



# Hawkes-driven stochastic volatility models: goodness-of-fit testing of alternative intensity specifications with S&P500 data

Iacopo Raffaelli<sup>1</sup> · Simone Scotti<sup>2</sup> · Giacomo Toscano<sup>3</sup> 

Accepted: 15 July 2022

© The Author(s) 2022

## Abstract

We introduce a novel stochastic volatility model with price and volatility co-jumps driven by Hawkes processes and develop a feasible maximum-likelihood procedure to estimate the parameters driving the jump intensity. Using S&P500 high-frequency prices over the period May 2007–August 2021, we then perform a goodness-of-fit test of alternative jump intensity specifications and find that the hypothesis of the intensity being linear in the asset volatility provides the relatively best fit, thereby suggesting that jumps have a self-exciting nature.

**Keywords** Stochastic Volatility · Co-Jumps · Hawkes processes · Maximum likelihood estimation · Goodness-of-fit testing

**JEL codes:** C14 · C51 · C52 · C58

## 1 Introduction

The volatility is arguably the most important attribute of a traded asset. Indeed, in the keystone Black and Scholes model, the volatility undoubtedly represents the most important parameter that a trader has to estimate in order to price a derivative claim. At the same time, the availability of efficient estimates of the volatility is also crucial for the econometrician who wishes to infer and forecast market features. It then sounds ironic that, in spite of decades of research, the question about the “right” volatility model is still at the heart of concerns for financial institutions and researchers.

---

✉ Giacomo Toscano  
[giacomo.toscano@unifi.it](mailto:giacomo.toscano@unifi.it)

Iacopo Raffaelli  
[iacopo.raffaelli@sns.it](mailto:iacopo.raffaelli@sns.it)

Simone Scotti  
[simone.scotti@unipi.it](mailto:simone.scotti@unipi.it)

<sup>1</sup> Scuola Normale Superiore, Pisa, Italy

<sup>2</sup> Department of Economics and Management, University of Pisa, Pisa, Italy

<sup>3</sup> Department of Economics and Management, University of Firenze, Firenze, Italy

Starting from the seventies, each decade proposed at least one novel paradigm and revisited the previous ones, claiming that the solution of the so-called “volatility puzzle” was in sight. In the seventies, Black and Scholes proposed to model the dynamics of asset prices via a geometric Brownian motion with constant volatility, but the Black Monday crash of 1987 exposed the pitfalls of such model in broad daylight. In the early eighties, Engle (1982) introduced the discrete-time ARCH model, which is capable of reproducing time-varying volatility and clustering. The model was then generalized by Bollerslev (1986), just a few years later. In the same years, Hull and White (1987) independently introduced a continuous-time log-normal model to reproduce the non-deterministic evolution and large fluctuations of asset volatilities. The nineties were instead the decade of the Heston model (Heston, 1993), which features a stochastic volatility process, possibly correlated with the price to reproduce leverage effects. The works of Bates (1996), Bates (2000) and Barndorff-Nielsen and Shephard (2001) in the early 2000s characterized the decade of stochastic volatility models with jumps, in particular common jumps in the volatility and the price. Finally, the last decade has seen the growth of rough volatility models, see El Euch et al. (2018).

All the previous models have clear advantages, but also some drawbacks. Early setups failed to reproduce stylized facts such as the volatility smile and the persistency and mean-reversion properties of the volatility, whereas more recent models may require the estimation of a large number of parameters, thereby inducing the problem of overfitting, or may fail to satisfy the Markov property. In this paper we present a self-exciting model for the price-volatility evolution which shares a large part of the advantages of existing setups but, at the same time, is also able to reproduce the volatility behavior parsimoniously, that is, with a small number of parameters, each one with a clear financial interpretation.

As pointed, e.g., by El Euch et al. (2018), the market microstructure of financial assets is well captured by point processes with a remarkably high degree of self-excitement, that is, asset price changes are quite often explained by previous changes. Under some hypotheses, El Euch et al. (2018) showed that their rescaled setup converges to a continuous-path volatility model driven by a fractional Brownian motion. However, this result contrasts with the remarkable study of Todorov and Tauchen (2011), who pointed out that the volatility activity is strictly smaller than 2 and thus is not consistent with a Brownian driver, either standard or fractional with Hurst parameter smaller than  $1/2$ , see Woerner (2011). The answer to the “volatility puzzle” by Todorov and Tauchen is that the volatility is driven by a pure jump process. Additionally, the authors highlighted the important role of common jumps in the underlying and its volatility, a feature which allows for the presence of leverage effects, in the case of a negative correlation. This feature is captured by the model proposed by Barndorff-Nielsen and Shephard (2001) (hereafter denoted by BNS). The crucial role of co-jumps is also supported by the empirical results presented in Broadie et al. (2007), where it is shown that an affine model, like the one by Heston, is rejected by historical option prices (even when it features volatility jumps with constant intensity), while a stochastic volatility model with contemporaneous jumps (denoted by SVCJ) is better supported empirically. Common large price-volatility fluctuations represent an important feature shared by the continuous-time BNS model and the discrete-time models in the ARCH family. The novelty of ARCH models was in fact represented by the feedback effect that links larger past shocks on the underlying to a higher volatility level. This property is clearly not captured by the Heston model and rough volatility models.

Our goal is then to generalize the continuous-time BNS model in order to accommodate for the feedback effect from the underlying to the volatility. Accordingly, we assume that the jump intensity is not constant but increases after each jump, following a Hawkes process, see Hawkes (1971). Under this hypothesis, one obtains jump clustering, see Bernis and Scotti

(2020). Moreover, the replacement of the Poisson process with a Hawkes driver in the BNS model allows reproducing the persistency of the volatility. Up to some extent, this replacement covers the same gap existing between ARCH and GARCH processes.

Unfortunately, the use of Hawkes processes comes with a main issue. Differently from the Poisson process, the Hawkes process is not Markov, unless the couple represented by the point process and its intensity is considered. This in turn causes an increase in the dimension of the model and thus in the number of parameters to be inferred. A possible strategy to overcome this problem, introduced in option pricing theory, see Bernis et al., (2021), is to force the intensity of the Hawkes process to coincide with the volatility. This way one reduces the size of the process and, consequently, the parameters to be estimated, along with the ensuing numerical instabilities. A Kolmogorov-Smirnov test performed with VIX data in (Bernis et al., 2021, Sect. 2) provides support to this hypothesis. Moreover, this hypothesis is coherent with the results presented in the econometric literature. For instance, Pan (2002) stresses that an “important feature of the jump-risk premium is that it depends on the market volatility”. Moreover, based on Broadie et al. (2007, Figure 1), one may observe that the VIX index was relatively low and flat during the period 1988–93, except for the rise linked with first Gulf War. The same figure indicates that the decade 1988–1997 featured almost no large fluctuations, compared with the period 1998–2002, which was instead characterized by five big crises and an average level of the VIX index equal to at least twice the value of the decade 1988–97. Similar considerations emerge from (Horst and Xu, 2019, Figure 2), where a more recent period was taken into account. In particular, no large fluctuations of the VIX were observed over the periods 2002–2007 and 2012–2014, characterized by financial stability. Instead, the intervals 2008–2011 and 2015–2018, corresponding to turmoil phases, were characterized by the presence of a large number of jumps and a much larger average level of the VIX.

In this paper we achieve two main results. First, we derive a feasible maximum-likelihood procedure to estimate the parameters driving the jump intensity of Hawkes-driven models from asset prices. The effectiveness of the procedure is tested with simulated data. Secondly, we evaluate the goodness-of-fit of alternative jump intensity specifications with empirical data. To do so, we focus on the S&P 500 index, sampled at the 5-minute frequency, over the period May 1, 2007–August 6, 2021. It emerges that the relatively best fit is obtained when the intensity of the Hawkes process is a linear transformation of the variance process. This result is coherent with the findings by Bernis et al., (2021). Note that this goodness-of-fit assessment is performed in the spirit of the study by Callegaro et al. (2019) on energy data, but is based on a more robust econometric strategy.

The paper is organized as follows. In Sect. 2 we outline the Hawkes-driven BNS model, while in Sect. 3 we derive a feasible maximum-likelihood procedure to estimate the intensity parameters of the latter. Sects. 4 and 5 contain, respectively, numerical and empirical analyses. Sect. 6 concludes.

