_{i,j}(t)} J_k$ . Let  $\nu^b(\cdot)$  be a density of probability defined on the same domain as  $\nu(\cdot)$ , the pdf of the jumps  $J$  under  $\mathbb{P}$ . We define the following log-ratio:

$$\phi(h, u) := \ln \left( h \frac{\nu^b(u)}{\nu(u)} \right) \quad (30)$$

where  $h \in \mathbb{R}^+$  and  $u$  is in the support of  $\nu(\cdot)$ . For any  $h_{i,j} \in \mathbb{R}^+$  such that  $h_{i,j} > -1$  for  $i, j = 0, \dots, n$ , the following compensated jump process

$$\begin{aligned}
M_{i,j}(t) &= \sum_{k=1}^{N_{i,j}(t)} \left( e^{\phi(h_{i,j}, J_k)} - 1 \right) - \int_0^t \lambda_{i,j}(s) \mathbb{E} \left( e^{\phi(h_{i,j}, J_k)} - 1 \mid \mathcal{F}_0 \right) ds \\
&= \sum_{k=1}^{N_{i,j}(t)} \left( e^{\phi(h_{i,j}, J_k)} - 1 \right) - \int_0^t \lambda_{i,j}(s) (h_{i,j} - 1) ds,
\end{aligned} \tag{31}$$

is a martingale by construction. The second equality in equation (31) follows from the relation:

$$\mathbb{E} \left( e^{\phi(h_{i,j}, J_k)} - 1 \mid \mathcal{F}_0 \right) = h_{i,j} \int_{-\infty}^{\infty} \left( \frac{\nu^b(u)}{\nu(u)} \right) \nu(u) du - 1.$$

We will later see that  $h_{i,j}$  is involved in the definition of a risk neutral measure. Notice that one could consider the generalization that  $h_{i,j}$  is an  $\mathcal{F}_t$ -adapted process but, in this case, the process  $\delta_t$  would no longer be a Markov chain under the equivalent measures.

In the next proposition, we will define an interesting family of equivalent measures and the law of the point process  $Z_{i,j}(t)$  under the new measure. This result follows from Girsanov theorem for semi-martingales, but we include the proof for comprehension in this specific case.

**Proposition 5.1.** *Let  $Z_{i,j}(t)$  be point processes defined by*

$$Z_{i,j}(t) := \sum_{k=1}^{N_{i,j}(t)} J_k,$$

for  $i, j = 0, \dots, n$ . If  $h_{i,j} > -1$ , the processes  $L_{i,j}(t)$  defined as follows

$$L_{i,j}(t) = \exp \left( \int_0^t \phi(h_{i,j}, J_s) dN_{i,j}(s) - \int_0^t \lambda_{i,j}(s) (h_{i,j} - 1) ds \right), \tag{32}$$

for  $i, j = 0, \dots, n$ , are Radon-Nikodym derivatives  $\frac{d\mathbb{P}^b}{d\mathbb{P}}$  from the real measure  $\mathbb{P}$  to a new probability measure  $\mathbb{P}^b$ . Under  $\mathbb{P}^b$ ,  $Z_{i,j}(t)$  is still a point process but its dynamics are modified as follows

$$Z_{i,j}(t) = \sum_{k=1}^{N_{i,j}^h(t)} J_k^b, \tag{33}$$

where  $N_{i,j}^h(t)$  is a counting process of intensity  $\lambda_{i,j}(t)h_{i,j}$  and  $J_k^b$  are i.i.d. jumps with pdf  $\nu^b(u)$ . According to the definition (11) of  $\lambda_{i,j}(t)$ , the matrix of transition probabilities of  $(\delta_t)_{t \geq 0}$  under  $\mathbb{Q}$  is equal to  $Q_0^{\mathbb{Q}} = (q_{i,j} h_{i,j})_{i,j=1:n}$ .

**Proof of Proposition 5.1.** From equation (31),  $M_{i,j}(t)$  is a martingale satisfying the SDE:

$$dM_{i,j}(t) = \left( e^{\phi(h_{i,j}, J)} - 1 \right) dN_{i,j}(t) - \lambda_{i,j}(t) (h_{i,j} - 1) dt.$$

We can construct a martingale  $L_{i,j}(t)$  with geometric dynamics given by:

$$\begin{aligned}
dL_{i,j}(t) &= L_{i,j}(t) dM_{i,j}(t) \\
&= L_{i,j}(t) \left( e^{\phi(h_{i,j}, J)} - 1 \right) dN_{i,j}(t) - L_{i,j}(t) \lambda_{i,j}(t) (h_{i,j} - 1) dt
\end{aligned}$$

If we apply Itô's lemma to the function  $\ln L_{i,j}(t)$ , the differential of  $\ln L_{i,j}(t)$  clearly equals

$$d \ln L_{i,j}(t) = \phi(h_{i,j}, J) dN_{i,j}(t) - \lambda_{i,j}(t) (h_{i,j} - 1) dt,$$

from which equation (32) follows.

The expectation of  $Z_{i,j}(t)$  under the measure  $\mathbb{P}^b$  defined by the Radon-Nikodym derivative  $L_{i,j}(t)$ , is given by

$$\mathbb{E}^{P^b} \left( e^{uZ_{i,j}(t)} | \mathcal{F}_0 \right) = \mathbb{E} \left( e^{\int_0^t (uJ_s + \phi(h_{i,j}, J)) dN_{i,j}(s) - \int_0^t \lambda_{i,j}(s) (h_{i,j} - 1) ds} | \mathcal{F}_0 \right). \quad (34)$$

If the filtrations of  $N_{i,j}(t)$  and  $\lambda_{i,j}(t)$  are momentarily denoted by  $\mathcal{G}_t^{i,j} \subset \mathcal{F}_t$  and  $\mathcal{H}_t^{i,j} \subset \mathcal{F}_t$ , using nested expectations allows to rewrite the expectation (34) as follows:

$$\begin{aligned} & \mathbb{E} \left( e^{\int_0^t (uJ_s + \phi(h_{i,j}, J)) dN_{i,j}(s) - \int_0^t \lambda_{i,j}(s) (h_{i,j} - 1) ds} | \mathcal{F}_0 \right) \\ &= \mathbb{E} \left( e^{-\int_0^t \lambda_{i,j}(s) (h_{i,j} - 1) ds} \mathbb{E} \left( e^{\int_0^t (uJ_s + \phi(h_{i,j}, J)) dN_{i,j}(s)} | \mathcal{G}_t^{i,j} \vee \mathcal{H}_t^{i,j} \vee \mathcal{F}_0 \right) | \mathcal{F}_0 \right) \\ &= \mathbb{E} \left( e^{-\int_0^t \lambda_{i,j}(s) (h_{i,j} - 1) ds} \mathbb{E} \left( \prod_{k=1}^{N_{i,j}(t)} \mathbb{E} \left( e^{uJ + \phi(h_{i,j}, J)} | \mathcal{G}_t^{i,j} \vee \mathcal{H}_t^{i,j} \vee \mathcal{F}_0 \right) | \mathcal{H}_t^{i,j} \vee \mathcal{F}_0 \right) | \mathcal{F}_0 \right). \end{aligned} \quad (35)$$

By definition of  $\phi(\cdot)$ , we have that  $\mathbb{E} \left( e^{uJ + \phi(h_{i,j}, J)} | \mathcal{G}_t^{i,j} \vee \mathcal{H}_t^{i,j} \vee \mathcal{F}_0 \right) = \int h_{i,j} e^{uz} \nu^b(z) dz = h_{i,j} \mathbb{E} \left( e^{uJ^b} \right)$ . Furthermore, conditionally to  $\mathcal{H}_t^{i,j} \vee \mathcal{F}_0$ ,  $N_{i,j}(t)$  is a inhomogeneous Poisson process with the following moment generating function:

$$\begin{aligned} \mathbb{E} \left( \prod_{k=1}^{N_{i,j}(t)} h_{i,j} \mathbb{E} \left( e^{uJ^b} \right) | \mathcal{H}_t^{i,j} \vee \mathcal{F}_0 \right) &= \mathbb{E} \left( e^{N_{i,j}(t) \ln [h_{i,j} \mathbb{E} (e^{uJ^b})]} | \mathcal{H}_t^{i,j} \vee \mathcal{F}_0 \right) \\ &= \exp \left( \int_0^t \lambda_{i,j}(s) \left( h_{i,j} \mathbb{E} \left( e^{uJ^b} \right) - 1 \right) ds \right). \end{aligned}$$

