Testing for Self-Excitation in Jumps

H. Peter Boswijk*
Amsterdam School of Economics
University of Amsterdam
and Tinbergen Institute

Roger J. A. Laeven†
Amsterdam School of Economics
University of Amsterdam, EURANDOM
and CentER

Xiye Yang‡
Department of Economics
Rutgers University

November 3, 2017

This paper extends the notion of self-excitation in jumps to a rich class of continuous time semimartingale models, proposes statistical tests to detect its presence in a discretely observed sample path at high frequency, and derives the tests' asymptotic properties. Our statistical setting is semiparametric: except for necessary parametric assumptions on the jump size measure, the other components of the semimartingale model are left essentially unrestricted. We analyze the finite sample performance of our tests in Monte Carlo simulations.

Keywords: Self-excitation; jumps; semimartingale; spot jump intensity; discrete sampling; high frequency data; financial crisis.

JEL classification: C12, C14.

MSC classification: Primary: 62F12, 62M05; Secondary: 60J75, 91B25.

*University of Amsterdam and Tinbergen Institute, PO Box 15867, Amsterdam, 1001 NJ, The Netherlands. Email: H.P.Boswijk@uva.nl, Phone: +31 (0)20 5254316.

†University of Amsterdam, EURANDOM and CentER, PO Box 15867, Amsterdam, 1001 NJ, The Netherlands. Email: R.J.A.Laeven@uva.nl, Phone: +31 (0)20 5254219.

‡Corresponding author. Rutgers University, 75 Hamilton Street, New Jersey Hall, New Brunswick, NJ 08901, USA. Email: xiyeyang@econ.rutgers.edu, Phone: +1 848 932 8655.

1 Introduction

Over the past few decades, continuous time stochastic processes with jumps have seen increasing interest in finance, economics, statistics, operations research, and other fields. The initial focus of studies that allow for discontinuities in continuous time financial models has been almost exclusively on the development of financial and economic theory and econometrics in the context of option pricing and risk management. To guarantee mathematical tractability, these early studies typically assume a relatively simple structure with respect to the jump components. The most commonly adopted assumption is that jump increments are serially independent; examples satisfying this assumption include jump-diffusions with compound Poisson jumps or, more generally, Lévy processes (see e.g., Merton (1976), Bertoin (1996), Sato (1999), Cont and Tankov (2003) and the references therein for further details).

In the most recent decade, with the increased availability of intra-daily financial data and powerful computational-statistical tools, another branch of the literature has emerged: first on statistically justifying the existence of jumps (see e.g., Lee and Mykland (2008), Aït-Sahalia and Jacod (2009b), and Aït-Sahalia et al. (2012)) and next on uncovering the subtle structures of jumps, including identifying jump activity (Aït-Sahalia and Jacod (2009a, 2011, 2012), Todorov and Tauchen (2010) and Jing et al. (2012b)), and jump tail behavior (Bollerslev and Todorov (2011, 2014)).

Most recently, Aït-Sahalia et al. (2015) (abbreviated as ACL hereafter) have investigated a new dependence structure among the jump components of a continuous time semimartingale model. These authors relax the assumption of independent discontinuous increments, by embedding a multivariate Hawkes process (referred to as a mutually exciting point process in the original papers by Hawkes (1971a,b)) into a continuous time model which remains an Itô semimartingale, with a drift and a diffusion component. Different from the compound Poisson process, the multivariate Hawkes process allows a jump in one component to raise the probability of future jumps occurring both in the same component (*self-excitation*) and in other components (*cross-excitation*). Intuitively, such a model specification helps to reproduce the observed time and space propagation of jumps during a financial crisis. Empirical findings of the ACL paper support the existence of *mutual excitation*, including both self-excitation and cross-excitation, among major stock markets around the world.[1] Hawkes processes, as pure point processes, have also been successfully applied to model excitation in other fields, such as genome analysis, seismology and neurophysiology, market microstructure, and crime (see e.g., Brillinger (1988), Bowsher (2007), Reynaud-Bouret and Schbath (2010), Mohler et al. (2011) and the references therein).

Yet by the parametric nature of the Hawkes process, these findings may be subject to the particular parametric model specification adopted. Moreover, the jump activity of the Hawkes process employed in the ACL paper remains finite, while the aforementioned papers about identifying jump activity point toward the presence of infinite-activity jumps (at least in the

[1] Besides, Christoffersen et al. (2012) propose a discrete time model, in which the jump intensity is a function of itself (with a one period lag) and of the jump size of the previous period's jump. Li and Zhang (2013) adopt a similar discrete time model for the jump intensity process and find that the jump intensity depends on the current relative stock price level, next to the past jump intensity.

stocks that are analyzed). Therefore, it is important to study the phenomenon of mutual excitation in jumps under model specifications that accommodate this feature, which is the objective of the present paper.

First, we extend the notion of self-excitation in jumps to a class of continuous time semimartingale models that stipulates the presence of infinite-activity jumps.[2] The key feature of this extended notion of self-excitation in jumps is that the intensity of jumps positively depends on the jumps themselves (see Section 2.1 for the formal definition). Second, we develop and implement new testing procedures, which are valid for a wide range of underlying data generating processes, to detect the presence of self-excitation in jumps. We adopt the commonly used asymptotic scheme in high frequency data analysis, in which the length of the sampling interval Δ_n shrinks to zero while the time horizon $T < \infty$ stays fixed. Under such a scheme, detecting self-excitation in jumps amounts to testing for common jumps between the process of interest and its latent stochastic jump intensity process, on the discretely observed path over $[0, T]$, with the additional restriction that the jump in the intensity process at the co-jumping time has positive sign. The testing problem is therefore path-wise and not based on an invariant distribution (as in classical time series analysis). In fact, we do not require the invariant distribution of the processes of interest to exist. Thus, the statistical setting in the current paper is fairly general: except for some necessary parametric assumptions on the jump size measure (see Assumption 2 below for details), implying infinite jump activity to ensure statistical consistency

[2]For ease of presentation, we restrict attention to the univariate self-exciting case. The generalization to the multivariate case is conceptually straightforward.

on a fixed time horizon of length $T < \infty$, the coefficients of the drift and diffusion terms and of the jump intensity process are allowed to be components of any Itô semimartingale, as long as some rather mild conditions (such as local boundedness) are satisfied.