## 2 The model

This section is devoted to the presentation of the general setup that will be tested empirically in Sect. 5. Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, where  $\mathbb{P}$  represents the historical probability measure, supporting the following objects:

- a standard Brownian motion  $\{W_t\}_{t \geq 0}$ ;
- an auxiliary square-integrable martingale  $\{M_t\}_{t \geq 0}$ ;

- a marked point process  $\{N_t\}_{t \geq 0}$  characterized by its counting measure  $\nu(dt, d\ell, dz)$  defined on  $\mathbb{R}_+^2 \times \mathbb{R}_*$ , whose jump times and marks are denoted by  $\{T_i, L_i, Z_i\}_{i \geq 1}$ .

In addition to  $\{N_t\}_{t \geq 0}$ ,  $\nu(dt, d\ell, dz)$  gives rise to two compound processes:

- $J_t^L := \sum_{i \geq 1} L_i \mathbb{I}_{T_i \leq t}$ , representing the non-negative jumps of the variance;
- $J_t^Z := \sum_{i \geq 1} Z_i \mathbb{I}_{T_i \leq t}$ , standing for the jumps of the equity.

Last but not least, let us further assume that the compensator of the counting measure  $\nu(dt, d\ell, dz)$  writes  $\lambda_t dt \nu(d\ell, dz)$  where  $\lambda := \{\lambda_t\}_{t \geq 0}$  refers to the right-continuous form of the jump intensity, and where  $\nu(d\ell, dz)$  is a probability distribution on the measurable space  $(\mathbb{R}_+ \times \mathbb{R}_*, \mathcal{B}(\mathbb{R}_+ \times \mathbb{R}_*))$ , with marginals  $\nu_L(d\ell) := \int_{z \in \mathbb{R}_*} \nu(d\ell, dz)$  and  $\nu_Z(dz) := \int_{\ell \in \mathbb{R}_+} \nu(d\ell, dz)$ .

We specify the evolution of the log-price  $X := \{X_t\}_{t \geq 0}$  of the asset and its spot variance  $V := \{V_t\}_{t \geq 0}$  as follows:

$$\begin{cases} dX_t = \eta_t dt + \sqrt{V_t} dW_t - \frac{1}{2} V_t dt + dJ_t^Z \\ dV_t = \kappa(\phi - V_t) dt + dM_t + dJ_t^L \\ \lambda_t = \mu + \alpha \sum_{i \in \mathbb{N}} e^{-\beta(t-T_i)} \mathbb{I}_{T_i \leq t} + \theta V_t \end{cases}, \quad (1)$$

where  $\eta := \{\eta_t\}_{t \geq 0}$  represents the instantaneous drift of the log-price under  $\mathbb{P}$ ,  $V_0 = v > 0$  and  $X_0 = x \in \mathbb{R}$  denote the initial conditions,  $\beta, \kappa, \phi$  are positive constants and  $\mu, \alpha, \theta$  are non-negative constants.

## 2.1 Examples

The intensity process  $\lambda$  specified in (1) includes a large class of models studied in literature. We refer in particular to the following interesting instances, which will be tested empirically in Sect. 5.

### Poisson setup

If  $\alpha = \theta = 0$ , the intensity is constant and equal to  $\mu$ . In this case the model coincides with the BNS model, see Barndorff-Nielsen and Shephard (2001), provided that the two marks  $(Z, L)$  are the same, up to a leverage parameter. Similarly, if the mark  $L$  is identically zero (i.e., the volatility never jumps) and  $M$  is a Brownian motion correlated with  $W$ , one obtains the model by Bates (1996). Moreover, if the volatility is set to be constant, one finds the model by Merton (1976). Finally, if the law of the mark  $Z$  is assumed to be a double exponential, one gets the model by Kou (2002).

### Hawkes setup

If  $\theta = 0$ , the intensity is a Hawkes process, see Hawkes (1971), that is, the intensity starts at the value  $\mu$  and increases by a fixed amount. In light of the growing interest on clustering effects, the financial literature has recently focused on Hawkes processes, see, for instance, Bernis et al. (2018), El Euch et al. (2018), Fićura (2017) and the references in Bernis and Scotti (2020). The main drawback of this setup is that the process is intrinsically three-dimensional, since the triplet  $(X, V, \lambda)$  is Markovian whereas the couple  $(X, V)$  is not. Basically all previously cited models with constant intensity could be extended to the Hawkes paradigm.

**Table 1** The different jump intensity specifications studied in the paper under the model illustrated in (1)

setup \ parameters	$\mu$	$\alpha$	$\theta$
Poisson	Positive	Null	Null
Hawkes	Positive	Positive	Null
Pure self-exciting	Null	Null	Positive
Hybrid self-exciting	Positive	Null	Positive

### Pure self-exciting setup

A way to preserve the Markov property of the couple  $(X, V)$ , while adding the clustering effect, is to assume that the process  $V$  plays both the role of the volatility and the intensity of jumps. Accordingly, we assume that  $\mu = \alpha = 0$ , so that the intensity  $\lambda$  coincides with the variance  $V$ , up to a multiplicative constant. Under the mild hypothesis that the quadratic variation of the auxiliary martingale  $M$  is proportional to  $V$ , the model belongs to the exponentially affine class, see Duffie et al. (2000). The volatility process  $V$  belongs to the class of branching processes, see Dawson and Li, (2012) for the explicit representation. The first example of a pure self-exciting jump model is, to the best of our knowledge, the model by Jiao et al., (2017) for interest rates. A similar model has been applied to power markets and volatilities, see Jiao et al. (2018), Jiao et al. (2021). These last three models are based on alpha-stable processes. The main advantage of this framework is its parsimony, since the volatility process  $V$  plays a two-fold role and thus the number of parameters is importantly reduced.

### Hybrid self-exciting setup

The main drawback of the pure self-exciting setup, as the name suggests, is the perfectly endogenous nature of the fluctuations. By definition, external, i.e., non-financial, events play no role. To overcome this issue, in the hybrid setup we relax the constraint on  $\mu$ , i.e., we only assume that  $\alpha = 0$ , so that the intensity  $\lambda$  is affine in the variance  $V$ . The latter then belongs to the class of branching processes with immigration, see Dawson and Li, (2012). A recent application of this setup is proposed in Bernis et al., (2021).

We resume the previous examples in the Table 1

## 2.2 Main properties

We start by showing that the model satisfies the Markov property.

**Proposition 1** (Markov property) *The intensity process  $\lambda$  satisfies the Markov property, and hence is an autonomous process, if one of the three following conditions is verified:*

1. **Hawkes setup:**  $\theta = 0$ .
2. **Pure self-exciting setup:**  $\alpha = 0$  and the process  $M$  is such that  $V$  is a Markov process.
3. **Same mean-reversion speed case:**  $\kappa = \beta$  and  $M \equiv 0$ .

Moreover, in the last specification, the process  $\{N_t\}_{t \geq 0}$  is a marked Hawkes process with jump intensity satisfying

$$\lambda_t = \mu + \theta V_0 + \int_0^t \int_{\mathbb{R}_+} \int_{\mathbb{R}_*} e^{-\beta(t-s)} (\alpha + \theta \ell) v(ds, d\ell, dz)$$

and the process is stable if  $\beta > \alpha + \theta \bar{\ell}$ .

**Proof** For the self-exciting case, the result is immediate. For the Hawkes case, the result is deduced by the fact that  $\lambda$  is the strong solution of the third SDE in (1). The interesting case is the last one, where we assume, without loss of generality, that  $\alpha > 0$  and  $\theta > 0$ . We compute the differential equation satisfied by  $\lambda$ , that reads

$$\begin{aligned} d\lambda_t &= -\beta\alpha \left( \sum_{i \in \mathbb{N}} e^{-\beta(t-T_i)} \mathbb{I}_{T_i \leq t} \right) dt + \alpha dN_t + \theta dV_t \\ &= -\beta\alpha \left( \sum_{i \in \mathbb{N}} e^{-\beta(t-T_i)} \mathbb{I}_{T_i \leq t} \right) dt + \theta\kappa(\phi - V_t)dt + \theta dM_t + \left( \alpha dN_t + \theta dJ_t^L \right). \end{aligned}$$

The drift in the previous equation depends both on  $V$  and on the sequence of jumps times  $\{T_i\}_{i \in \mathbb{N}}$ . However, if  $\beta = \kappa$ , the previous differential equation can be written in the form

$$d\lambda_t = \beta(\mu + \theta\phi - \lambda_t)dt + \left( \alpha dN_t + \theta dJ_t^L \right), \quad (2)$$

that is,  $\lambda$  is a mean-reverting process driven by a marked point process with marks  $\{\alpha + \theta L_i\}_{i \geq 1}$ . The fact that  $\{N_t\}_{t \geq 0}$  is a marked Hawkes process is a direct consequence of (Bernis and Scotti 2020, Definition 5.1) with  $\hat{\alpha}(\ell, z) = \alpha + \theta\ell$ , see also (Bernis et al. 2018, Definition 2). The stability property is obtained by applying (Brémaud and Massoulié, 1996, Sect. 4, eq. (14)).  $\square$

The sharp-eyed reader may notice that the process  $\lambda$  satisfies an SDE similar to that satisfied by  $V$  in (1). A natural question then arises about the interest of considering two feedback effects to capture clustering. The next remark highlights the differences between these two feedback mechanisms, while the empirical analysis in Sect. 5 will test if both effects are needed or one is sufficient to achieve a good fit with empirical data.