We infer from this last equation that the expectation (35) is equal to

$$\begin{aligned} & \mathbb{E} \left( e^{\int_0^t uJ_s + \phi(h_{i,j}, J) dN_{i,j}(s) - \int_0^t \lambda_{i,j}(s) (h_{i,j} - 1) ds} | \mathcal{F}_0 \right) \\ &= \mathbb{E} \left( e^{-\int_0^t \lambda_{i,j}(s) (h_{i,j} - 1) ds + \int_0^t \lambda_{i,j}(s) \left( h_{i,j} \mathbb{E} (e^{uJ^b}) - 1 \right) ds} | \mathcal{F}_0 \right) \\ &= \mathbb{E} \left( \exp \left( \int_0^t h_{i,j} \lambda_{i,j}(s) \left( \mathbb{E} \left( e^{uJ^b} \right) - 1 \right) ds \right) | \mathcal{F}_0 \right) \end{aligned}$$

which turns out to be the moment generating function of  $Z_{i,j}(t)$  under the equivalent measure  $\mathbb{P}^b$ .  $\blacksquare$

To avoid arbitrage opportunities, financial derivatives are priced under an equivalent risk neutral measure under which discounted (non-dividend paying) asset prices are martingales. In the remainder of this section, we consider a financial market composed of two assets: a risk free cash account and a stock. The interest rate depends on  $\delta_t$  and is defined as  $r_t = \delta_t \bar{r}^\top$  where  $\bar{r} = (r_0, \dots, r_n)^\top \in \mathbb{R}^{n+1}$ . The stock price  $S_t$  is defined by equation (17). By construction, the risk neutral measure is not unique. We consider Radon-Nikodym derivatives of the following form

$$\begin{aligned} L_t &= \prod_{i,j=0}^n \exp \left( \int_0^t \phi(h_{i,j}, J_s) dN_{i,j}(s) - \int_0^t \lambda_{i,j}(s) \left( \mathbb{E}(h_{i,j} | \mathcal{F}_0) - 1 \right) ds \right) \\ &\quad \times \exp \left( -\frac{1}{2} \int_0^t \beta_s^2 ds + \int_0^t \beta_s dW_s \right). \end{aligned} \quad (36)$$

where  $(\beta_t)_{t \geq 0}$  is a  $\mathcal{F}_t$  measurable process. The last factor in the definition of this Radon-Nikodym derivative implies that

$$dW_t^\beta = dW_t + \beta_t dt$$

is a Brownian motion under the new equivalent measure. The next proposition establishes the dynamics of  $S_t$  under such a new martingale measure, which will be denoted by  $\mathbb{Q}$ .

**Proposition 5.2.** *The dynamics of the asset price under the equivalent measure  $\mathbb{Q}$  defined by the Radon-Nikodym derivative (36) equals*

$$\begin{aligned} \frac{dS_t}{S_t} &= (\mu_t - \sigma_t \beta_t) dt + \sum_{i=0}^n \sum_{j=0, j \neq i}^n 1_{\{j > i\}} \lambda_{i,j}(t) \left( h_{i,j} \mathbb{E} \left( e^{J^b} - 1 \right) dt - \mathbb{E} \left( e^J - 1 \right) dt \right) \\ &\quad + \sigma_t dW_t^\beta + \sum_{i=0}^n \sum_{j=0, j \neq i}^n 1_{\{j > i\}} \left( (e^{J^b} - 1) dN_{i,j}^h(t) - h_{i,j} \lambda_{i,j}(t) \mathbb{E} \left( e^{J^b} - 1 \right) dt \right). \end{aligned}$$

**Proof of Proposition 5.2.** Notice that equation (16) defining  $d \ln S_t$  can be rewritten as follows

$$\begin{aligned} dY_t = d \ln S_t &= \left( \mu_t - \frac{1}{2} \sigma_t^2 - \sigma_t \beta_t \right) dt + \sigma_t (dW_t + \beta_t dt) \\ &\quad + \sum_{i=0}^n \sum_{j=0, j \neq i}^n 1_{\{j > i\}} \lambda_{i,j}(t) \left( h_{i,j} \mathbb{E} \left( e^{J^b} - 1 \right) dt - \mathbb{E} \left( e^J - 1 \right) dt \right) \\ &\quad + \sum_{i=0}^n \sum_{j=0, j \neq i}^n 1_{\{j > i\}} \left( dZ_{i,j}(t) - h_{i,j} \lambda_{i,j}(t) \mathbb{E} \left( e^{J^b} - 1 \right) dt \right). \end{aligned}$$

Applying Itô's lemma to the function  $f(Y_t) = e^{Y_t}$ , leads to the dynamics under  $\mathbb{Q}$  given by

$$\begin{aligned} \frac{dS_t}{S_t} &= (\mu_t - \sigma_t \beta_t) dt + \sum_{i=0}^n \sum_{j=0, j \neq i}^n 1_{\{j > i\}} \lambda_{i,j}(t) \left( h_{i,j} \mathbb{E} \left( e^{J^b} - 1 \right) dt - \mathbb{E} \left( e^J - 1 \right) dt \right) \\ &\quad + \sigma_t dW_t^\beta + \sum_{i=0}^n \sum_{j=0, j \neq i}^n 1_{\{j > i\}} \left( (e^{J^b} - 1) dN_{i,j}^h(t) - h_{i,j} \lambda_{i,j}(t) \mathbb{E} \left( e^{J^b} - 1 \right) dt \right). \end{aligned}$$

■

Given that under the risk neutral measure, all assets earn on average the risk free rate, one easily obtains the condition that ensures that  $L_t$  defines a pricing measure:

**Corollary 5.3.** *An equivalent measure defined by the Radon-Nykodym derivative (36) is a risk neutral measure if and only if  $(\beta_t)_{t \geq 0}$ ,  $h_{i,j} > -1$  for  $i, j = 0, \dots, n$  and  $\nu^b(\cdot)$  satisfy the following constraint:*

$$\delta_t \bar{r}^\top = (\mu_t - \sigma_t \beta_t) + \sum_{i=0}^n \sum_{j=0, j \neq i}^n 1_{\{j > i\}} \lambda_{i,j}(t) \left( h_{i,j} \mathbb{E} \left( e^{J^b} - 1 \right) - \mathbb{E} \left( e^J - 1 \right) \right).$$

## 6 European option pricing

Let us consider European call and put options of maturity  $T$ , written upon an underlying price process  $(S_t)_t$  as given by (17). In the following, we express their payoff and strike as functions of the log-return  $\ln(\frac{S_T}{S_0})$  and of the log-strike  $k = \ln\left(\frac{K}{S_0}\right)$ . We assume that the Markov chain process  $(\delta_t)_{t \geq 0}$  is observable.