In a nutshell, the testing procedures we develop can be summarized as follows. First, we select appropriate functionals of the process of interest and its stochastic jump intensity that yield distinct values on the disjoint outcome sets of self-excitation in jumps and its complement of no self-excitation in jumps. Next, using the truncation technique introduced by Mancini (2001), we identify the intervals on the observed sample paths that contain jumps. Then, we design local jump intensity estimators around each such interval. Thus, we obtain estimators of the chosen functionals. Finally, the tests are based on the distinct limiting behaviors of these estimators on the disjoint outcome sets of self-excitation in jumps and of no self-excitation in jumps, which we describe by deriving the corresponding central limit theorems. Both the null hypothesis of "no self-excitation in jumps" and of "self-excitation in jumps" are considered. The mathematical details of the new limit theories we develop are delicate.

Two relevant papers on testing for common arrivals of jumps are Jacod and Todorov (2009, 2010); the former considers the problem of testing for common jumps in two observable continuous time processes; the latter considers testing for common jumps in an observable continuous time process and its latent volatility process. Our testing problem differs from the testing problems in these papers in at least two important respects. First, Jacod and Todorov (2009, 2010) mainly only consider the question of whether or not there are common jumps, and do in most of their work not discriminate between the *signs* of the co-jumps, as is required for our problem. Therefore, the main null and alternative hypotheses in these two papers are of a different nature than in the current paper.

Second, and more importantly, the estimation methods underlying the tests differ substantially among the three papers. The core difference lies in how the target (log-price, volatility or jump intensity) process is related to what can be observed. In Jacod and Todorov (2009), since both processes of interest are observable, the discretely observed values become natural estimators of the continuous spot (local) ones. By contrast, non-trivial effort is required to deal with the estimation of the latent volatility process in Jacod and Todorov (2010). But the estimation of the latent stochastic jump intensity, as required in this paper, is quite different from, and more demanding than, the estimation of the latent volatility process: the diffusion component contains two parts, a standard Brownian motion, the distribution of which is well-known, and its time-varying loading factor, the volatility process. As previous work has shown, after removing jumps, local averages of squared returns consistently estimate spot volatilities. However, the fact that the jump size measure, which jointly with the stochastic jump intensity process determines the jump measure, is not fully known (by contrast to the law of the standard Brownian motion) complicates the problem we are facing in the present paper. Therefore, additional efforts (and structural assumptions) are needed to estimate the jump activity index, which controls the shape of the jump size measure, and then filter out the stochastic jump intensities. This involves the derivation of the corresponding consistency and functional central limit results. Besides, while volatility has been intensively studied, much fewer results are known about the estimation of stochastic jump intensities. Among the few papers on stochastic jump intensity estimation with high frequency data are Aït-Sahalia and Jacod (2009a) and Jing et al. (2012b), who have investigated the problem of estimating the *integrated* jump intensity, as an interim step in the estimation of the jump activity index. We investigate *local* jump intensity estimation, as an interim step in our testing procedure.

To assess the finite sample performance of our tests, we conduct Monte Carlo simulations. They show that our tests have good power and size properties under both null hypotheses considered, in a realistic setting. Upon applying our tests to 30 real time-series of high frequency individual stock returns data, we find evidence of self-excitation in asset price jumps, particularly during the financial and European debt crises.

The remainder of this paper is organized as follows. In Section 2 we introduce our definition of self-excitation in jumps in the setting of general Itô semimartingales and state our main model assumptions and our testing problem. In Section 3 we introduce and derive the limiting behaviors of our estimators of chosen functionals of the process of interest and its jump intensity. In Section 4 we describe the asymptotic behavior of our tests under two distinct null hypotheses. In Section 5 we present the design and results of our Monte Carlo experiments. Conclusions are in Section 6. All proofs, some examples, additional Monte Carlo experiments, and the description of our empirical analysis are collected in an online appendix, see Boswijk et al. (2017).

2 Definition, Assumptions and Testing Problem

2.1 Self-excitation in jumps: A definition

The notions of self- and mutually exciting point processes were first introduced by Hawkes (1971a,b). Most recently, ACL employed a multivariate Hawkes process to reproduce time and space propagation during a financial crisis. The phenomenon of self-excitation in jumps can alternatively be generated by an affine jump-diffusion model in which the jump intensity is affine in the diffusive volatility and the latter contains jumps that arrive at the same time as the price jumps; see Pan (2002), Andersen et al. (2015) and the references therein.

Although the specifications of the above two models are different, they share the same feature that the occurrence of price jumps, whether directly or through the volatility channel, exerts a positive feedback on the price jump intensity process, hence induces a higher conditional probability of the occurrence of future price jumps. Such (local and clustered) increases of tail risk constitute exactly the phenomenon we are interested in, and we regard them as the key feature of the notion of self-excitation in jumps.[3] While these two models assume the jump activity to be finite, we are going to generalize the concept of self-excitation to a semimartingale setting including infinite activity jumps.

Let X_t be the logarithmic price of a financial asset at time t. Throughout this paper, we assume that the log-price follows an Itô semimartingale de-

[3] As explained by Aït-Sahalia et al. (2015), this feature plays a similar role for jumps as ARCH (Engle (1982)) does for volatility.

fined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$. Its *Grigelionis decomposition* (cf. Jacod and Protter (2011), Theorem 2.1.2) is given by

$$\begin{aligned} X_t = \ & X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s \\ & + h(\delta) * (\mu^X - \nu^X)_t + (\delta - h(\delta)) * \mu_t^X, \end{aligned} \tag{2.1}$$

where b and σ are real-valued optional processes, W is a standard Brownian motion, h is a truncation function, μ^X is a Poisson random measure on $\mathbb{R}_+ \times E$ with compensator ν^X, and δ is a predictable function. Recent empirical evidence and statistical procedures suggest to use pure jump processes to model the price process (see e.g. Todorov (2015), Todorov and Tauchen (2010), Jing et al. (2012a) and Kong et al. (2015)). This would simplify the problem at hand. Yet to achieve more generality, we proceed with the above representation. The testing procedures we describe below also apply to the pure jump case.