**Remark 1** There are two main differences between the Hawkes and the pure self-exciting setups. The first difference relates to the size of jumps. Indeed, the latter is fixed and equal to  $\alpha$  in the Hawkes framework. From the financial standpoint, this means that the effect on the probability of a new jump is the same, independently of the level of the price and/or volatility. A dichotomy then arises between the standard evolution of the model, captured by continuous-path processes, and the occurrence of unusual events, described by jumps. Consequently, a potential challenge emerges in relation to the precise identification of the jump times and sizes. This difficulty is smoothed away in the self-exciting setup, where the intensity is proportional to the volatility.

The other main difference is more subtle. The intensity is a hidden process and shares this drawback with the volatility. However, the volatility can be accurately estimated with high-frequency data, see Aït-Sahalia and Jacod (2014) and references therein. Moreover, if one assumes that the model is exponentially affine (see Duffie et al. (2000)), then the variance-swap rate is affine in the spot variance (see Kallsen et al. (2011)), and this affine link can be exploited to empirically approximate volatility values indirectly. In fact, in the case of the S&P500, variance-swap rates are quoted on the market in the form of the VIX index. Mancino et al. (2020) test the affine link empirically and find a remarkable match. In sum, even if the volatility is theoretically hidden, it could be considered as (nearly) observable in practice. In contrast, the intensity is hard to be inferred, as many approximations and hypotheses may be needed.

From here onwards, the Markov property will be assumed and we will replace  $\kappa$  with  $\beta$  when both  $\alpha$  and  $\theta$  are strictly positive.

**Lemma 1** (Compensated version) *The process  $(X, V, \lambda)$  admits the following compensated version*

$$\begin{cases} dX_t = (\eta_t + \bar{z}\lambda_t) dt + \sqrt{V_t} dW_t - \frac{1}{2} V_t dt + d\tilde{J}_t^Z \\ dV_t = \kappa(\phi - V_t) dt + dM_t + dJ_t^L = [\kappa(\phi - V_t) + \bar{\ell}\lambda_t] dt + dM_t + d\tilde{J}_t^V, \\ d\lambda_t = \tilde{\beta}(\tilde{\mu} - \lambda_t) dt + \theta dM_t + (\alpha + \theta L_{N_t}) d\tilde{N}_t \end{cases}, \quad (3)$$

where  $(\tilde{N}, \tilde{J}^Z, \tilde{J}^V)$  is the compensated process associated with  $(N, J^Z, J^V)$  and  $\tilde{\beta} := \beta - \alpha - \theta\bar{\ell}$ ,  $\tilde{\mu} := \beta(\mu + \theta\phi)\tilde{\beta}^{-1}$ .

The proof is a direct computation and is then omitted, see also (Bernis et al., 2021, Lemma 1) for a similar result. The main interest of the previous lemma is to exploit it to demonstrate the actual speed of mean reversion. We resume the result in the next corollary, which is a direct application of the previous lemma and (Jiao et al., 2017, Proposition 3.7).

**Corollary 1** (Mean-reverting and ergodicity) *We have the following mean-reversion and ergodicity conditions for the variance and intensity processes.*

*Variance process:* The variance process  $V$  is mean-reverting if  $\kappa > \bar{\ell}$ . Its mean level reads  $\frac{\kappa\phi}{\kappa - \bar{\ell}}$ . The actual speed of mean reversion is  $\kappa - \bar{\ell}$ . Moreover, the process is exponentially ergodic under the same hypothesis.

*Intensity process:* Assume that  $\alpha > 0$  and  $\theta > 0$ . Then the intensity is mean-reverting if  $\kappa = \beta > \alpha + \theta\bar{\ell}$  and its mean level reads  $\frac{\beta(\mu + \theta\phi)}{\beta - \alpha - \theta\bar{\ell}}$ . The actual speed of mean reversion is  $\beta - \alpha - \theta\bar{\ell}$ . Moreover, the process is exponentially ergodic under the same hypothesis.

**Proposition 2** (Exponentially affine class) *The intensity process  $\lambda$  solution of (2) is a continuous-state branching process with immigration. As a consequence, the triplet  $(X, V, \lambda)$  is an exponentially affine model. Moreover, this characterization is preserved when adding the noise  $M$ , provided that its instantaneous quadratic variation is proportional to  $\lambda$  itself, that is,  $M_t = \sigma_M \int_0^t \sqrt{\lambda_s} dW_s^M$ , where  $W^M$  is an auxiliary Brownian motion, and the intensity process reads*

$$d\lambda_t = \beta(\mu + \theta\phi - \lambda_t) dt + \theta\sigma_M \sqrt{\lambda_t} dW_t^M + (\alpha dN_t + \theta dJ_t^L). \quad (4)$$

**Proof** First consider the case when  $M \equiv 0$ . The SDE satisfied by  $\lambda$  verifies the conditions in Dawson and Li, (2012) and then belongs to the class of continuous-state branching processes with immigration (the proof is essentially the same as in Bernis et al., 2021, Proposition 1). Moreover, it is non-negative. The affine property of the triplet  $(X, V, \lambda)$  is deduced from the characterization of the intensity process  $\lambda$  as a continuous-state branching process with immigration, see (Dawson and Li, 2006, Theorems 5.1, 5.2 and 6.2).

Consider now the case when  $M$  has instantaneous quadratic variation proportional to  $\lambda$ . Thanks to the Lévy characterization theorem, see for instance (Ikeda and Watanabe, 1998, Theorem II.7.1), we write  $M_t = \sigma_M \int_0^t \sqrt{\lambda_s} dW_s^M$ , where  $W^M$  is an auxiliary Brownian motion. The SDE (4) has a unique solution, see (Li and Ma, 2008, Proposition 2.1) and belongs to the class of continuous-state branching processes with immigration, in accordance with (Dawson and Li, 2012, Theorem 3.1). Also, we remark that  $(X, V, \lambda)$  is a Markov process. Finally, by applying Itô's lemma with jumps, we find the evolution of  $\lambda$  as given in (4).  $\square$

From here onwards, we will assume the affine structure.

**Proposition 3** (Positivity of the process) *The intensity process  $\lambda$  is non-negative. It is moreover strictly positive if and only if:*

*Hawkes setup:  $\mu > 0$  when  $\theta = 0$  and  $\alpha > 0$ .*

*Pure self-exciting setup:  $2\kappa\phi > \sigma_M^2\theta$  when  $\alpha = \mu = 0$  and  $\theta > 0$ .*

*General case:  $2\beta(\mu + \phi\theta) > \sigma_M^2\theta^2$  when  $\mu > 0$  and  $\theta > 0$  and  $\beta = \kappa$ .*

*The positivity property has to be considered in the probability sense, that is, the probability of the intensity reaching zero is either zero if the previous conditions are satisfied or is strictly positive otherwise.*

The proof of Proposition 3 is postponed to Appendix 1.

### 3 Feasible maximum-likelihood estimation

In this section we describe a feasible procedure to estimate the intensity  $\lambda$  appearing in (1) which is based on likelihood maximization. First of all, we obtain the log-likelihood function related to  $\lambda$ , assuming the knowledge of the volatility process and jump times and sizes. Then, given that the jumps and the volatility are hidden processes, we replace them with the corresponding estimated processes based on the following strategy. First, we detect jump times via the non-parametric test by Lee and Mykland (2008); secondly, we exploit the bipower variation estimator by Barndorff-Nielsen and Shepard (2004) and its numerical derivative to obtain non-parametric estimates of, respectively, the integrated and the spot variance.