For the sake of simplicity, the risk free rate is assumed constant. The available information is then carried by the filtration  $(\mathcal{F}_t)_{t \geq 0}$ . The prices at time  $t$  of call and put options, denoted by  $C(t, k, \delta_t)$  and  $P(t, k, \delta_t)$ , are functions of the log-strike  $k$ . It is well-known that prices are equal to their expected discounted payoffs under the risk neutral measure (using  $\mathbb{E}^{\mathbb{Q}}$  as notation), and therefore if the risk neutral density at time  $t \leq T$  of the log-return  $\ln \frac{S_T}{S_0} | \mathcal{F}_t$  is denoted by  $f_{t,T}(x, \delta_t)$ :

$$\begin{aligned} C(t, k, \delta_t) &= \mathbb{E}^{\mathbb{Q}} \left( e^{-r(T-t)} (S_0 e^{X_T} - K)_+ | \mathcal{F}_t \right) \\ &= S_0 \int_k^{+\infty} e^{-r(T-t)} (e^x - e^k) f_{t,T}(x, \delta_t) dx, \\ P(t, k, \delta_t) &= \mathbb{E}^{\mathbb{Q}} \left( e^{-r(T-t)} (K - S_0 e^{X_T})_+ | \mathcal{F}_t \right) \\ &= S_0 \int_{-\infty}^k e^{-r(T-t)} (e^k - e^x) f_{t,T}(x, \delta_t) dx \end{aligned}$$

where  $r \in \mathbb{R}^+$  is assumed to be the constant risk free rate. As  $C(\cdot)$  (resp.  $P(\cdot)$ ) tends to  $S_t$  (resp.  $-S_t$ ) when  $k \rightarrow -\infty$  (resp.  $k \rightarrow +\infty$ ),  $C(\cdot)$  and  $P(\cdot)$  are not square integrable with respect to  $k$  and their Fourier transforms are not defined. For this reason, we consider the modified call and put prices denoted by  $c(k) = e^{\epsilon k} C(t, k, \delta_t)$ ,  $p(k) = e^{\epsilon k} P(t, k, \delta_t)$ , for which the Fourier transform exists for some damping factor  $\epsilon$  ( $\epsilon > 1$  for the call and  $\epsilon < -1$  for the put). The Fourier transforms of  $c(k)$  and  $p(k)$  are defined as follows:

$$\begin{aligned} \mathcal{FC}(\omega) &= \int_{-\infty}^{\infty} e^{i\omega k} c(k) dk, \\ \mathcal{FP}(\omega) &= \int_{-\infty}^{\infty} e^{i\omega k} p(k) dk. \end{aligned}$$

Recalling that  $\Upsilon_{t,s}(\omega) = \mathbb{E}^{\mathbb{Q}} (e^{\omega X_s} | \mathcal{F}_t)$  follows from Proposition 3.1 with parameters under  $\mathbb{Q}$ , a direct calculation leads to the same expressions for  $\mathcal{FC}(\omega)$  and  $\mathcal{FP}(\omega)$ :

$$\mathcal{FC}(\omega) = \mathcal{FP}(\omega) = \frac{S_0 e^{-r(T-t)}}{(i\omega + \epsilon)^2 + (i\omega + \epsilon)} \Upsilon_{t,T}(i\omega + \epsilon + 1),$$

except that  $\epsilon$  is positive (resp. negative) for the call (resp. put). The values of call options are then obtained by inverting the Fourier transform:

$$C(t, k, \delta_t) = \frac{S_0 e^{-\epsilon k - r(T-t)}}{\pi} \int_0^{\infty} e^{-i\omega k} \frac{\Upsilon_{t,T}(i\omega + \epsilon + 1)}{(i\omega + \epsilon)^2 + (i\omega + \epsilon)} d\omega. \quad (37)$$

As the same expressions hold for puts, except that  $\epsilon < 0$ , we exclusively focus on call options in the remainder of this section. The naive approach consists in calculating numerically the integral in equation (37). Setting  $\omega_m = \Delta_\omega(m-1)$  and letting  $M$  be the number of steps used in the Discrete Fourier Transform (DFT) as in Carr and Madan (1999), an approximation of the call price is then given by:

$$C(t, k, \delta_t) \approx \frac{S_0 e^{-\epsilon k - r(T-t)}}{\pi} \sum_{m=1}^M e^{-i\omega_m k} \varrho_m \left[ \frac{\Upsilon_{t,T}(i\omega_m + \epsilon + 1)}{(i\omega_m + \epsilon)^2 + (i\omega_m + \epsilon)} \right] \Delta_\omega, \quad (38)$$

where  $\varrho_m = \frac{1}{2} \mathbf{1}_{\{m=1\}} + \mathbf{1}_{\{m \neq 1\}}$ . An judicious choice for the discretization steps in equation (38), allows us to use a Fast Fourier Transform algorithm to speed up calculations. This point is detailed in the following proposition.

**Proposition 6.1.** Let  $M$  be the number of steps used in the Discrete Fourier Transform (DFT) and  $\Delta_k = \frac{2k_{max}}{M-1}$ , be the step of discretization. Let us denote  $\varrho_m = \frac{1}{2}1_{\{m=1\}} + 1_{\{m \neq 1\}}$ ,  $\Delta_\omega = \frac{2\pi}{M\Delta_k}$  and  $\omega_m = (m-1)\Delta_\omega$ . The values of  $C(t, k, \delta_t)$  at points  $k_j = -\frac{M}{2}\Delta_k + (j-1)\Delta_k$  are approximated by

$$C(k_j) \approx \frac{2S_0 e^{-\epsilon k - r(T-t)}}{M \Delta_k} \times \operatorname{Re} \left( \sum_{m=1}^M \delta_m \left( \frac{\Upsilon_{t,T}(i\omega_m + \epsilon + 1)}{(i\omega_m + \epsilon)^2 + (i\omega_m + \epsilon)} (-1)^{m-1} \right) e^{-i\frac{2\pi}{M}(m-1)(j-1)} \right). \quad (39)$$

This last relation can be computed with a fast Fourier transform algorithm.

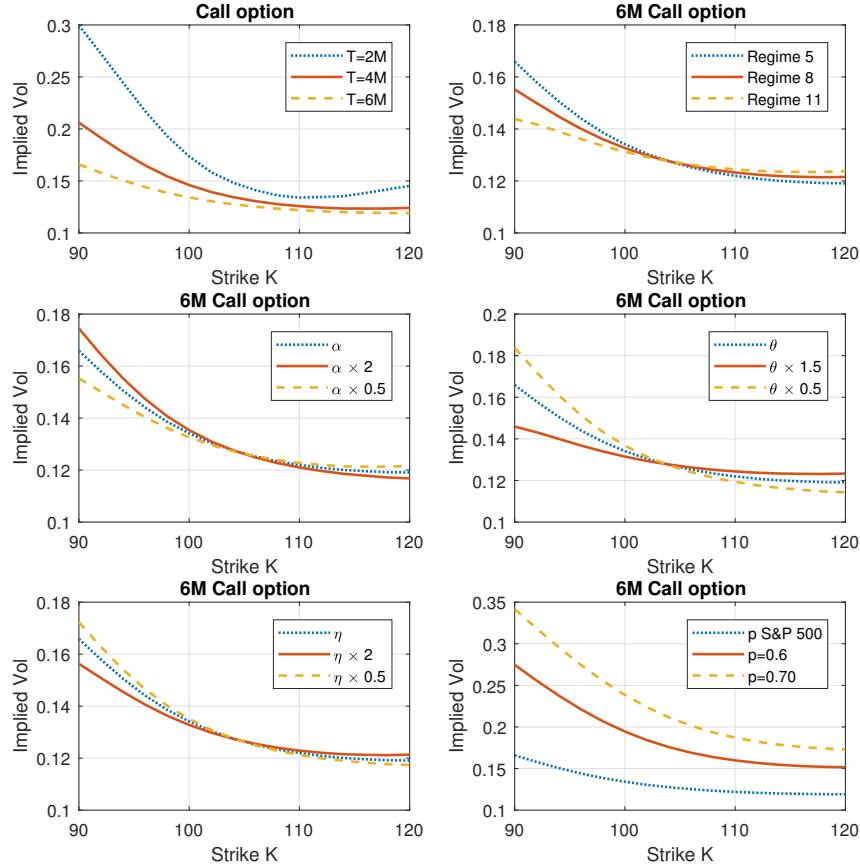


Figure 3: Implied volatility of European call options, for different strikes and sets of parameters. The risk free rate is constant and equal to 2%.