In what follows, we impose the following very mild assumption on b and σ, but leave their dynamics unspecified. Note, in particular, that they can be random and exhibit jumps.

Assumption 1. *The processes $b = (b_t)$ and $\sigma = (\sigma_t)$ are locally bounded.*

To factor out the jump intensity process, we make the following structural assumption on ν^X:

Assumption 2. *Assume $\nu^X(\omega, dt, dx) = dt \otimes F_t(\omega, dx)$, with F_t a predictable random measure. Furthermore, suppose there are three (nonrandom) numbers $\beta \in (0, 2)$, $\beta' \in [0, \beta)$ and $\gamma > 0$, and a locally bounded process*

$L_t \geq 1$, *such that, for all* (ω, t), $F_t = F'_t + F''_t$, *where:*

(a) $F'_t(\omega, dx) = f'_t(x)\, \lambda_{t-}(\omega)dx$, $0 < \lambda_t(\omega) \leq L_t$ *and*

$$f'_t(x) = \frac{1 + |x|^\gamma h_t(x)}{|x|^{1+\beta}}, \quad 1 + |x|^\gamma h_t(x) \geq 0, \quad |h_t(x)| \leq L_t. \tag{2.2}$$

(b) F''_t *is a measure that is singular with respect to* F'_t *and satisfies*

$$\int_{\mathbb{R}} (|x|^{\beta'} \wedge 1) F''_t(dx) \leq L_t. \tag{2.3}$$

(c) The process $\lambda = (\lambda_t)$ *is a positive Itô semimartingale given by*

$$\begin{aligned} \lambda_t = \lambda_0 + \int_0^t b'_s ds + \int_0^t \sigma'_s dW_s + \int_0^t \sigma''_s d\tilde{B}_s \\ + \delta' * \mu^X_t + \delta'' * \mu^{\perp X}_t, \end{aligned} \tag{2.4}$$

where $\tilde{B}$ *is a standard Brownian motion independent of* W, *and* $\mu^{\perp X}$ *is orthogonal to* μ^X.[4] *Furthermore,* δ' *and* δ'' *are predictable.*

We now state our definition of self-excitation in jumps (examples can be found in the online appendix):

Definition 1. *An outcome* $\omega \in \Omega$ *exhibits self-excitation in jumps on the interval* $(0, T]$ *if* λ *jumps positively at some (not necessarily all) price jump*

[4]The "orthogonality" between two random measures means here that (almost surely) they share no common Dirac masses lying on the same "vertical" line $\{t\} \times E$. Or, in other words, they have no common jump times.

times, i.e.

$$\sum_{0\le t\le T} \Delta\lambda_t(\omega)\, 1_{\{\Delta\lambda_t(\omega)\ge 0,\, \Delta X_t(\omega)\neq 0\}} > 0. \tag{2.5}$$

Note that we still need the condition $\Delta X_t(\omega) \neq 0$ in the above indicator function to exclude positive jumps of λ that are not associated with jumps of X. A first alternative hypothesis to self-excitation in jumps is:

$$\sum_{0\le t\le T} |\Delta\lambda_t(\omega)|\, 1_{\{\Delta X_t(\omega)\neq 0\}} = 0. \tag{2.6}$$

That is, X and λ don't have common jumps on this particular path (but they may for another path). A second alternative hypothesis is:

$$\sum_{0\le t\le T} \Delta\lambda_t(\omega)\, 1_{\{\Delta\lambda_t(\omega)\le 0,\, \Delta X_t(\omega)\neq 0\}} < 0. \tag{2.7}$$

In this case, jumps tend to reduce their intensities, hence inducing a *self-dampening* property. This alternative shows clearly that the sign of the intensity jumps matters for our definition of self-excitation in jumps.

We conclude this subsection with several remarks.

Remark 1. *Assumption 2 requires the presence of infinite-activity jumps in X. This is to guarantee that the spot values of λ can be consistently estimated from a local window (see* (3.6) *below). This differentiates the current paper from the previous literature on self-exciting processes. Note that in the finite jump activity case, the number of jumps in any finite interval is finite. Hence one does not have enough observations to estimate the spot jump intensity*

λ. Even if one can have an infinite number of jumps by letting the time horizon go to infinity, only stationary moments of λ, not the spot values, can be estimated consistently in the finite jump activity case.

Remark 2. *When the jump activity is finite, as in the case of a Hawkes process, it is quite natural to call λ the stochastic jump intensity: it is stochastic and measures the conditional arrival rate of log-price jumps. In the case of infinite jump activity, λ no longer represents the conditional arrival rate of jumps. Hence, it might be more appropriate to refer to λ as a stochastic scale. However, for ease of presentation and with slight abuse of terminology, we do not distinguish between these two notions, and continue to refer to λ as the (stochastic) jump intensity in the sequel.*

Remark 3. *One may note that our specification of F'_t is similar to but slightly different from that in Aït-Sahalia and Jacod (2009a). That paper requires F'_t to be very close to the Lévy measure of a β-stable process (scaled by λ_t) on a random interval $(-z_t^-, z_t^+)$ with $1/L_t \leq z_t^\pm \leq 1$. We assume in (2.2) that F'_t behaves like a β-stable process on the whole real line. To understand the reason for this difference, note that log-price jumps outside such an interval are of finite activity, hence do not affect the estimation of β asymptotically. Therefore, for the purpose of estimating the degree of jump activity one can leave the jump size distribution unspecified outside that interval to achieve more generality. However, in the current paper, we need the stochastic intensity (or scale) λ to be defined in the same way for any potential jump size, irrespective of whether large or small. Indeed, we require large (finite activity) and small jumps to share the same (infinite* *activity) jump measure and compensator.*

Remark 4. *We allow for additional randomness in λ by introducing $\tilde{B}$ and $\mu^{\perp X}$ in (2.4). The part (c) of Assumption 2 is arguably weak and is satisfied for virtually all parametric jump specifications used in the literature to date (cf. e.g., Todorov (2010) and Bollerslev and Todorov (2011)), including affine jump-diffusion models.*