**Proposition 4** *Assume to observe the point process  $N$ , whose intensity is the process  $\lambda$  in (1), until time  $t$ . The log-likelihood function is given by*

$$\mathcal{L}(t; \vartheta) = -\mu t + \frac{\alpha}{\beta} \sum_{i=1}^{N_t} \left( e^{-\beta(t-T_i)} - 1 \right) - \theta \int_0^t V_s ds + \sum_{i=1}^{N_t} \log \left( \mu + \alpha \sum_{j < i} e^{-\beta(T_i - T_j)} + \theta V_{T_i-} \right), \quad (5)$$

where  $\vartheta := [\mu, \alpha, \beta, \theta]$  is the parameter vector.

**Proof** Starting from the results on point processes by Rubin (1972), it is:

$$\mathcal{L}(t; \vartheta) = - \int_0^t \lambda_s(\vartheta) ds + \int_0^t \log \lambda_{s-}(\vartheta) dN_s.$$



Hence:

$$\begin{aligned}
 \mathcal{L}(t; \vartheta) &= - \int_0^t \left( \mu + \alpha \sum_{i \in \mathbb{N}} e^{-\beta(s-T_i)} \mathbb{I}_{T_i \leq s} + \theta V_s \right) ds \\
 &\quad + \int_0^t \log \left( \mu + \alpha \sum_{i \in \mathbb{N}} e^{-\beta(s-T_i)} \mathbb{I}_{T_i < s} + \theta V_{s-} \right) dN_s \\
 &= -\mu t - \alpha \sum_{i \in \mathbb{N}} \int_0^t e^{-\beta(s-T_i)} \mathbb{I}_{T_i \leq s} ds - \theta \int_0^t V_s ds \\
 &\quad + \sum_{i=1}^{N_t} \log \left( \mu + \alpha \sum_{j \in \mathbb{N}} e^{-\beta(T_i-T_j)} \mathbb{I}_{T_j < T_i} + \theta V_{T_i-} \right) \\
 &= -\mu t + \frac{\alpha}{\beta} \sum_{i=1}^{N_t} \left( e^{-\beta(t-T_i)} - 1 \right) - \theta \int_0^t V_s ds \\
 &\quad + \sum_{i=1}^{N_t} \log \left( \mu + \alpha \sum_{j < i} e^{-\beta(T_i-T_j)} + \theta V_{T_i-} \right).
 \end{aligned}$$

Note that the only technical step involves interchanging the summation and the integration at the second line. The interchange is a consequence of the fact that the total number of jumps is finite. In fact, since the activity of the jump process is finite, then all the summations are almost surely finite over the fixed interval  $[0, t]$ . We then apply the stochastic Fubini theorem, see e.g., (Eberlein and Kallsen, 2020, Theorem 3.29).  $\square$

We point out that the maximum-likelihood estimator is consistent and asymptotically normal, see Ozaki (1979) and Ogata (1988). In particular, the Fischer Information matrix is explicitly computed in Ozaki (1979) under the hypothesis that the jump size is constant. Bernis and Scotti (2020) provide the extension to the case of random a jump size, i.e. the marked Hawkes case. The interested reader could also refer to Kirchner (2017) or to Lemonnier and Vayatis (2014) for a new estimation method based on machine learning.

Clearly, the maximization of (5) is unfeasible, unless accurate estimates of the integrated volatility on  $[0, t]$ , i.e.,  $\int_0^t V_s ds$ , and the spot volatility values  $V_t$ ,  $t \in \{T_i\}_{i=1, \dots, N_t}$ , are available. The literature on financial econometrics offers very reliable tools to detect jumps and reconstruct the volatility, based on high-frequency data.

To detect the instants at which jumps occurred, one can apply the non-parametric test by Lee and Mykland (2008). To the best of our knowledge, this is the only test that allows detecting jumps on arbitrarily small intervals. We recall that the model (1) postulates price-volatility co-jumps, so it is sufficient to the test for the presence of jumps in the price. As the output of the test, we obtain a sequence  $\{T_i^*\}_{i=1, \dots, N_t^*}$  of reconstructed jump times. Specifically, assume that, for a sufficiently large value of the sample size  $n$ , the price process is observable on the equally-spaced grid with mesh equal to  $\delta_n := t/n$  over the interval  $[0, t]$ . The jump test can be then performed on all “small” intervals  $((j-1)\delta_n, j\delta_n]$ ,  $j = 1, \dots, n$ , and, if the  $i$ -th jump is detected on  $((j-1)\delta_n, j\delta_n]$ , we say that  $T_i^* = j\delta_n$ , while  $T_{i-}^* = (j-1)\delta_n$ . The reader may refer to Lee and Mykland (2008) for further details.

As for the estimation of the integrated volatility, we propose to use a non-parametric consistent estimator in the presence of jumps, namely the bipower variation by Barndorff-

Nielsen and Shepard (2004). Let  $t_{i,n} := i\delta_n$ ,  $i = 0, \dots, n$ . The bipower variation reads

$$BPV_{[0,t],n} := \frac{\pi}{2} \sum_{i=2}^n |X_{t_{i,n}} - X_{t_{i-1,n}}| |X_{t_{i-1,n}} - X_{t_{i-2,n}}|. \quad (6)$$

Barndorff-Nielsen et al. (2006) show that  $BPV_{[0,t],n}$  converges stably in law to  $\int_0^t V_s ds$  as  $n \rightarrow \infty$ , with rate  $n^{1/2}$ . A local consistent non-parametric estimator of the spot volatility can be obtained via the numerical differentiation of (6) (see Aït-Sahalia and Jacod (2014)) and reads, for  $s \in (t_{i-1,n}, t_{i,n}]$ ,

$$\widehat{V}_{s,n} := \frac{\frac{\pi}{2} \sum_{j=2}^{k_n} |X_{t_{i+j,n}} - X_{t_{i+j-1,n}}| |X_{t_{i+j-1,n}} - X_{t_{i+j-2,n}}|}{k_n \delta_n}, \quad (7)$$

where  $k_n$  is a sequence of positive integers such that  $k_n \rightarrow \infty$  and  $k_n \delta_n \rightarrow 0$  as  $n \rightarrow \infty$ . Note that we do not need to reconstruct the entire volatility trajectory on the grid of mesh size  $\delta_n$ , but we only need local estimates on the grid  $\{T_{i-}^*\}_{i=1, \dots, N_t^*}$ .

The feasible version of (5) then reads:

$$\begin{aligned} \mathcal{L}^*(t, n; \vartheta) = & -\mu t + \frac{\alpha}{\beta} \sum_{i=1}^{N_t^*} \left( e^{-\beta(t-T_i^*)} - 1 \right) - \theta BPV_{[0,t],n} \\ & + \sum_{i=1}^{N_t^*} \log \left( \mu + \alpha \sum_{j<i} e^{-\beta(T_i^*-T_j^*)} + \theta \widehat{V}_{T_i^*,n} \right). \end{aligned} \quad (8)$$

In general, if volatility values are accurately estimated and jumps times are carefully detected, one may obtain a reliable approximation of the values of the parameter vector  $\vartheta$  that maximize the unfeasible log-likelihood in (5) via the numerical maximization of its feasible version in (8), for a fixed time horizon  $t$  and a sufficiently large sample size  $n$ . However, as one can easily verify, an analytic solution to the maximization problem of the unobservable log-likelihood is attainable in some special cases. For instance, if  $\alpha = \theta = 0$  (i.e., the Poisson setup is assumed), then, for any  $t > 0$ ,

$$\operatorname{argmax}_{\mu \in \mathbb{R}^+} \mathcal{L}(t; \mu) = \frac{N_t}{t},$$

which can be approximated by

$$\hat{\mu} := \frac{N_t^*}{t}; \quad (9)$$

instead, if  $\mu = \alpha = 0$  (i.e., the pure self-exciting setup is assumed), then, for any  $t > 0$ ,

$$\operatorname{argmax}_{\theta \in \mathbb{R}^+} \mathcal{L}(t; \theta) = \frac{N_t}{\int_0^t V_s ds},$$

which could be approximated by

$$\hat{\theta} := \frac{N_t^*}{BPV_{[0,t],n}}. \quad (10)$$

## 4 Simulation study

In this section, we illustrate the results of a numerical exercise where we investigated the performance of the maximum-likelihood estimation procedure described in Sect. 3 with

simulated data from model (1). The main goal is to show that the procedure detailed in Sect. 3 leads to efficient estimates of the parameters appearing in the examples described in Sect. 2.1.