To conclude this section, we study the sensitivity of the implied volatility surface to modifications of the jump dynamics. We use the parameters reported in Table 2 to evaluate European options. The initial stock price is  $S_0 = 100$  and the strike  $K$  ranges from 90 to 120. The upper left graph of Figure 3 presents the smiles of implied volatility for 1, 2 and 6 months options when  $\delta_0 = e_5$ . These smiles are asymmetric and the curvature is inversely proportional to the maturity. The upper right graph emphasizes the importance of the initial state on the smile. In a high regime, the probability that a shock occurs increases. As jumps are on average negative, “In the Money” call options become cheaper and the smile of volatility flattens. The mid left graph shows that reducing  $\alpha$  has a similar impact: as the Markov chain

reverts slower than in the initial case, the risk of observing several negative jumps increases, causing a fall in call prices. On the other hand, reducing  $\theta$  or  $\eta$  raises prices and implied volatility as jumps (on average negative) are less frequent. The lower right graph of Figure 3 underlines the important role of the parameter  $p$  on the volatility surface. When  $p$  increases, jumps are mostly positive and the probability that the option will be exercised at expiry moves up, leading to higher prices and higher implied volatility.

## 7 Perpetual Binary options

To the best of our knowledge, there are no explicit results about hitting times of a threshold for a self-exciting jump diffusion. Therefore, pricing path dependent options in presence of jumps clustering leads to the need to run Monte Carlo simulations. Interest rates being low in Europe and US since the end of the credit crunch of 2008, the pricing of perpetual path dependent options requires then to consider a relatively long time horizon for simulations, which is computationally time consuming. Working with the SESJD model offers in this case an interesting alternative. Jumps being triggered by changes of regime of a hidden Markov chain, we can find the Laplace transform of the SESJD model hitting times with the fluid embedding technique of Rogers (1994) and adapted by Jiang and Pistorius (2008) for phase-type jump diffusions. In this section, we illustrate how this method can be applied to the pricing of perpetual binary options.

Recall that  $X_t$  is the log-return of  $S_t$  as defined in equation (22) and assume that the Markov chain  $(\delta_t)_{t \geq 0}$  is observable. Perpetual high or low binary options have an infinite time horizon and respectively deliver a payoff equal to  $\delta_\tau^\top \bar{h}$  where  $\bar{h} = (\bar{h}_0, \dots, \bar{h}_n) \in \mathbb{R}^{n+1}$  where the stopping time  $\tau$  for a low and high binary option is respectively defined by  $\tau = \inf\{t : X_t \leq k\}$  or  $\tau = \inf\{t : X_t \geq k\}$  for a certain level  $k \in \mathbb{R}^+$ . The value of these binary options is equal to the expected discounted cash-flow under the risk neutral measure (to lighten notations, the expectation under the risk neutral measure is denoted by  $\mathbb{E}(\cdot)$  in the following):

$$\begin{aligned} B^{high}(X_t, \delta_t) &= \mathbb{E} \left( e^{-\int_t^\tau \delta_s^\top \bar{r} ds} (\delta_\tau^\top \bar{h}) | \mathcal{F}_t \right), \quad \tau = \inf\{t : X_t \geq k\} \\ B^{low}(X_t, \delta_t) &= \mathbb{E} \left( e^{-\int_t^\tau \delta_s^\top \bar{r} ds} (\delta_\tau^\top \bar{h}) | \mathcal{F}_t \right), \quad \tau = \inf\{t : X_t \leq k\} \end{aligned}$$

where  $\bar{r} = (r_0, \dots, r_n) \in \mathbb{R}^{n+1}$  is the risk free rate in each phase. From Buffington and Elliott (2002), we know that

$$\mathbb{E} \left( e^{-\int_t^T \delta_s^\top \bar{r} ds} \delta_T | \mathcal{F}_t \right) = \delta_t^\top e^{(Q_0 - \text{diag}(\bar{r}))(T-t)}.$$

Conditionally to the filtration of the hitting time  $\tau$ , we then have that

$$\begin{aligned} \mathbb{E} \left( e^{-\int_t^\tau \delta_s^\top \bar{r} ds} (\delta_\tau^\top \bar{h}) | \mathcal{F}_t \right) &= \mathbb{E} \left( \mathbb{E} \left( e^{-\int_t^\tau \delta_s^\top \bar{r} ds} \delta_\tau^\top | \mathcal{F}_t \vee \tau \right) \bar{h} | \mathcal{F}_t \right) \\ &= \mathbb{E} \left( \delta_t^\top e^{(Q_0 - \text{diag}(\bar{r}))(\tau-t)} \bar{h} | \mathcal{F}_t \right). \end{aligned}$$

Next, let us define the stopping time  $\zeta$  which is the first jumping time of a process  $N_t^\tau$  with intensity  $\lambda_t^\tau = \int_0^t \delta_s^\top \bar{r} ds$ . If we denote  $Q_r = Q_0 - \text{diag}(\bar{r})$ , the above expectation may then be rewritten as

$$\mathbb{E} \left( \delta_t^\top e^{Q_r(\tau-t)} \bar{h} | \mathcal{F}_t \right) = \mathbb{E} \left( \mathbf{1}_{\tau \leq \zeta} (\delta_\tau^\top \bar{h}) | \mathcal{F}_t \right),$$

and therefore, the following equality holds:

$$\mathbb{E} \left( e^{-\int_t^\tau \delta_s^\top \bar{r} ds} (\delta_\tau^\top \bar{h}) | \mathcal{F}_t \right) = \mathbb{E} \left( \mathbf{1}_{\tau \leq \zeta} (\delta_\tau^\top \bar{h}) | \mathcal{F}_t \right).$$

We denote by  $(\gamma_t)_{t > 0}$  an irreducible continuous time Markov chain, defined on a finite state space  $E \cup \partial$  where  $\partial$  is an absorbing state ( $\gamma_t$  enters this state at time  $\zeta$ , when  $N_t^\tau$  jumps).  $\gamma_t$  is a vector that takes

values in the set of unit vectors of dimension  $3(n+1)+1$ . The first  $3(n+1)$  elements correspond to  $E$  and the last one to  $\partial$ . And we define a new stochastic process  $A_t$  which is the fluid embedding process of  $X_t$ , as follows:

$$A_t = A_0 + \int_0^t \gamma_s^\top \bar{m} ds + \int_0^t \gamma_s^\top \bar{s} dW_s$$

where  $\bar{m}$  and  $\bar{s}$  are vectors of dimension  $3(n+1)+1$ , that will be defined later. The generator of  $\gamma_t$ , restricted to  $E$ , is the next  $3(n+1) \times 3(n+1)$  matrix:

$$Q_\gamma = \begin{pmatrix} -D^- & D^- & O \\ B^- & C & B^+ \\ O & D^+ & -D^+ \end{pmatrix}$$

where  $D^-$ ,  $D^+$ ,  $B^-$ ,  $B^+$  are  $(n+1) \times (n+1)$  matrix defined by

$$\begin{aligned} D^- &= \text{diag}(-\rho^- 1_{n+1}) \quad , \quad D^+ = \text{diag}(\rho^+ 1_{n+1}) \\ B^- &= ((1-p) q_{i,j} 1_{j>i})_{i,j=1,\dots,n} \quad , \quad B^+ = (p q_{i,j} 1_{j>i})_{i,j=1,\dots,n} \\ C &= (q_{i,j} 1_{j \leq i})_{i,j=1,\dots,n} - \text{diag}(\bar{r}) \end{aligned}$$