Remark 5. *We can further decompose the intensity process λ into two parts, corresponding to negative and positive log-price jumps, respectively: $\lambda_{t-} = \lambda_{t-}^{(+)} 1_{\{x>0\}} + \lambda_{t-}^{(-)} 1_{\{x<0\}}$. The extension of our estimation and testing results to $\lambda^{(+)}$ and $\lambda^{(-)}$ is straightforward.*

Remark 6. *The rationale behind Definition 1 is that a log-price jump can (potentially) increase the arrival intensity (or the scale factor) of future jumps. Similarly, one can define excitation from X to another log-price process Y, by requiring that the jump intensity of Y satisfies the part (c) of Assumption 2 (with the same μ^X) and condition (2.5) holds for Y. If the reverse effect is also true, then we have mutual excitation in jumps; see e.g., Dungey et al. (2015) for a specific application.*

Definition 1 allows both negative and positive jumps of X, and whether large or small, to be accompanied by a positive jump of λ. We note that, despite the presence of self-excitation, a standard correlation of X and λ may well be insignificant, because the positive and negative terms in the correlation definition may cancel out.

2.2 The testing problem

Suppose we observe, over a period of total length T (measured in years), every $\Delta_n = T/n$ units of time the log-price $X_{j\Delta_n}$, for $j \in \mathbb{N}$, without measurement error. Let $\Delta_j^n X := X_{j\Delta_n} - X_{(j-1)\Delta_n}$. Our aim is to decide whether or not there is self-excitation in jumps.

It is worth stressing that, given the discretely observed path over $[0, T]$, the best we can achieve in terms of estimation and testing, is to make inference about a specific outcome $\omega \in \Omega$, not about the underlying process dynamics in general. The reason is that, even if we would be in the ideal situation in which the paths of $t \mapsto X_t$ and $t \mapsto \lambda_t$ were fully observed over the time interval $[0, T]$, which would *de facto* mean that we knew whether or not they jumped simultaneously, we still wouldn't know anything about other paths not-observed (see Example A.1 in the online appendix). Consequently, if we cannot reject the null hypothesis of "no self-excitation" on the observed path, it does not mean that the underlying model cannot generate the self-exciting property: it might be a result of "bad luck". In other words, the tests we propose here are constructed on a path-wise basis. Any conclusion will be drawn with respect to the observed path only, not with respect to the underlying model. Thus, we avoid the inconsistency problem discussed in Corradi et al. (2014).

In principle, we allow any log-price jump, whether the small ones and/or the large ones, to trigger the self-excitation effect. But in practice, it may be also interesting to test more specifically whether log-price jumps larger than a certain value can excite their own intensity process. More generally, we can restrict our attention to the jumps of X in a subset $\mathscr{E}$ of $\mathbb{R}$. For example, the sets $\mathscr{E} = (\epsilon, \infty)$ and $\mathscr{E} = (-\infty, -\epsilon)$ correspond to positive and negative log-price jumps of magnitude larger than $\epsilon > 0$, respectively. And their union corresponds to log-price jumps of (absolute) size larger than ϵ. Let $\Omega_T^{\mathrm{j},\mathscr{E}} = \{\omega : \exists\, t \in (0,T],\ \text{s.t. } \Delta X_t(\omega) \in \mathscr{E} \setminus \{0\}\}$ and note that there is no need to consider the possibility of self-excitation in jumps if no price jumps belong to this set (see again Example A.1 in the online appendix). Consider the following disjoint subsets of $\Omega_T^{\mathrm{j},\mathscr{E}}$:

$$\begin{aligned}
\Omega_T^{\mathscr{E},+} &= \Big\{\omega \in \Omega_T^{\mathrm{j},\mathscr{E}} : \sum_{0\le t\le T} \Delta\lambda_t(\omega)\, 1_{\{\Delta\lambda_t(\omega)\ge 0,\, \Delta X_t(\omega)\in\mathscr{E}\setminus\{0\}\}} > 0\Big\},\\
\Omega_T^{\mathscr{E},-} &= \Big\{\omega \in \Omega_T^{\mathrm{j},\mathscr{E}} : \sum_{0\le t\le T} \Delta\lambda_t(\omega)\, 1_{\{\Delta\lambda_t(\omega)\le 0,\Delta X_t(\omega)\in\mathscr{E}\setminus\{0\}\}} < 0\Big\},\\
\Omega_T^{\mathscr{E},0} &= \Big\{\omega \in \Omega_T^{\mathrm{j},\mathscr{E}} : \sum_{0\le t\le T} |\Delta\lambda_t(\omega)|\, 1_{\{\Delta X_t(\omega)\in\mathscr{E}\setminus\{0\}\}} = 0\Big\}.
\end{aligned} \tag{2.8}$$

Then we can test whether the outcome ω belongs to $\Omega_T^{\mathscr{E},\mathrm{self}} := \Omega_T^{\mathscr{E},+}$ or $\Omega_T^{\mathscr{E},\mathrm{no}} := \Omega_T^{\mathscr{E},-} \cup \Omega_T^{\mathscr{E},0}$. When $\mathscr{E} = \mathbb{R}$, we will omit it from the superscript for simplicity.

3 Objective Functionals and Limit Theorems

3.1 Functionals of jumps and intensity

Recall the notation introduced in Section 2.2 and observe that if we would exclude the possibility of self-dampening in jumps, i.e., assign zero probability to Ω_T^- (which we will not in general), then the problem is effectively reduced to testing for co-jumps between X and λ. Therefore, we first briefly discuss the basic idea of testing for common jumps and next modify it to meet our needs.