For the simulation, we made the following assumptions on the compound processes  $J^Z$  and  $J^L$ . The marks  $Z$  and  $L$  were assumed to follow two exponential distributions, with parameters  $\gamma_Z$  and  $\gamma_L$ , respectively. Further, the sign of the equity marks  $Z$  was assumed to be negative with probability 0.8, since, empirically, jumps are most often negative, see, e.g., Bernis et al., (2021). Finally, for all simulated paths, we set the initial conditions  $X(0) = \log(1000)$  and  $V(0) = \phi$ . Overall, we simulated observations from the following four parametric setups, each corresponding to one of the relevant jump intensity specifications illustrated in Sect. 2.1:

- Poisson setup:  $(\eta, \kappa, \phi, \mu, \gamma_Z, \gamma_L) = (0.01, 0.4, 0.05, 10, 0.05, 0.005)$ ;
- Hawkes setup:  $(\eta, \kappa, \phi, \mu, \alpha, \beta, \gamma_Z, \gamma_L) = (0.01, 0.4, 0.05, 7.5, 8.5, 8.4, 0.05, 0.005)$ ;
- Pure Self-Exciting setup:  $(\eta, \kappa, \phi, \theta, \gamma_Z, \gamma_L) = (0.01, 0.4, 0.05, 65, 0.05, 0.005)$ ;
- Hybrid Self-Exciting setup:  $(\eta, \kappa, \phi, \mu, \theta, \gamma_Z, \gamma_L) = (0.01, 0.4, 0.05, 6.5, 180, 0.05, 0.005)$ .

For each setup, we simulated 200 monthly trajectories of price-volatility observations. A month is assumed to encompass 21 days of 6.5 hours each, matching the actual length of a trading day on the US market. Observations were simulated on the equally-spaced grid of mesh size equal to 5 minutes. The simulation scheme for an Hawkes process is adapted from Dassios and Zhao (2013) and Bernis et al., (2021). The explicit algorithm is presented in Appendix 1.

For the detection of jump times and the estimation of integrated and spot volatility values, needed to compute the feasible log-likelihood (8), we used the largest  $n$  available, that is,  $n$  corresponding to the frequency  $\delta_n = 5$  minutes. The value of the tuning parameter  $k_n$  in (7) is set equal to 390. However, parameter estimates appear to be fairly robust to any  $k_n$  in the range 350 – 430. The use of relatively long local windows for the efficient estimation of the spot volatility is discussed, e.g., in Lee and Mykland (2008), Aït-Sahalia et al. (2013), Toscano and Recchioni (2021). Note that, to compute spot volatility values, we used “left” local windows of length  $k_n \delta_n$ , thus a sufficient number of extra observations was simulated at the beginning of each monthly trajectory to accommodate for the necessary burn-in period. For the implementation of the jump test by Lee and Mykland, we followed the practical recommendations given in Sections 1.3 and 1.4 of Lee and Mykland (2008). In particular, note that the test was performed at the 1% significance level.

In order to estimate the intensity parameters in the Poisson and pure-self exciting setups, we used, respectively, the closed-form estimators given in Eq. (9) and (10). Instead, in the case of the Hawkes and hybrid self-exciting frameworks, we resorted to numerical optimization. Specifically, we used the interior-point method for the maximization of the feasible log-likelihood in Eq. (8), with a strict-positivity constraint on the parameters to be estimated. As initial parameter values, we used the vectors  $(\mu = N_t^*/t, \theta = N_t^*/BPV_{n,[0,t]})$  and  $(\mu = N_t^*/t, \alpha = 1, \beta = 0.9)$ , respectively. However, the numerical optimization program appeared to be rather insensitive to the selection of initial parameter values. Table 2 reports the sample average of parameters estimates obtained over an increasing number of monthly samples. Sample standard deviations range from approximately 120% of the true value for sample estimates of  $\beta$  to approximately 60% for sample estimates of  $\mu$  in the Poisson framework and are quite stable with respect to the value of  $M$ . Based on Table 2, the performance of the feasible maximum-likelihood procedure proposed appears to be satisfactory, as estimates present relatively small biases, at least for  $M$  greater or equal than 100. These simulation

**Table 2** Sample average of parameter estimates obtained over an increasing number  $M$  of simulated monthly samples of the model (1) under four different intensity setups: Poisson (P), Hawkes (H), Pure Self-Exciting (PSE) and Hybrid Self-Exciting (HSE)

Setup	Parameter	True value	$M = 50$	$M = 100$	$M = 150$	$M = 200$
P	$\mu$	10	9.201	9.455	9.5800	9.880
H	$\mu$	7.5	6.492	7.403	7.840	7.777
	$\alpha$	8.5	10.132	9.013	8.908	8.757
	$\beta$	8.4	6.334	7.560	8.122	8.214
PSE	$\theta$	65	62.370	65.121	64.973	65.037
HSE	$\mu$	6.5	5.963	6.126	6.3872	6.561
	$\theta$	180	176.336	181.3146	179.011	179.578

results support the robustness of the empirical results shown in Sect. 5, where the sample used covers 163 months.

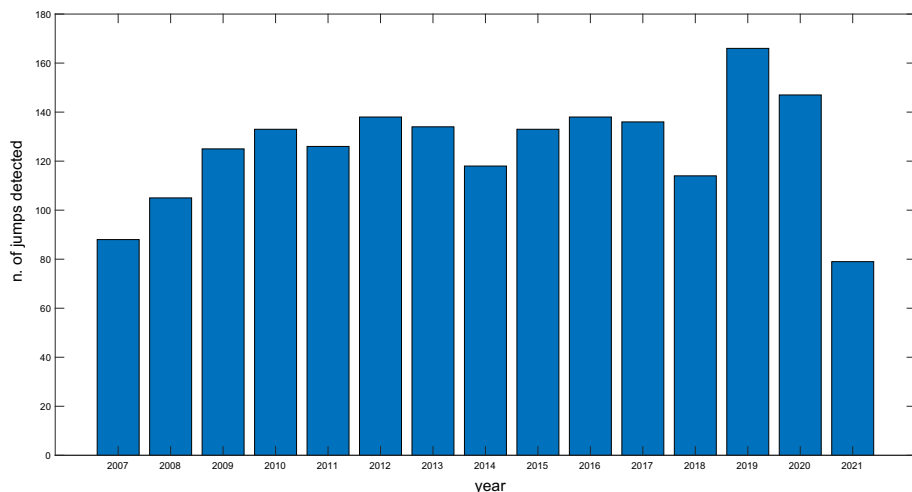
## 5 Empirical study

In this section, we illustrate the results of an empirical study where, based on the maximum-likelihood procedure described in Sect. 3, we reconstructed jump-intensity parameters under the alternative setups described in Sect. 2.1 in order to perform a goodness-of-fit test. For this study we used the series of 5-minute trade prices of the S&P500 index, recorded during trading hours. Specifically, we used the prices recorded between 9.30 a.m. and 4 p.m. Days with early closures were discarded. The period considered for the analysis is May 1, 2007 - August 6, 2021.