with  $O$  being a  $(n+1) \times (n+1)$  null matrix. Notice that the generator  $Q_\gamma$  of  $\gamma_t$  is totally different from the generator of Jiang and Pistorius (2008) because jumps occur in our model exclusively when  $\delta_t$  changes of regime and not between times of regime shifts. The state space  $E$  can be fractioned as  $E = E^+ \cap E_0 \cap E^-$ . States in  $E^+$  and  $E^-$  respectively correspond to positive and negative jumps, whereas states of  $E_0$  are inherited from  $\delta_t$ . The vectors  $\bar{m}$  and  $\bar{s}$  are defined by:

$$\begin{aligned} \bar{m} = (m_i)_{i=0:3n+3} &= \begin{cases} -1 & i \in \{0, \dots, n\} \\ \tilde{\mu}_{i-n-1} & i \in \{n+1, \dots, 2n+1\} \\ +1 & i \in \{2n+2, \dots, 3n+2\} \\ 0 & i = 3n+3 \end{cases} \\ \bar{s} = (s_i)_{i=0:3n+3} &= \begin{cases} 0 & i \in \{0, \dots, n\} \\ \sigma_{i-n-1} & i \in \{n+1, \dots, 2n+1\} \\ 0 & i \in \{2n+2, \dots, 3n+2\} \\ 0 & i = 3n+3 \end{cases} \end{aligned}$$

where  $\tilde{\mu}_{i-n-1}$  is the drift of  $X_t$  when  $\delta_t = e_{i-n-1}$ , such as defined by equation (21). Within this approach, any sample path of  $X_t$  that represents discontinuities at jump times may be converted into a continuous path of  $A_t$ . Let us denote

$$\begin{aligned} T_0(t) &:= \int_0^t 1_{\{\gamma_s \in E_0\}} ds \quad , \\ T_0^{-1}(u) &:= \inf\{t \geq 0 : T_0(t) > u\} \quad , \end{aligned}$$

where  $T_0(t)$  is the total time spent in  $E_0$  by the Markov chain  $\gamma_t$  up to time  $t$ . By construction  $(A \circ T_0^{-1}, \gamma \circ T_0^{-1})$  has the same distribution as  $(X_t, \delta_t)$ . Then

$$\begin{aligned} \mathbb{E} \left( e^{-\int_t^\tau \delta_s^\top \bar{r} ds} (\delta_\tau^\top h) | \mathcal{F}_t \right) &= \mathbb{E} \left( 1_{\tau \leq \zeta} (\delta_\tau^\top \bar{h}) | \mathcal{F}_t \right) \\ &= \mathbb{E} \left( 1_{\bar{\tau} \leq \zeta} (\gamma_{\bar{\tau}}^\top \tilde{h}) | \mathcal{F}_t \right) \end{aligned}$$



where  $\tilde{\tau}$  is respectively for a binary low and high options given by

$$\begin{aligned}\tilde{\tau} &= \inf \{t \geq 0; \gamma_t \in E_0 \text{ and } A_t \leq k\} \\ \tilde{\tau} &= \inf \{t \geq 0; \gamma_t \in E_0 \text{ and } A_t \geq k\}\end{aligned}$$

and

$$\tilde{h} = \left(\tilde{h}_i\right)_{i=0:3n+3} = \begin{cases} \bar{h}_{i-n-1} & i \in \{n+1, \dots, 2n+1\} \\ 0 & \text{otherwise} \end{cases}$$

We now define up-crossing and down-crossing ladders,  $\gamma_t^+$  and  $\gamma_t^-$  as follows

$$\gamma_t^+ = \gamma_{\tau_t^+} \quad \gamma_t^- = \gamma_{\tau_t^-},$$

where

$$\begin{aligned}\tau_t^+ &= \inf \{s \geq 0 : A_s > t\}, \\ \tau_t^- &= \inf \{s \geq 0 : A_s < t\}.\end{aligned}$$

By construction,  $\gamma_t^+$  and  $\gamma_t^-$  are Markov chains with respective state spaces  $E^+ \cup E^0$  and  $E^- \cup E^0$ . Their generators are  $2(n+1) \times 2(n+1)$  matrix denoted by  $Q^+$  and  $Q^-$ . And the initial distributions are  $(n+1) \times 2(n+1)$  matrix,  $\eta^+$  and  $\eta^-$  such that

$$\begin{aligned}\eta^+(i, j) &= P(\gamma_0^+ = e_j, \tau_0^+ \leq \zeta | \gamma_0 = e_i) \quad e_i \in E^-, e_j \in E^0 \cup E^+, \\ \eta^-(i, j) &= P(\gamma_0^- = e_j, \tau_0^- \leq \zeta | \gamma_0 = e_i) \quad e_i \in E^+, e_j \in E^- \cup E^0,\end{aligned}$$

Let  $\Sigma = \text{diag}((\bar{s}_i)_{i=0:3n+2})$  and  $V = \text{diag}((\bar{m}_i)_{i=0:3n+2})$  be  $3(n+1) \times 3(n+1)$  matrices.

**Proposition 7.1.**  $(\eta^+, Q^+, \eta^-, Q^-)$  are solutions of the following matrix equations

$$\frac{1}{2}\Sigma^2 W^+ (Q^+)^2 - V W^+ Q^+ + Q_\gamma W^+ = O, \quad (40)$$

$$\frac{1}{2}\Sigma^2 W^- (Q^-)^2 + V W^- Q^- + Q_\gamma W^- = O, \quad (41)$$

where  $O$  is a  $3(n+1) \times 2(n+1)$  null matrix and

$$W^+ = \begin{pmatrix} \eta^+ \\ I_{(n+1)} & O_{(n+1)} \\ O_{(n+1)} & I_{(n+1)} \end{pmatrix} \quad W^- = \begin{pmatrix} I_{(n+1)} & O_{(n+1)} \\ O_{(n+1)} & I_{(n+1)} \\ \eta^- \end{pmatrix}.$$

Furthermore for  $X_t = x$  and  $k > x$ , we have that

$$\begin{aligned}B^{\text{high}}(X_t, \delta_t) &= \mathbb{E} \left( \mathbf{1}_{\tau_k^+ \leq \zeta} (\gamma_k^+ \bar{h}^+) | \mathcal{F}_t \right) \\ &= \gamma_t^\top W^+ \exp(Q^+(k - X_t)) h^+, \end{aligned} \quad (42)$$

and for  $X_t = x$  and  $k < x$ , we have that

$$\begin{aligned}B^{\text{low}}(X_t, \delta_t) &= \mathbb{E} \left( \mathbf{1}_{\tau_k^- \leq \zeta} (\gamma_k^- \bar{h}^-) | \mathcal{F}_t \right) \\ &= \gamma_t^\top W^- \exp(Q^-(X_t - k)) h^-, \end{aligned} \quad (43)$$

where

$$\bar{h}^+ = (\bar{h}_i^+)_{i=0:2n+2} = \begin{cases} \bar{h}_{i-n-1} & i \in \{n+1, \dots, 2n+1\} \\ 0 & \text{otherwise} \end{cases}.$$

and

$$\bar{h}^- = (\bar{h}_i^-)_{i=0:2n+2} = \begin{cases} \bar{h}_i^- & i \in \{0, \dots, n\} \\ 0 & \text{otherwise} \end{cases}.$$

**Proof of Proposition 7.1.** We just sketch the proof of relations (40) and (42) and refer to Jiang and Pistorius (2008) for details. By construction,

$$\mathbb{E} \left( \mathbf{1}_{\bar{\tau} \leq \zeta} (\gamma_{\bar{\tau}}^\top \bar{h}) | \mathcal{F}_t \right) = \mathbb{E} \left( \mathbf{1}_{\tau_k^+ \leq \zeta} (\gamma_k^+ \bar{h}^+) | \mathcal{F}_t \right).$$

On the other hand, given that  $\gamma_t^+$  is a Markov chain with generator  $Q^+$ , with an initial distribution  $W^+$ , we immediately infer that

$$\mathbb{E} \left( \mathbf{1}_{\tau_k^+ \leq \zeta} (\gamma_k^+ \bar{h}^+) | \mathcal{F}_t \right) = \gamma_t^\top W^+ \exp(Q^+(k - X_t)) h^+$$