In general, when testing for co-jumps between two processes X and Y,[5] with $Y \in \mathbb{R}_+^*$ where $\mathbb{R}_+^* = (0, \infty)$, one can consider the following process:

$$U(H)_t = \sum_{0 \leq s \leq t} H(X_{s-}, X_{s+}, Y_{s-}, Y_{s+}) \, 1_{\{\Delta X_s \in \mathscr{E} \setminus \{0\}\}}. \tag{3.1}$$

The idea then consists in choosing a function H on $\mathbb{R} \times \mathbb{R} \times \mathbb{R}_+^* \times \mathbb{R}_+^*$ such that $U(H)_T$ behaves distinctly when X and Y do or do not co-jump within the interval $[0, T]$. For instance, one may choose a function H such that

$$H(x_1, x_2, y_1, y_2) = 0 \Longleftrightarrow x_1 = x_2 \text{ or } y_1 = y_2. \tag{3.2}$$

Consequently, $U(H)_T \neq 0$ on the set where X and Y co-jump within the time interval $[0, T]$ and $U(H)_T = 0$ elsewhere.[6]

Now think of $Y \equiv \lambda$. If we do not make the *ex ante* assumption that self-dampening in jumps is excluded, then, to achieve our goal, we can modify

[5] At this stage, we simply discuss how one can identify common jumps when one has full knowledge of the paths of X and Y on the interval $[0, T]$. We don't deal with the (un)observability of X and Y yet. This will, however, of course be a key issue to be dealt with below.

[6] A complication here is provided by the possibility that X and Y have co-jumps with different signs so that the sum $U(H)_T$ happens to be zero. However, we argue that this event is unlikely to occur. In fact, if the jump sizes are mutually independent and can take infinitely many values, then the probability of this event is exactly zero.

H to satisfy either one of the following conditions

$$H(x_1, x_2, y_1, y_2) \neq 0 \Longleftrightarrow x_1 \neq x_2 \text{ and } y_1 < y_2, \tag{3.3}$$

$$H(x_1, x_2, y_1, y_2) \begin{cases} > 0 & \text{if } x_1 \neq x_2 \text{ and } y_1 < y_2, \\ = 0 & \text{if } x_1 = x_2 \text{ or } y_1 = y_2, \\ < 0 & \text{if } x_1 \neq x_2 \text{ and } y_1 > y_2. \end{cases} \tag{3.4}$$

With the former choice, $U(H)_T$ yields the same result on both $\Omega_T^{\mathscr{E},-}$ and $\Omega_T^{\mathscr{E},0}$. With the latter choice, a full spectrum to distinguish between the three different outcome sets in (2.8) is provided.

In the current paper, the log-price process is discretely observable while the intensity process λ is completely latent. Therefore, we will estimate the local (or spot) jump intensity, yielding $\hat{\lambda}$, and then "approximate" $U(H)$ by the observable process

$$\begin{aligned} U(H, k_n)_t = \sum_{i=k_n+1}^{[t/\Delta_n]-k_n} & H(X_{(i-1)\Delta_n}, X_{i\Delta_n}, \widehat{\lambda}(k_n)_{i-k_n-1}, \widehat{\lambda}(k_n)_i) \\ & \times \; 1_{\{|\Delta_i^n X| > \alpha \Delta_n^{\varpi}\}} \times 1_{\{|\Delta_i^n X| \in \mathscr{E} \setminus \{0\}\}}. \end{aligned} \tag{3.5}$$

Here, our "local estimator" of the spot jump intensity is given by:

$$\widehat{\lambda}(k_n)_i = \frac{\Delta_n^{\varpi \widehat{\beta}}}{k_n \Delta_n} \sum_{j=i+1}^{i+k_n} g\left(\frac{|\Delta_j^n X|}{\alpha \Delta_n^{\varpi}}\right) \frac{\alpha^{\widehat{\beta}}}{C_{\widehat{\beta}}(1)}, \tag{3.6}$$

where the parameter $\widehat{\beta}$ can be obtained by using the methods in Aït-Sahalia and Jacod (2009a) and/or Jing et al. (2012b). The conditions on ϖ and the positive integer k_n are:

$$0 < \varpi < \frac{1}{2}, \quad \frac{1}{K} \leq k_n \Delta_n^{\rho} \leq K \quad \text{with } 0 < \rho < 1,\, 0 < K < \infty. \tag{3.7}$$

Furthermore, the function g satisfies the following assumption:

Assumption 3. *The function g satisfies either one of the following conditions: (a) $g = g_0(x) = 1_{\{x>1\}}$; (b) $g(x) = |x|^p$ if $|x| \leq a$ for some constant $a > 0$ and even integer $p > 2$, and $g(x)$ is even, non-negative, bounded and smooth with bounded and Lipschitz continuous first order derivative.*

Under these conditions and with g of the indicator type, the continuous component of $\Delta_j^n X$ will be truncated out with probability approaching one in the limit as $n \to \infty$. While small jumps may also be truncated out, inducing a truncation error, our results show that this truncation error is negligible in the limit.

Note that if the only jump that occurs is contained in the interval $[(i-1)\Delta_n, i\Delta_n)$, then the estimated values for λ_{i-k_n-1} and λ_i are both zero. Consequently, $U(H, k_n)_T = 0$ for all n, hence we naturally conclude that there is no self-excitation.

Remark 7. *This "local" estimator is obtained by "differentiating" the integrated estimator $\widehat{\Lambda}$ given in Proposition C.1 in the online appendix, which summarizes the relevant results in Aït-Sahalia and Jacod (2009a) and Jing et al. (2012b).*

Remark 8. *Although the idea of constructing the spot jump intensity estimator seems to be natural, when plugging it into $U(H, k_n)_t$, a nontrivial* *challenge arises in deriving the corresponding central limit theory (CLT). Since we allow every log-price jump to excite its intensity process, for each of them, we need other jumps within its left and right local window of length* $k_n\Delta_n$ *to estimate the pre- and post-jump spot intensity levels. However, we need to do the same for the other jumps within these local windows. Thus, those estimators for the spot intensities will overlap with each other. And, more importantly, we cannot use the usually adopted technique (e.g., Jacod and Todorov (2010)) by first focusing on jumps of magnitude larger than* $1/m$ *for some integer* $m \geq 1$*, which are sufficiently separated from each other as long as* n *is large enough, and then letting* m *go to infinity. To make an analogy, the challenge here is somewhat similar to what mixing sequences brings (compared to IID sequences) to classical time series analysis.*

3.2 Main theorems

We now derive the asymptotic behavior of $U(H, k_n)$ defined in (3.5). Intuitively, the asymptotic behavior of the estimator $U(H, k_n)$ of the functional $U(H)$ (such as its convergence rate and limiting variance) is essentially determined by the estimators of the processes X and λ, rather than by the form of the function H, as long as H satisfies some regularity conditions (on its smoothness and boundedness). Recall that the difference between spot volatility estimates and discretely observed log-prices distinguishes Jacod and Todorov (2010) from Jacod and Todorov (2009). By the same token, spot intensity estimates differentiate the current paper from Jacod and Todorov (2009, 2010).