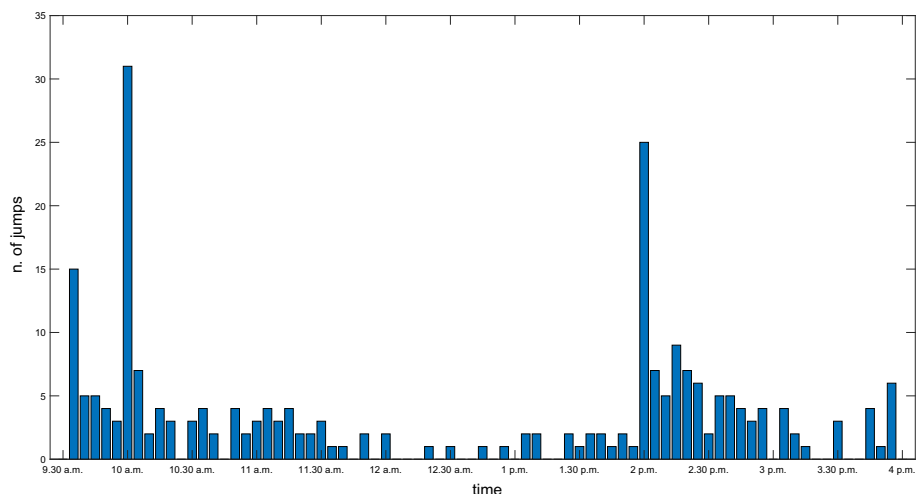
### 5.1 Preliminary jump detection

As a preliminary step of our empirical study, it is of paramount importance to assess whether the frequency of jumps is large enough to justify the use of the model illustrated in (1). Accordingly, by means of the methodology by Lee and Mykland (2008), we tested for the presence of jumps on 5-minute intervals at 1% significance interval. The number of jumps detected per year is illustrated in Fig. 1. For 2007 and 2021, we considered only the periods May-December and January-August, respectively. Based on the observation of Fig. 1, it is evident that jumps play an important role, as each full year considered displays at least 100 jumps. On average, each month analyzed contains approximately 13 jumps. Note that, given the well-documented presence of leverage effects, see, e.g., Ait-Sahalia et al. (2013) and references therein, we are assuming that price jumps are accompanied by volatility jumps (see also the result by Todorov and Tauchen (2011) and Bernis et al., (2021)).

Moreover, some interesting insights can be obtained by looking at the reconstructed intra-day pattern of the jump frequency. As one would expect, the majority of jumps (namely, 1420, corresponding to 85.39% of the total number of jumps detected) takes place between the closing time (16 p.m.) and the opening time (9.30 a.m.) of the next day, as companies typically release their earnings reports when the market is closed. However, if one restricts the horizon to the trading day, as we do in Fig. 2, it is possible to notice that the jump frequency peaks at 10 a.m. and 2 p.m., when new information affecting macroeconomic variables is usually released by, e.g., governments and central banks.



**Fig. 1** S&P500 over the period May 1, 2007–August 6, 2021: frequency of jumps on 5-minute intervals per year



**Fig. 2** S&P500 over the period May 1, 2007– August 6, 2021: daily pattern of the jump frequency

## 5.2 Goodness-of-fit testing of alternative intensity specifications

The goodness-of-fit of the different jump intensity specifications discussed in Sect. 2.1 can be compared via the computation of the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC), see, respectively, Schwarz (1978) and Akaike (1974). To do so, however, one needs first to obtain parameter estimates. In this regard, Table 3 reports the sample average of parameter estimates, along with sample standard deviations, computed over 163 samples of S&P500 prices with length 21 days, spanning the period May 1, 2007 - August 6, 2021. For the estimation, we followed the feasible maximum-likelihood procedure illustrated in Sect. 3. For selection of tuning parameters and initial conditions, necessary

**Table 3** Sample average and standard deviations of parameter estimates, related to four alternative jump intensity setups under model (1), obtained from 163 21-day long samples of 5-minute prices of the S&P500 index spanning the period May 1, 2007–August 6, 2021

Setup	Parameter	Average estimate (std. dev.)
Poisson	$\mu$	156.590 (35.623)
Hawkes	$\mu$	4.500 (9.907)
	$\alpha$	40.359 (9.276)
	$\beta$	2.909 (10.758)
Pure Self-Exciting	$\theta$	15186.733 (13548.571)
Hybrid Self-Exciting	$\mu$	156.589 (35.623)
	$\theta$	0.003 (0.003)

**Table 4** Sample average of AIC and BIC values, related to four alternative jump intensity setups under model (1), obtained from 163 21-day long samples of 5-minute prices of the S&P500 index spanning the period May 1, 2007–August 6, 2021

Setup	AIC	BIC
Poisson	108.470	116.884
Hawkes	181.298	198.180
Pure Self-Exciting	98.636	104.050
Hybrid Self-Exciting	110.470	121.298

for the estimation, we followed the recommendations offered by the simulation exercise in Sect. 4.

Comparing the results obtained under the Poisson setup to those obtained under the Hawkes setup, we observe that the intrinsic intensity  $\mu$  is reduced dramatically to compensate for the self-exciting component of the intensity. However, the estimated speed of mean-reversion  $\beta$  is really small, corresponding to a half-life of about one trimester. We also remark that the estimate of  $\alpha$  is larger than that of  $\beta$ , so that the stability condition is violated. The pure self-exciting framework yields an estimated intensity proportional to the variance of order  $10^4$ . Recalling that the mean value of the VIX is around 10% over the estimation period, we have a reconstructed average intensity which is coherent with the data. Finally, when the hybrid setup is considered, the optimization gives an important weight to the immigration component  $\mu$ , compared to the branching structure  $\theta$ . This is in contrast with a large empirical literature, see Aït-Sahalia et al. (2015), Bates (2000), Bernis and Scotti (2020), Callegaro et al. (2019), Fontana et al. (2021), Horst and Xu, (2019), Jiao et al. (2021), Mancino et al. (2020), Todorov and Tauchen (2011). A possible explanation is that the sample covers only two large crises, i.e., the financial crisis of 2008 and the COVID crisis of 2020. Outside of these two turbulent periods, the volatility is relatively flat; in particular, the VIX index shows an impressive persistence of low values. Low VIX values re-appear also after the peak of the COVID crisis: precisely, three months after the historical peak of 82, reached in March 2020, the VIX reverted to quotes below the value of 40.

Table 4 contains the average values of BIC and AIC over the same 163 21-day samples. Based on the minimization of the AIC and BIC, it appears that the pure-self exciting setup provides the best fit among the intensity specifications considered. Note that this setup minimizes AIC and BIC for all individual samples, not just on average. Both criteria give the same ranking, that is (from best to worst): pure self-exciting, Poisson, hybrid self-exciting and, finally, Hawkes.

Some remarks are in order. First, sample data appear to strongly reject the Hawkes setup, as it provides by far the worst performance in terms of AIC and BIC. Also, it is worth noting that this setup is the one with the largest number of intensity parameters to be estimated (three) and thus is the most likely to be sensitive to numerical instabilities. Overall, the rejection can then be justified both on empirical and theoretical grounds. The rejection of the Hawkes framework contrasts with the result by Callegaro et al. (2019), which is obtained on data from energy markets.

Secondly, notice that AIC and BIC values achieved under the hybrid self-exciting setup are close to the ones obtained under the Poisson setup. This suggests that the addition of a self-exciting effect to the Poisson model via the introduction of the parameter  $\theta$  (whose estimated value is, besides, relatively small, see Table 3) does not importantly increase the goodness-of-fit. Instead, it may be interpreted just as unnecessary over-parametrization, relative to the Poisson model. In fact, the AIC, which attributes less importance to overparametrization compared to the BIC, gives almost equivalent results in relation to the goodness-of-fit of the Poisson and hybrid self-exciting frameworks.

Finally, the pure self-exciting setup minimizes both the AIC and BIC for each month analyzed, outperforming the Poisson setup. It is worth noting that the pure self-exciting setup features the same number of parameters as the Poisson setup (namely, one): thus, the best AIC and BIC values are explained only by larger log-likelihood values. This result is in line with the empirical results obtained on daily data by Bernis et al., (2021), who test both a Poisson (i.e., constant) intensity and an intensity proportional to the square of the VIX, finding that a constant jump intensity is clearly rejected by the data, while an intensity proportional to the VIX achieves an impressive fit.

## 6 Conclusions

In this paper we obtained two main results. First, after introducing a novel stochastic volatility model with price and volatility co-jumps driven by Hawkes processes, we developed a feasible maximum-likelihood procedure to estimate the parameters driving the jump intensity. This procedure exploits the jump test by Lee and Mykland (2008), which allows testing for the presence of jumps on small intervals, and jump-robust volatility estimators based on the bipower variation. Secondly, using S&P500 5-minute prices over the period May 1, 2007 - August 6, 2021, we evaluated the goodness-of-fit of alternative jump-intensity specifications and found that the hypothesis of the intensity being linear in the volatility provides the relatively best fit, thereby hinting towards a self-exciting nature for the jumps. This result supports and extends the empirical findings by Todorov and Tauchen (2011) and Bernis et al., (2021).