(we substitute the initial time scale by the random clock  $\tau_t^+ = \inf\{s \geq 0 : A_s > t\}$ ). On the other hand,  $V_t^+ = \mathbb{E} \left( \mathbf{1}_{\tau_k^+ \leq \zeta} (\gamma_k^+ \bar{h}^+) | \mathcal{F}_t \right)$  clearly is a martingale, by its definition as conditional expectation. Let us denote

$$f(e_i, x) := e_i^\top W^+ \exp(Q^+(k - x)) h^+,$$

An application of Itô's lemma leads to the following relation for  $x \leq k$ :

$$\frac{1}{2} s_i^2 \frac{\partial^2}{\partial x^2} f(e_i, x) + m_i \frac{\partial}{\partial x} f(e_i, x) + \sum_j q_{ij}^+ (f(e_i, x) - f(e_j, x)) = 0.$$

Given that  $h^+$  is arbitrary, this last equation corresponds well to the system of equations (40). ■

Solving equations (40) and (41) is challenging. When the number of regimes is limited to 3 or 4, a method based on eigenvalues and eigenvectors of  $Q^+$  and  $Q^-$ , as proposed in Le Courtois and Su (2017) is numerically efficient. Unfortunately, for Markov chain with a high number of states, this approach becomes unstable because it requires to calculate the determinant of a badly conditioned matrix of big dimensions. However, it is possible to reduce the dimension of this problem if we recall that  $\Sigma$  et  $V$  are diagonal matrices:

**Proposition 7.2.** *The matrix  $Q^+$  may be rewritten as*

$$Q^+ = \begin{pmatrix} G_1 & G_2 \\ D^+ & -D^+ \end{pmatrix} \quad (44)$$

where  $G_1, G_2$  are  $(n+1) \times (n+1)$  matrices that satisfy the following system of matrix equations:

$$\begin{aligned} O_{(n+1)} &= \eta_{E^0}^+ G_1 + \eta_{E^+}^+ D^+ - D^- \eta_{E^0}^+ + D^- \\ O_{(n+1)} &= \eta_{E^0}^+ G_2 - \eta_{E^+}^+ D^+ - D^- \eta_{E^+}^+ \\ O_{(n+1)} &= \frac{1}{2} \text{diag}(\bar{\sigma}^2) (G_1^2 + G_2 D^+) - \text{diag}(\bar{\mu}) G_1 + B^- \eta_{E^0}^+ + C \\ O_{(n+1)} &= \frac{1}{2} \text{diag}(\bar{\sigma}^2) (G_1 G_2 - G_2 D^+) - \text{diag}(\bar{\mu}) G_2 + B^- \eta_{E^+}^+ + B^+ \end{aligned} \quad (45)$$

with  $\eta^+ = \{\eta_{E^0}^+, \eta_{E^+}^+\} \cdot \eta_{E^0}^+$  and  $\eta_{E^+}^+$  are here  $(n+1) \times (n+1)$  matrix.

**Proof of Proposition 7.2.** By definition of  $W^+$ , we have that

$$Q_\gamma W^+ = \begin{pmatrix} -D^- \eta_{E^0}^+ + D^- & -D^- \eta_{E^+}^+ \\ B^- \eta_{E^0}^+ + C & B^- \eta_{E^+}^+ + B^+ \\ D^+ & -D^+ \end{pmatrix}$$

and

$$VW^+ = \begin{pmatrix} -\eta_{E^0}^+ & -\eta_{E^+}^+ \\ \text{diag}(\bar{\mu}) & O \\ O & I_{(n+1)} \end{pmatrix}, \Sigma^2 W^+ = \begin{pmatrix} O_{(n+1)} & O_{(n+1)} \\ \text{diag}(\bar{\sigma}^2) & O_{(n+1)} \\ O_{(n+1)} & O_{(n+1)} \end{pmatrix}.$$

If we assume that

$$Q^+ = \begin{pmatrix} G_1 & G_2 \\ G_3 & G_4 \end{pmatrix}, (Q^+)^2 = \begin{pmatrix} G_1^2 + G_2 G_3 & G_1 G_2 + G_2 G_4 \\ G_1 G_3 + G_3 G_4 & G_2 G_3 + G_4^2 \end{pmatrix}$$

then

$$VW^+ Q^+ = \begin{pmatrix} -\eta_{E^0}^+ G_1 - \eta_{E^+}^+ G_3 & -\eta_{E^0}^+ G_2 - \eta_{E^+}^+ G_4 \\ \text{diag}(\bar{\mu}) G_1 & \text{diag}(\bar{\mu}) G_2 \\ G_3 & G_4 \end{pmatrix}$$

and

$$\Sigma^2 W^+ (Q^+)^2 = \begin{pmatrix} O_{(n+1)} & O_{(n+1)} \\ \text{diag}(\bar{\sigma}^2) (G_1^2 + G_2 G_3) & \text{diag}(\bar{\sigma}^2) (G_1 G_2 + G_2 G_4) \\ O_{(n+1)} & O_{(n+1)} \end{pmatrix}$$

Injecting these expressions in equation (40) leads to the result. ■

The same result holds for  $Q^-$ :

**Proposition 7.3.** *The matrix  $Q^-$  may be rewritten as*

$$Q^- = \begin{pmatrix} -D^- & D^- \\ G_3 & G_4 \end{pmatrix} \quad (46)$$

where  $G_3, G_4$  are  $(n+1) \times (n+1)$  matrices that satisfy the following system of matrix equations:

$$\begin{aligned} O_{(n+1)} &= \frac{1}{2} \text{diag}(\bar{\sigma}^2) (-D^- G_3 + G_3 G_4) + \text{diag}(\bar{\mu}) G_3 + B^- + B^+ \eta_{E^-}^- \\ O_{(n+1)} &= \frac{1}{2} \text{diag}(\bar{\sigma}^2) (D^- G_3 + G_4^2) + \text{diag}(\bar{\mu}) G_4 + C + B^+ \eta_{E^0}^- \\ O_{(n+1)} &= -\eta_{E^-}^- D^- + \eta_{E^0}^- G_3 - D^+ \eta_{E^-}^- \\ O_{(n+1)} &= \eta_{E^-}^- D^- + \eta_{E^0}^- G_4 + D^+ - D^+ \eta_{E^0}^- \end{aligned}$$

with  $\eta^- = \{\eta_{E^-}^-, \eta_{E^0}^-\}$  and  $\eta_{E^-}^-$  are here  $(n+1) \times (n+1)$  matrices.

To conclude this section, we study the sensitivity of perpetual binary options to modifications of the jump dynamics. We hereto use the parameters reported in Table 2. The upper left graph of Figure 4 shows prices of perpetual high binary options, for different log-strikes  $k$ . Whatever the regime of  $\delta_t$ , the price is inversely proportional to the strike because a higher strike postpones on average the exercise time of the binary option. For the same reason, the option value slightly falls when  $\delta_t$  is in a regime in which jumps have a higher probability to occur since in this case the exercise of the option will be delayed as jumps are on average negative. The right-hand graphs reveal indeed that increasing  $\eta$  or  $\theta$  drives down the prices, which can mainly be explained by the fact that the risk of observing several negative jumps increases and therefore the exercise of the option is delayed. On the contrary, if  $p$  increases, jumps are mostly positive and the period before the option exercise time becomes shorter, which causes a rise in option prices.