In fact, the estimators of the processes X and λ determine not only the asymptotic behavior, but also the requirements on the function H. In the ideal case in which both X and λ are continuously observable, we can directly tell to which set in (2.8) the outcome belongs, based on the observed (rather than estimated) value of $U(H)$. Therefore, no additional assumptions on H are needed. If at least one of X and λ is only discretely observable, for instance λ, then we need some assumptions on the smoothness and boundedness of the function H in its third and fourth arguments to derive the LLN and CLT. Lastly, if at least one of X and λ is not even discretely observable, as is the case considered in the present paper, then we need stronger assumptions on H, which involve its higher order derivatives, in order to construct valid testing procedures.

To describe under which assumptions on H the estimator $U(H, k_n)$ is consistent, we introduce the family $\mathcal{R}$ of subsets R (cf. Jacod and Todorov (2010)):

$$R \in \mathcal{R} \Longleftrightarrow \begin{cases} R \text{ is open, with a finite complement;} \\ D = \{x : \mathbb{P}(\exists s > 0 \text{ with } \Delta X_s = x) > 0\} \subset R. \end{cases} \tag{3.8}$$

Then, define $\mathcal{D}(R) := \{(x_1, x_2) \in \mathbb{R}^2 : x_2 - x_1 \in R\}$. We want to consider a function form such as $H(x_1, x_2, y_1, y_2) = h_1(x_1, x_2)1_{\{|x_2 - x_1| > \epsilon\}} h_2(y_1, y_2)$, where $\epsilon > 0$ and h_1, h_2 are continuous functions. Obviously, such a function H is not continuous on $\mathbb{R}^2 \times \mathbb{R}_+^{*2}$. So we are going to restrict it to the domain $\mathcal{D}(R) \times \mathbb{R}_+^{*2}$.

Theorem 1. *Suppose that Assumptions 1, 2 and 3 hold. Let ϖ and ρ satisfy condition (3.7) and let $\rho > 1 - \varpi\beta$. Furthermore, let H be a Borel function* *on $\mathbb{R}^2 \times \mathbb{R}_+^{*2}$, which is continuous at each point of $\mathcal{D}(R) \times \mathbb{R}_+^{*2}$ for some $R \in \mathcal{R}$ and satisfies (3.2). Then $U(H, k_n)_t \xrightarrow{\mathbb{P}} U(H)_t$ in the Skorokhod topology, as soon as one of the following conditions are satisfied for some $\epsilon > 0$: (a) $H(x_1, x_2, y_1, y_2) = 0$ for $|x_1 - x_2| \leq \epsilon$; (b) $|H(x_1, x_2, y_1, y_2)| \leq K|x_1 - x_2|^\beta(1 + y_1 + y_2)$, if $|x_1 - x_2| \leq \epsilon$ for some $K > 0$.*

The above consistency result is not enough for testing purposes. In fact, we will need a joint CLT for the pair of processes $(U(H, k_n)_t, U(H, wk_n)_t)$ for some integer $w \geq 2$ when testing against the null hypothesis of self-excitation, because in that case the test statistic will be constructed from the difference between two estimators of $U(H)$. Hence, to maintain generality, we will analyze the joint convergence of the above pair of processes.

To describe the limiting processes we will derive, we need additional notation.[7] We let $(\Omega', \mathcal{F}', \mathbb{P}')$ be an auxiliary space endowed with four sequences, (W_p^-), (W_p^+), $(W_p'^-)$ and $(W_p'^+)$, of independent normal random variables and introduce the following extension $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ of $(\Omega, \mathcal{F}, \mathbb{P})$: $\widetilde{\Omega} = \Omega \times \Omega'$, $\widetilde{\mathcal{F}} = \mathcal{F} \otimes \mathcal{F}'$, and $\widetilde{\mathbb{P}} = \mathbb{P} \otimes \mathbb{P}'$. All variables or processes defined on Ω and Ω' can be trivially extended to $\widetilde{\Omega}$. Next, we denote by $(T_p)_{p \geq 1}$ a sequence that exhausts the jumps of X (Jacod and Shiryaev (2003), I.1.32). This choice may not be unique, but that will not affect the results we are going to derive. Additionally, let $\mathcal{T}^t = \{p \geq 1 : T_p \leq t\}$. Finally, denote by $(\widetilde{\mathcal{F}})$ the smallest filtration that contains $(\mathcal{F}_t)$ and is such that the variables (W_p^-), (W_p^+), $(W_p'^-)$, $(W_p'^+)$ are $\widetilde{\mathcal{F}}_{T_p}$-measurable.

Now, we are ready to introduce two càdlàg adapted processes, $\mathcal{U}_t$ and

[7] We refer to Jacod and Todorov (2010) and I.2.2 in Jacod and Protter (2011) for more details on the space extension that follows.

$\mathcal{U}'_t$, on the extended filtered space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, (\widetilde{\mathcal{F}})_{t\geq 0}, \widetilde{\mathbb{P}})$:

$$\begin{cases} \mathcal{U}_t = & \sum_{p\in\mathcal{T}^t} \big(H'_3(X_{T_p-}, X_{T_p+}, \lambda_{T_p-}, \lambda_{T_p}) W_p^- \\ & \qquad +H'_4(X_{T_p-}, X_{T_p+}, \lambda_{T_p-}, \lambda_{T_p}) W_p^+\big), \\ \mathcal{U}'_t = & \sum_{p\in\mathcal{T}^t} \big(H'_3(X_{T_p-}, X_{T_p+}, \lambda_{T_p-}, \lambda_{T_p}) W_p^{\prime -} \\ & \qquad +H'_4(X_{T_p-}, X_{T_p+}, \lambda_{T_p-}, \lambda_{T_p}) W_p^{\prime +}\big). \end{cases} \tag{3.9}$$