It is worth noting that our empirical results on US equity data contrast with the analysis by Callegaro et al. (2019), which was conducted on data from energy markets. This may be explained by the fact that equity markets are highly endogenous while energy markets are more dependent on exogenous events (e.g., power plant failures, lack of wind or solar energy, etc...). Specifically, the results by Callegaro et al. (2019) are consistent with the typically observed clustering of external events and thus support the use of the Hawkes setup to model the intensity of jumps in energy markets. In other words, energy prices seem not to represent an efficient synthesis of the activity of market players, but rather they appear to be highly exposed to external factors. Instead, in our paper we empirically showed that this is not the case for equity markets, where the variance of traded assets represents a proper

synthetic indicator of the level of endogenous activity by market players and thus can be deemed as a very good predictor of the intensity of large fluctuations.

Similar considerations apply also to other markets. For instance, in the case of fixed-income and credit markets, the interest rate and the default intensity may represent a good proxy for the intensity of large fluctuations, as assumed in the setups discussed in Fontana et al. (2021) and Bernis et al. (2018).

**Funding** Open access funding provided by Università degli Studi di Firenze within the CRUI-CARE Agreement.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

### A Proof of Proposition 3

We focus on the general case, as the pure self exciting setup could be easily deduced from it. According to (Duhalde et al., 2014, Theorem 2), we have that zero is an inaccessible boundary point if and only if

$$\int_1^\infty \exp\left(\int_1^z \frac{\Phi(x)}{\Psi(x)} dx\right) \frac{dz}{\Psi(z)} = \infty,$$

where  $\Phi$  and  $\Psi$  are respectively the immigration and branching mechanism of the process  $\lambda$ , i.e., (4). In our particular setup, the immigration rate is  $\Phi(x) := \beta(\mu + \theta\phi)x$ . According to equation (5) in Duhalde et al., (2014), the branching mechanism reads

$$\Psi(x) := \beta x + \frac{1}{2}\sigma_M^2\theta^2 x^2 + \int_0^\infty \left(e^{-x(\alpha+\theta\ell)} - 1\right) \nu_L(d\ell).$$

We first show that the condition is necessary. Let  $\Psi^*(x) := \beta x + \frac{1}{2}\sigma_M^2\theta^2 x^2$ . Since  $x \geq 0$ , we have  $\Psi^*(x) \geq \Psi(x)$  and then

$$\int_1^\infty \exp\left(\int_1^z \frac{\Phi(x)}{\Psi(x)} dx\right) \frac{dz}{\Psi(z)} \geq \int_1^\infty \exp\left(\int_1^z \frac{\Phi(x)}{\Psi^*(x)} dx\right) \frac{dz}{\Psi^*(z)}.$$

A standard computation gives that the right hand term is infinite if and only if  $2\beta(\mu + \theta\phi) \geq \sigma_M^2\theta^2$ . In particular, note that the couple  $(\Psi^*, \Phi)$  corresponds to a CIR process, for which it is well-known that the right-hand term is infinite if and only if  $2\beta(\mu + \theta\phi) \geq \sigma_M^2\theta^2$ .

Conversely, recalling that the jump activity is finite, consider the interval  $[T_i, T_{i+1})$ . The restricted evolution of the intensity (4) reads

$$d\lambda_t = \beta(\mu + \theta\phi) dt - \tilde{\beta}\lambda_t dt + \theta\sigma_M\sqrt{\lambda_t} dW_t^M,$$

that is, we get the CIR evolution with mechanisms  $(\tilde{\Psi}^*, \Phi)$ , where  $\tilde{\Psi}^*(x) = \Psi^* + (\tilde{\beta} - \beta)x$ . Thus  $\lambda$  stays positive between two jumps and it is then always positive, since jump sizes are positive. Moreover, since the jump activity is finite, the jump time sequence  $\{T_i\}_{i \geq 1}$  goes to infinity. Therefore we have that the condition is both sufficient and necessary.  $\square$



## B Simulation scheme

In order to simulate the SDEs in (1), we adapted the schemes proposed in Dassios and Zhao (2013) and Bernis et al., (2021). The only approximation is to replace the increment of the martingale  $M_t$  by a standard Euler-Maruyama method. However, for the purpose of the paper, we have assumed  $M \equiv 0$  in order to have an exact simulation scheme.

In the next algorithm we detail the discretization scheme in its whole generality.

---

### Algorithm 1 Simulation scheme

---

**Input:**  $X_0, V_0, \eta, \lambda_0, \mu, \alpha, \beta, \theta, \gamma_L, \gamma_Z, \kappa, \phi, \{t_0, t_1, \dots, t_n := T\}$

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

---

Set  $k = 0$  and  $\delta_j = t_j - t_{j-1}$

**for**  $j = 1 : n$  **do**

Draw a uniform random variable  $U$

Set  $\bar{X}_{t_{j-1}} = X_{t_{j-1}}, \bar{V}_{t_{j-1}} = V_{t_{j-1}}, \bar{\lambda}_{t_{j-1}} = \lambda_{t_{j-1}}$  and  $\bar{\delta}_j$

Compute  $\hat{V}_j = \bar{\delta}_j \phi + \frac{1}{\kappa} (\bar{V}_{t_{j-1}} - \phi) (e^{-\kappa \bar{\delta}_j} - 1)$  and

$\hat{\Lambda}_j := \int_{t_{j-1}}^{t_j} \tilde{\lambda}_s ds = \bar{\delta}_j \mu + \frac{1}{\kappa} (\bar{\lambda}_{t_{j-1}} - \mu) (e^{-\beta \bar{\delta}_j} - 1) + \theta \hat{V}_j$

**while**  $\hat{\Lambda}_j < -\ln(U)$  **do**

Draw  $(L, Z)$  in accordance with the joint law

Draw  $G \sim \mathcal{N}(0, 1)$

Define  $\Delta T$  such that  $\int_{t_{j-1}}^{t_j} \hat{\lambda}_s ds = -\ln(U)$  as in Dassios and Zhao (2013)

Draw  $\Delta M_j := M_{t_{j-1}+\Delta T} - M_{t_{j-1}} | \bar{V}_{t_{j-1}}, \bar{\lambda}_{t_{j-1}}$

$\bar{V}_{t_{j-1}} \leftarrow \phi + (\bar{V}_{t_{j-1}} - \phi) e^{-\kappa \Delta T} + L$

$\bar{\lambda}_{t_{j-1}} \leftarrow \mu + (\bar{\lambda}_{t_{j-1}} - \mu) e^{-\beta \Delta T} - 1 + \alpha$

Compute  $\tilde{V}_j = \Delta T \phi + \frac{1}{\kappa} (\bar{V}_{t_{j-1}} - \phi) (e^{-\kappa \Delta T} - 1)$

$\bar{X}_{t_{j-1}} \leftarrow \bar{X}_{t_{j-1}} + \eta \Delta T - Z + \sqrt{\tilde{V}_j} G - \frac{1}{2} \tilde{V}_j$

$\bar{\delta}_j \leftarrow \bar{\delta}_j - \Delta T$  and draw  $U$

$\hat{\Lambda}_j \leftarrow \bar{\delta}_j \mu + \frac{1}{\kappa} (\bar{\lambda}_{t_{j-1}} - \mu) (e^{-\beta \bar{\delta}_j} - 1) + \theta \hat{V}_j$

**end**

Draw  $G \sim \mathcal{N}(0, 1)$

Draw  $\Delta M_j := M_{t_{j-1}+\bar{\delta}_j} - M_{t_{j-1}} | \bar{V}_{t_{j-1}}, \bar{\lambda}_{t_{j-1}}$

$V_{t_j} = \phi + (\bar{V}_{t_{j-1}} - \phi) e^{-\kappa \bar{\delta}_j}$

$\lambda_{t_j} = \mu + (\bar{\lambda}_{t_{j-1}} - \mu) e^{-\beta \bar{\delta}_j}$

Compute  $\tilde{V}_j = \bar{\delta}_j \phi + \frac{1}{\kappa} (\bar{V}_{t_{j-1}} - \phi) (e^{-\kappa \bar{\delta}_j} - 1)$

$X_{t_j} = \bar{X}_{t_{j-1}} \eta \bar{\delta}_j + \sqrt{\tilde{V}_j} G - \frac{1}{2} \tilde{V}_j$