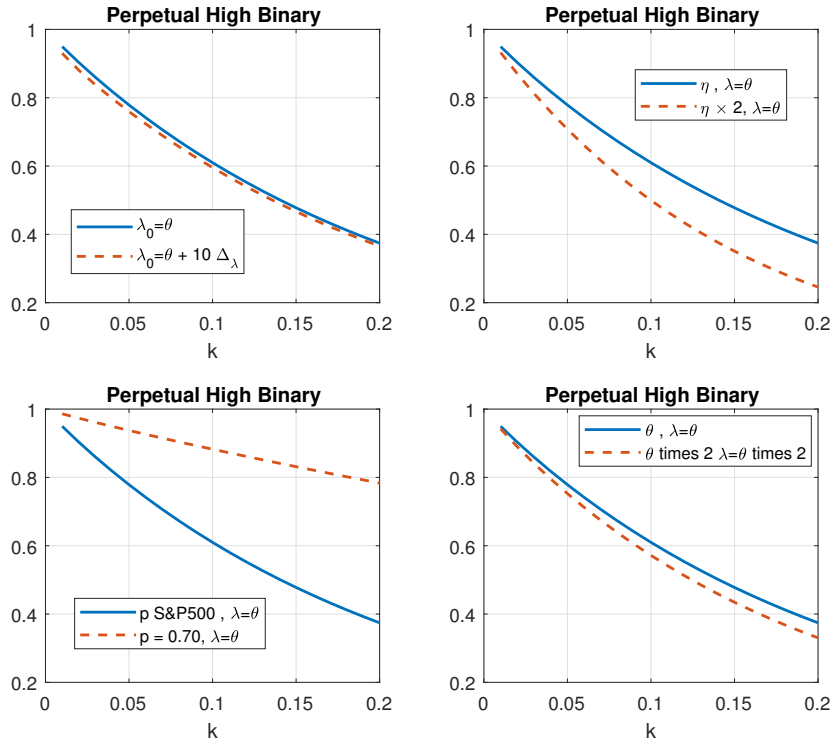


Figure 4: Comparison of binary option prices for different sets of parameters. The risk free rate is constant and set to 4%.

## 8 Conclusions

This article proposes the SESJD model as a new alternative to Hawkes processes, for modelling the jump clustering in financial time series. This approach is based on a switching regime jump diffusion. Whereas pure switching jump diffusions fail to duplicate the clustering of jumps because memory-less exponential random variables model the length of staying in a certain regime, the SESJD model does not present this drawback for two reasons. Firstly, we assume that jumps are synchronized to transitions times of a Markov chain with ordered states. Secondly, the matrix of transition probabilities is designed such that when the chain moves to a higher state, the probability of climbing again in the scale of states rises instantaneously.

The SESJD model is a very flexible model with several advantages. Contrary to Hawkes jump diffusions, the SESJD model is easy to calibrate with an enhanced Hamilton filter. Next, it is well defined under different measure changes, and in particular under a risk-neutral measure. Therefore, the model can be easily used for option pricing, both of European and exotic type. In particular, the fluid embedding technique of Rogers (1994) leads to a closed-form expression for the Laplace transform of the hitting time of a SESJD and therefore leads to explicit pricing formula for e.g. perpetual binary options.

## Appendix A

We illustrate the convergence of the pgf of the point process  $\tilde{N}_t$  towards the pgf of the Hawkes process  $N_t$  with a numerical example. We consider the following parameters for  $(N_t)_{t \geq 0}$ :  $\alpha = 8$ ,  $\theta = 10$ ,  $\eta = 5$ . To construct  $(\tilde{N}_t)_{t \geq 0}$ , we choose  $\Delta_t = \frac{1}{200}$ ,  $n = 10m$  and  $m = \{10, 20, 30, 40\}$ . Figure 5 shows the

pgf's of  $\tilde{N}_t$  and  $N_t$ , for  $t = 1$  year and  $\lambda_0 = \theta + \eta$ , whereas Table 3 reports the  $L^2$ -norms of the vector  $\mathbb{E}(u^{\tilde{N}_t} | \mathcal{F}_0) - \mathbb{E}(u^{N_t} | \mathcal{F}_0)$ , for  $u$  ranging from 0.01 to 1.01 by steps of 0.025 and for different values of  $\lambda_0$ . These results tend to confirm the convergence of the pgf of  $\tilde{N}_t$  to the pgf of  $N_t$  when the number of states of  $(\delta_t)_{t \geq 0}$  increases.

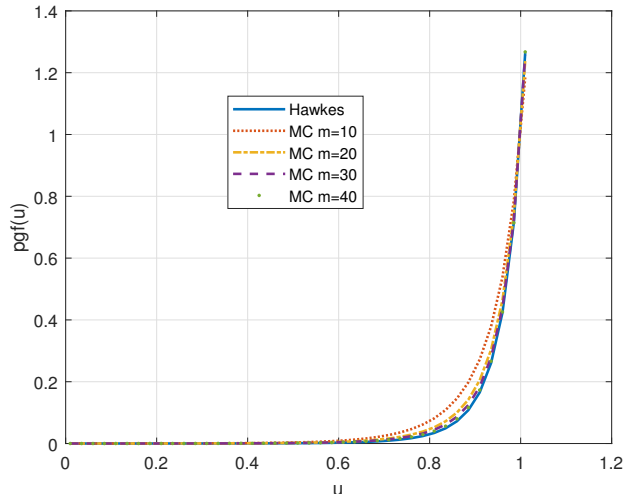


Figure 5: Comparison of  $\mathbb{E}(u^{\tilde{N}_t} | \mathcal{F}_0)$  and  $\mathbb{E}(u^{N_t} | \mathcal{F}_0)$  for  $u = 0$  to 1. Parameters of  $N_t$ :  $\alpha = 8$ ,  $\theta = 10$ ,  $\eta = 5$ . Parameters of  $\tilde{N}_t$ :  $\Delta_t = \frac{1}{200}$ ,  $t = 1$  year,  $n = 10m$  and  $m = 10, 20, 30, 40$  and  $\lambda_0 = \theta + \eta$ .

$\lambda_0$	$m = 10$	$m = 20$	$m = 30$	$m = 40$
$\theta$	0.2478	0.0921	0.0426	0.0199
$\theta + \frac{\lambda}{2}$	0.2672	0.1051	0.0512	0.0237
$\theta + \lambda$	0.2694	0.0980	0.0438	0.0201
$\theta + \frac{3\lambda}{2}$	0.2565	0.0899	0.0349	0.0169
$\theta + 2\lambda$	0.2375	0.0808	0.0290	0.0174

Table 3:  $L^2$ -norms of the vector  $\mathbb{E}(u^{\tilde{N}_t} | \mathcal{F}_0) - \mathbb{E}(u^{N_t} | \mathcal{F}_0)$ , for  $u$  ranging from 0.01 to 1.01 by steps of 0.025 for different values of  $\lambda_0$  and levels of discretization  $m$ .

## Appendix B

Hawkes jump are detected with a “peaks over threshold” (POT) procedure that is an enhanced version of the procedure of Embrechts et al. (2011). The discrete record of  $T$  observations of log-returns, equally spaced by a time step  $\Delta$  of one day of trading is denoted  $\{x_1, x_2, \dots, x_T\}$ . A jump is believed to occur if the return is above or below some thresholds. These thresholds, denoted  $g(\beta_1)$  and  $g(\beta_2)$ , depend on the lag between observations and on two confidence levels,  $\beta_1$   $\beta_2$ . To determine thresholds, we fit by log-likelihood maximization, a pure Gaussian process :  $x_i \sim \mu\Delta + \sigma W_\Delta$ . If  $\Phi(\cdot)$  denotes the cumulative distribution function (cdf) of a standard normal,  $g(\beta_1)$ ,  $g(\beta_2)$  are set to the  $\beta_1$  and  $\beta_2$  percentiles of the Brownian motion:  $g(\beta_i) = \sigma\sqrt{\Delta}\Phi^{-1}(\beta_i)$ . When a jump is detected, the variation of prices is assumed

equal to the jump size:

$$\left\{ (x_i - \mu\Delta) \sim J_i \quad (x_i - \mu\Delta) > g(\beta_1) \text{ or } < g(\beta_2) . \right.$$