Here, H'_j stands for the first partial derivative w.r.t. its j-th argument. Conditionally on $\mathcal{F}$, these two processes are independent, centered Gaussian martingales (cf. (3.8) in Jacod and Todorov (2010)). In addition, the $\mathcal{F}$-conditional variances of $Z_p^- = W_p^-$ or $W_p^{\prime -}$ and $Z_p^+ = W_p^+$ or $W_p^{\prime +}$ are

$$\widetilde{\mathbb{E}}\left((Z_p^-)^2 \mid \mathcal{F}\right) = \lambda_{T_p-} \frac{\alpha^\beta C_\beta(2)}{\big(C_\beta(1)\big)^2} \quad \text{and} \quad \widetilde{\mathbb{E}}\left((Z_p^+)^2 \mid \mathcal{F}\right) = \lambda_{T_p} \frac{\alpha^\beta C_\beta(2)}{\big(C_\beta(1)\big)^2},$$

where the expectations with tildes are defined on the extended probability space. See Proposition C.1 in the online appendix for the definition of $C_\beta(k)$.

Moreover, we need stronger assumptions on the function H:

Assumption 4. *H is C^1 on $\mathcal{D}(R) \times \mathbb{R}_+^{*2}$ and, for some $q, q'_c, q'_d > \beta$,*

(a) One of (3.2), (3.3) or (3.4) holds;

(b) $\frac{1}{|x_2-x_1|^q} H(x_1, x_2, y_1, y_2)$, $\frac{1}{|x_2-x_1|^q} H'_3(x_1, x_2, y_1, y_2)$ and $\frac{1}{|x_2-x_1|^q} H'_4(x_1, x_2, y_1, y_2)$ are locally bounded on $\mathcal{D}(R) \times \mathbb{R}_+^{*2}$;

(c) $\frac{1}{|x|^{q'_c}|x_2-x_1|^{q'_d}} |H(x_1 + x, x_2, y_1, y_2) - H(x_1, x_2, y_1, y_2)|$ and $\frac{1}{|x|^{q'_c}|x_2-x_1|^{q'_d}} |H(x_1, x_2 + x, y_1, y_2) - H(x_1, x_2, y_1, y_2)|$ are locally bounded

for any $x \in \mathbb{R}$.

Theorem 2. *Let* $\varpi \in (0, 1/2)$. *Suppose that Assumptions 1, 2, 3 and 4 hold with* $p > \frac{1-\varpi\beta}{1/2-\varpi}$, $q > 2$ *and* $q'_c, q'_d \geq 1$, *and condition (3.7) is satisfied with*

$$1 - \varpi\beta < \rho < (1 - \varpi\beta) + 2\phi' \wedge 2\phi'' \wedge \frac{1}{2}\varpi\beta, \tag{3.10}$$

where ϕ' *and* ϕ'' *are defined in the online appendix (Lemmas C.2 and C.3). If, in addition, one of the following conditions is satisfied:*

(a) $H(x_1, x_2, y_1, y_2) = 0$ *for* $|x_1 - x_2| \leq \epsilon$, *where* $\epsilon > 0$;

(b) $2\varpi\beta < 1$ *and* $\rho \leq 1 - \varpi\beta + (q'_c - 1 + \varpi(2q'_d - \beta)) \wedge 2\varpi(q \wedge (q'_c + q'_d) - \beta)$;

then, for any fixed $t > 0$, *the vector*

$$\sqrt{\frac{k_n \Delta_n}{\Delta_n^{\varpi\widehat{\beta}}}} \Big(U(H, k_n)_t - U(H)_t, U(H, wk_n)_t - U(H)_t \Big) \tag{3.11}$$

converges stably in law to the random vector $(\mathcal{U}_t, \frac{1}{w}(\mathcal{U}_t + \sqrt{w-1}\,\mathcal{U}'_t))$.

Remark 9. *The choice of* ρ *and* ϖ *controls the convergence rate, which is* $(\rho + \varpi\beta - 1)/2$ *in the CLT above. Aït-Sahalia and Jacod (2009a) and Jing et al. (2012b), who deal with integrated rather than local jump intensity estimation, find a convergence rate equal to* $\varpi\beta/2$, *which is larger than the rate found in the CLT above. The intuition behind this is essentially the same as for the case of integrated versus local volatility estimation. Since we are using a shrinking time span* $k_n\Delta_n$, *instead of a fixed nonzero one as in Aït-Sahalia and Jacod (2009a) and Jing et al. (2012b), the number of* *observations used in each local window is much smaller, resulting in a slower convergence rate. However, we note that the fewer local observations can be compensated in part by a weaker restriction on ϖ. In Aït-Sahalia and Jacod (2009a), the restriction on ϖ is $\varpi < 1/(2+\beta) \wedge 2/(5\beta)$. So the "universal" value of ϖ that is admissible for all $\beta \in (0,2)$ is $1/5$. In Jing et al. (2012b), the condition is weakened to $\varpi < 1/(2+\beta)$. Hence, a conservative choice is $\varpi < 1/4$. Actually, their results correspond to the case $\rho = 1$. And since in our case $\rho < 1$, the restriction on ϖ can be even weaker. This means that we can use a smaller threshold (but still significantly larger than most Brownian increments) and utilize more information from small increments.*

It may happen that the limiting processes $\mathcal{U}$ and $\mathcal{U}'$ in the above CLT vanish. A simple situation is when X and λ have no common jumps and additionally the following *degeneracy condition* holds:

$$y_1 = y_2 \Longrightarrow |H_3'(x_1, x_2, y_1, y_2)| + |H_4'(x_1, x_2, y_1, y_2)| = 0. \tag{3.12}$$

In such a case, we need a higher order CLT, which requires stronger smoothness assumptions on H.

Assumption 5. *The function H satisfies Assumption 4 and*

(a) $H(x_1, x_2, y_1, y_2)$ is C^1 in (x_1, x_2) and C^2 in (y_1, y_2) on $\mathcal{D}(R) \times \mathbb{R}_+^{*2}$;

(b) $\frac{1}{|x_2 - x_1|^{q''}} H_{ij}''(x_1, x_2, y_1, y_2)$ is locally bounded on $\mathcal{D}(R) \times \mathbb{R}_+^{*2}$ for $i, j = 3, 4$ and $q'' > \beta$.