**end**

---

## References

- Aït-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.
- Aït-Sahalia, Y., Fan, J., & Li, Y. (2013). The leverage effect puzzle: disentangling the sources of bias at high frequency. *Journal of financial economics*, 109(1), 224–249.
- Aït-Sahalia, Y., & Jacod, J. (2014). *High-frequency financial econometrics*. Princeton University Press.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.
- Bakshi, G., Cao, C., & Chen, Z. (1997). Empirical performance of alternative option pricing models. *Journal of Finance*, 52, 2003–2049.
- Barndorff-Nielsen, O., Gravesen, S. E., Jacod, J., & Shepard, N. (2006). Limit theorems for bipower variation in financial econometrics. *Econometric Theory*, 22(4), 677–719.
- Barndorff-Nielsen, O., & Shephard, N. (2001). *Modelling by Levy processes for financial econometrics* (pp. 283–318). Birkhäuser: In Lévy Processes.
- Barndorff-Nielsen, O., & Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistics Society*, 63(2), 167–241.
- Barndorff-Nielsen, O., & Shepard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of financial econometrics*, 2(1), 1–37.
- Bates, D. (1996). Jump and stochastic volatility: exchange rate processes implicit in Deutsche mark options. *Review of Financial Studies*, 9, 69–107.
- Bates, D. (2000). Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics*, 94, 181–238.
- Bernis, G., Salhi, K., & Scotti, S. (2018). Sensitivity analysis for marked Hawkes processes: application to CLO pricing. *Mathematics and Financial Economics*, 12(4), 541–559.
- Bernis, G., & Scotti, S. (2020). Clustering effects via Hawkes processes. In *From probability to finance*, (pp. 145–181). Springer.
- Bernis, G., Brignone, R., Scotti, S., & Sgarra, C. (2021). A gamma Ornstein-Uhlenbeck model driven by a Hawkes Process. *Mathematics and Financial Economics*, 15, 747–773.
- Bessy-Roland, Y., Boumezoued, A., & Hillairet, C. (2021). Multivariate Hawkes process for cyber insurance. *Annals of Actuarial Science*, 15(1), 14–39.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3), 307–327.
- Brémaud, P., & Massoulié, L. (1996). Stability of nonlinear Hawkes processes. *The Annals of Probability*, 24(3), 563–1588.
- Broadie, M., Chernov, M., & Johannes, M. (2007). Model specification and risk premia: evidence from futures options. *Journal of Finance*, 62(3), 1453–1490.
- Callegaro, G., Gaigi, M., Scotti, S., & Sgarra, C. (2017). Optimal investment in markets with over and under-reaction to information. *Mathematics and Financial Economics*, 11(3), 299–322.
- Callegaro, G., Mazzoran, A., & Sgarra, C. (2019). A self-exciting modelling framework for forward prices in power markets. *Applied stochastic models in business and industry*. <https://doi.org/10.1002/asmb.2645>
- Dassios, A., & Zhao, H. (2013). Exact simulation of Hawkes process with exponentially decaying intensity. *Electronic Communications in Probability*, 18, 1–13.
- Dawson, D. A., & Li, Z. (2006). Skew convolution semigroups and affine Markov processes. *The Annals of Probability*, 34(3), 1003–1142.
- Dawson, D. A., & Li, Z. (2012). Stochastic equations, flows and measure-valued processes. *The Annals of Probability*, 40(2), 813–857.
- Duhalde, X., Foucart, C., & Ma, C. (2014). On the hitting times of continuous-state branching processes with immigration. *Stochastic Processes and their Applications*, 124(12), 4182–4201.
- Duffie, D., Pan, J., & Singleton, K. (2000). Transform analysis and asset pricing for affine jump diffusions. *Econometrica*, 68(6), 1343–1376.
- Eberlein, E., & Kallsen, J. (2020). *Mathematical Finance*. Springer finance.
- El Euch, O., Fukasawa, M., & Rosenbaum, M. (2018). The microstructural foundations of leverage effect and rough volatility. *Finance and Stochastics*, 22(2), 241–280.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, p. 987–1007.
- Engle, R.F., Patton, A.J. (2007). What good is a volatility model?. In *Forecasting volatility in the financial markets*, (47–63). Butterworth-Heinemann.
- Eraker, B. (2004). Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *Journal of Finance*, 59(3), 1367–1404.

- Eraker, B., Johannes, M., & Polson, N. (2003). The impact of jumps in volatility and returns. *Journal of Finance*, 58, 1269–1300.
- Errais, E. (2019). Pricing insurance premia: a top down approach. *Annals of Operations Research*, p. 1–16.
- Fičura, M. (2017). Forecasting jumps in the intraday foreign exchange rate time series with Hawkes processes and logistic regression. In *New Trends in Finance and Accounting*, (pp. 125–137). Springer.
- Fontana, C., Gnoatto, A., & Szulda, G. (2021). Multiple yield curve modelling with CBI processes. *Mathematics and Financial Economics*, 15(3), 579–610.
- Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58, 83–90.
- Hawkes, A. G., & Oakes, D. (1974). A cluster process representation of a self-exciting process. *Journal of applied probability*, 11, 493–503.
- Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2), 327–343.
- Horst, U., & Xu, W. (2019). The microstructure of stochastic volatility models with self-exciting jump dynamics. *ArXiv e-print*, 1911, 12969.
- Hull, J., & White, A. (1987). The pricing of options on assets with stochastic volatility. *Journal of Finance*, 42, 281–300.
- Ikeda, N., & Watanabe, S. (1998). Stochastic differential equations and diffusion processes 2nd edition, North-Holland mathematical library.
- Kirchner, M. (2017). An estimation procedure for the Hawkes process. *Quantitative Finance*, 17(4), 571–595.
- Jiao, Y., Ma, C., & Scotti, S. (2017). Alpha-CIR model with branching processes in sovereign interest rate modeling. *Finance and Stochastics*, 21(3), 789–813.
- Jiao, Y., Ma, C., Scotti, S., Sgarra, C. (2019). A branching process approach to power markets. *Energy Economics*, 79, 144–156.
- Jiao, Y., Ma, C., Scotti, S., & Zhou, C. (2021). The alpha-heston stochastic volatility model. *Mathematical finance*, 31, 943–978.
- Kallsen, J., Muhle-Karbe, J., & Voß, M. (2011). Pricing options on variance in affine stochastic volatility models. *Mathematical Finance*, 21, 627–641.
- Kou, S. G. (2002). A jump-diffusion model for option pricing. *Management Science*, 48, 1086–1101.
- Lee, S. S., & Mykland, P. A. (2008). Jumps in financial markets: a new nonparametric test and jump dynamics. *The review of financial studies*, 21(6), 2535–2563.
- Lemonnier, R., & Vayatis, N. (2014). Nonparametric Markovian learning of triggering kernels for mutually exciting and mutually inhibiting multivariate Hawkes processes. In W. Daelemans & K. Morik (Eds.), *Machine Learning and Knowledge Discovery in Databases* (pp. 161–176). Berlin: Springer.
- Li, Z. (2020). Continuous-state branching processes with immigration. In *from probability to finance* (pp. 1–69). Springer.
- Li, Z., & Ma, C. (2008). Catalytic discrete state branching models and related limit theorems. *Journal of Theoretical Probability*, 21(4), 936–965.
- Mancino, M. E., Scotti, S., & Toscano, G. (2020). Is the variance swap rate affine in the spot variance? Evidence from S&P500 data. *Applied Mathematical Finance*, 27(4), 288–316.
- Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of financial economics*, 3(1–2), 125–144.
- Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical association*, 83(401), 9–27.
- Ozaki, T. (1979). Maximum likelihood estimation of Hawkes self-exciting processes. *Annals of the institute of statistical mathematics*, 31, 145–155.
- Pan, J. (2002). The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of Financial Economics*, 63, 3–50.
- Protter, P. (1992). *Stochastic integration and differential equations*. Springer.
- Rubin, I. (1972). Regular point processes and their detection. *IEEE Transactions on Information Theory*, 18, 547–557.
- Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461–464.
- Todorov, V., & Tauchen, G. (2011). Volatility jumps. *Journal of business and economic statistics*, 29, 356–371.
- Toscano, G., & Recchioni, M. C. (2022). Bias-optimal vol-of-vol estimation: the role of window overlapping. *Decisions in Economics and Finance*, 45(1), 137–185.
- Woerner J.H.C. (2011). Analyzing the fine structure of continuous time stochastic processes. In: *Seminar on stochastic analysis, random fields and applications VI. progress in probability*, vol. 63, Springer.