Levels of confidence,  $\beta_1$  and  $\beta_2$  are optimized such that the skewness and the kurtosis of  $x_i$  for periods without jump are close to these of a normal distribution. For the S&P 500, we find that  $\beta_1$  and  $\beta_2$  are respectively equal to 94% and 91%. The skewness and kurtosis of returns for days without detected jumps are equal to 0.008 and 2.95. The volatility of the sample from which we eliminate jumps is 11.88%. Once that jumps are detected, the sample path of  $(\lambda_t)_t$  for a given set of parameters is approximated by:

$$\Delta\lambda_i = \alpha(\theta - \lambda_{i-1})\Delta + \eta 1_{jump\ at\ t_i} .$$

When  $\Delta$  is small, the probability of observing a jump in the  $i^{th}$  interval of time is equal to  $\lambda_i\Delta$ . Jumps and intensities can then be calibrated by maximizing the log-likelihood of jumps distribution and of  $\lambda_t$  as follows:

$$\begin{cases} (\rho^-, \rho^+, p) & = \arg \max \sum_{i=1}^n \log \nu(x_i | \rho^-, \rho^+, p) 1_{jump\ at\ t_i} \\ (\alpha, \eta, \theta, \lambda_0) & = \arg \max \sum_{i=1}^n ((\log(\lambda_i\Delta)) 1_{jump\ at\ t_i} + (\log(1 - \lambda_i\Delta)) 1_{no\ jump\ at\ t_i}) \end{cases} ,$$

where  $\nu(\cdot)$  is the pdf of double-exponential jumps. The results of this calibration are reported in table 2.

## Acknowledgement

We thank the anonymous referee and the editor for their comments. Donatien Hainaut also thanks the Chair ‘‘Data Analytics and Models for insurance’’ of BNP Paribas Cardiff, hosted by ISFA (Universit  Claude Bernard, Lyon) and managed by the ‘‘Fondation Du Risque’’. Griselda Deelstra acknowledges support of the FNRS grant PDR ‘‘Risk management in finance and insurance’’ 2016-2018.

## References

- [1] Abate J. , Whitt. W. 2006. "A Unified Framework for Numerically Inverting Laplace Transforms." *INFORMS Journal of Computing*, 18.4 , 408-421.
- [2] Ait-Sahalia Y., Cacho-Diaz J., Laeven R.J.A. 2015. Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, 117, 585-606.
- [3] Bacry E., Delattre S., Hoffmann M., Muzy J.F., 2013. Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, 13, 65–77.
- [4] Bacry E., Mastromatteo I., Muzy J.F. 2015. Hawkes Processes in Finance. *Market Microstructure and Liquidity* 1(1), pp 1-59.
- [5] Boswijk P., Laeven R.J.A, Lalu A. 2015. Asset returns with self-exciting jumps: option pricing and estimation with a continuum of moments. *Working paper*.
- [6] Buffington and Elliott (2002) American options with regime switching. *International Journal of Theoretical and Applied Finance*, 5, 497-514.
- [7] Boyarchenko S., Levendorskii S. 2009. American Options in Regime-Switching Models. *SIAM Journal on Control and Optimization*, 48:3, 1353-1376.
- [8] Carr. P. Madan D. 1999. Option Valuation Using the Fast Fourier Transform. *Journal of Computational Finance*, 2, 61-73.

- [9] Carr P., Wu L. 2016. Leverage Effect, Volatility feedback, and self-exciting market disruptions. *Forthcoming in Journal of Financial and Quantitative Analysis*.
- [10] Chavez-Demoulin V., McGill J., 2012. High-frequency financial data modeling using Hawkes processes. *Journal of Banking and Finance*, 36, 3415–3426. 6.
- [11] Chen K., Poon S.-H. 2013. Variance swap premium under stochastic volatility and self-exciting jumps. *working paper SSRN-id2200172*.
- [12] Chourdakis K. 2005. Switching Lévy models in continuous time: Finite distributions and option pricing. *Working paper SSRN: 838924*.
- [13] Dassios, A., Jang, J. 2003. Pricing of catastrophe reinsurance and derivatives using the Cox process with shot noise intensity. *Finance and Stochastic*, 7, 73–95.
- [14] Dassios A., Zhao H. 2011. A dynamic contagion process. *Advances in Applied Probability*, 43, 814–846
- [15] Dassios, A. and Zhao, H. 2014. A Markov chain model for contagion. *Risks*, 2(4):434–455.
- [16] Embrechts P., Liniger T., Lu L., 2011, Multivariate Hawkes processes: an application to financial data. *Journal of Applied Probability*, 48 (A), 367–378.
- [17] Errais E., Giesecke K., Goldberg L., 2010. Affine Point Processes and Portfolio Credit Risk. *SIAM Journal on Financial Mathematics*, 1, 642-665.
- [18] Ficura M. 2015. Modelling Jump Clustering in the Four Major Foreign Exchange Rates Using High-Frequency Returns and Cross-Exciting Jump Processes. *Procedia Economics and Finance* 25 208 – 219
- [19] Fulop A., Li J., Yu J. 2015. Self-exciting jumps, learning, and asset pricing implications. *Review of Financial Studies*, 28, 876-912. DOI: 10.1093/rfs/hhu078
- [20] Guidolin M., Timmermann A. 2008. International Asset Allocation under Regime Switching, Skew, and Kurtosis Preferences. *Review of Financial Studies*. 21 (2), 889-935.
- [21] Hainaut D. 2016 a. A model for interest rates with clustering effects. *Quantitative Finance*, 16(8), 1203-1218.
- [22] Hainaut D. 2016 b. Impact of volatility clustering on equity indexed annuities. *Insurance: Mathematics and Economics*, 71, 367-381.
- [23] Hainaut D., Colwell D. 2016. A structural model for credit risk with switching processes and synchronous jumps. In: *The European Journal of Finance*, Vol. 22, no.11, p. 1040-1062.
- [24] Hainaut D., Moraux F. 2018. Hedging of options in presence of jump clustering. Forthcoming in *Journal of Computational Finance*.
- [25] Hamilton J.D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57 (2), 357-384.
- [26] Hawkes A. 1971. Spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society Series B*, 33, 438-443. DOI : 10.1093/biomet/58.1.83
- [27] Heston, S. L. 1993. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies*. 6 (2): 327–343.

- [28] Honda, T. 2003. Optimal portfolio choice for unobservable and regime-switching mean returns. *Journal of Economic Dynamics and Control* 28(1), 45–78.
- [29] Jiang Z., Pistorius M.R., 2008. On perpetual American put valuation and first-passage in a regime-switching model with jumps. *Finance and Stochastics*, 12(2), 331–355.
- [30] Kou, S. G. 2002. A jump-diffusion model for option pricing. *Management Science*, 48(8):1086–1101.
- [31] Kou S.G., Wang H. 2003. First passage times of a jump diffusion process. *Advances in Applied Probability*. 35(2)
- [32] Le Courtois O., Su X., 2017. Structural Pricing of CoCos and Deposit Insurance with Regime Switching and Jumps. *Working paper*, AFFI International Conference 2017.
- [33] Levendorskiĭ S. 2008. American and European options in multi-factor jump-diffusion models, near expiry. *Finance and Stochastics*, 12:4, 541-560.
- [34] Rangel J. 2011. Macroeconomic news, announcements, and stock market jump intensity dynamics. *Journal of Banking and Finance*, 35 (5), 1263-1276. DOI: 10.1016/j.jbankfin.2010.10.009
- [35] Rogers L.C.G. 1994. Fluid models in queueing theory and Wiener-Hopf factorization of Markov chains. *Annals of Applied Probability*, 4 (2), 390–413.
- [36] Schoutens W. 2006. Exotic options under Lévy models: An overview. *Journal of Computational and Applied Mathematics*, 189:1-2, 526-538.
- [37] Yong M. K., Xu S.W. 2017. Pricing Vulnerable Options with Jump Clustering. *The journal of futures markets*. 37:12, 1155-1178.