Theorem 3. *Suppose that the same assumptions as in Theorem 2 hold, but with Assumption 4 replaced by 5 with $q'' > 2$, and assume furthermore that* *the degeneracy condition (3.12) holds. If, in addition, one of the following conditions is satisfied:*

(a) $H(x_1, x_2, y_1, y_2) = 0$ *for* $|x_1 - x_2| \leq \epsilon$*, where* $\epsilon > 0$*;*

(b) $2\varpi\beta < 1$ *and* $\rho \leq 1 - \varpi\beta + (q'_c - 1 + \varpi(2q'_d - \beta))/2 \wedge \varpi(q \wedge (q'_c + q'_d) - \beta)$*;*

then, for any fixed $t > 0$*, the vector*

$$\frac{k_n \Delta_n}{\Delta_n^{\varpi\widehat{\beta}}} \Big(U(H, k_n)_t - U(H)_t, U(H, wk_n)_t - U(H)_t \Big) \tag{3.13}$$

converges stably in law to the random vector $(\overline{\mathcal{U}}_t, \overline{\mathcal{U}}'_t)$*, where*

$$\begin{cases} \overline{\mathcal{U}}_t = & \displaystyle\sum_{p \geq 1} \frac{1}{2} \big(H''_{33}(X_{T_p-}, X_{T_p+}, \lambda_{T_p-}, \lambda_{T_p})(W_p^-)^2 \\ & \quad + H''_{44}(X_{T_p-}, X_{T_p+}, \lambda_{T_p-}, \lambda_{T_p})(W_p^+)^2 \\ & \quad + 2H''_{34}(X_{T_p-}, X_{T_p+}, \lambda_{T_p-}, \lambda_{T_p}) W_p^- W_p^+ \big) 1_{\{T_p \leq t\}}, \\ \overline{\mathcal{U}}'_t = & \displaystyle\sum_{p \geq 1} \frac{1}{2w^2} \big(H''_{33}(X_{T_p-}, X_{T_p+}, \lambda_{T_p-}, \lambda_{T_p}) \\ & \qquad \times (W_p^- + \sqrt{w-1} W_p^{'-})^2 \\ & \quad + H''_{44}(X_{T_p-}, X_{T_p+}, \lambda_{T_p-}, \lambda_{T_p})(W_p^+ + \sqrt{w-1} W_p^{'+})^2 \\ & \quad + 2H''_{34}(X_{T_p-}, X_{T_p+}, \lambda_{T_p-}, \lambda_{T_p})(W_p^- + \sqrt{w-1} W_p^{'-}) \\ & \qquad \times (W_p^+ + \sqrt{w-1} W_p^{'+}) \big) 1_{\{T_p \leq t\}}. \end{cases} \tag{3.14}$$

One may refer to Theorems 2 and 3 as first and second order CLT, respectively, according to the order of the derivatives of H involved in the limiting processes.

The degeneracy condition (3.12) does not exclude the possibility that

$$y_1 > y_2 \Longrightarrow |H_3'(x_1, x_2, y_1, y_2)| + |H_4'(x_1, x_2, y_1, y_2)| = 0.$$

Indeed, if the above property holds and H additionally satisfies Assumption 5 with (3.3), then it is easy to show that (3.12) is also true. Importantly, again the two limiting processes, $\overline{\mathcal{U}}_t$ and $\overline{\mathcal{U}}_t'$, will vanish on the set of no self-excitation in this case. In fact, as long as (3.3) holds, even if we replace $H \in C^2$ in (y_1, y_2) in Assumption 5 by $H \in C^n$ in (y_1, y_2) for any integer $n > 2$, all the derivatives of H (if they exist) w.r.t. y_1 and y_2, not only the first order ones, are still identically zero on the area $\{y_1 \geq y_2\}$. This means that higher order limiting processes akin to $\overline{\mathcal{U}}_t$ and $\overline{\mathcal{U}}_t'$ will vanish on the set of no self-excitation, too.

Remark 10 (About microstructure noise)**.** *Many methods (e.g. subsampling, kernel smoothing and pre-averaging) have been proved to be very successful in eliminating the effects of microstructure noise when estimating volatility related quantities. However, according to Bücher and Vetter (2013), this may not be the case when estimating jump compensator related quantities. For example, they showed why the pre-averaging method doesn't yield a consistent estimator of the tail of a Lévy measure ν, i.e. $\nu([x, +\infty))$ for $x > 0$. Since the jump intensity is part of the measure ν, the existing methods may not work here as well. The (highly liquid) dataset we use in our empirical analysis (see Section B in the online appendix) doesn't exhibit any evidence of microstructure noise at the 10-min sampling frequency.*

4 Tests

For simplicity, we omit the superscript $\mathscr{E}$ in this section. The testing procedures described below apply to any non-trivial choice of $\mathscr{E}$, including $\mathscr{E} = \mathbb{R}$.

4.1 Testing the null hypothesis of "no self-excitation"

The null hypothesis of no self-excitation is given by $\omega \in \Omega_T^{\text{no}}$. Depending on whether or not we assume $\mathbb{P}(\Omega_T^-) = 0$ (e.g. $\delta' \geq 0$ in (2.4)), we will choose a function H that satisfies (3.2) or (3.4), respectively. In the sequel, when employing any theorem of Section 3, we will always assume that the associated function H satisfies the relevant assumptions stated in the theorem.

First of all, by Theorem 1, we have the following convergence result:

$$U(H,k_n)_T \xrightarrow{\mathbb{P}} U(H)_T \begin{cases} = 0, & \text{on the set } \Omega_T^0, \\ \neq 0, & \text{on the set } \Omega_T^+. \end{cases} \quad \text{or} \quad \begin{cases} \leq 0, & \text{on the set } \Omega_T^{\text{no}}, \\ > 0, & \text{on the set } \Omega_T^{\text{self}}. \end{cases}$$

If the degeneracy condition (3.12) does not hold, for instance when we choose H as in (A.2), which satisfies (3.4), we can make use of Theorem 2 to construct the critical region. In this case, note that the limiting process $\mathcal{U}_t$ of $U(H,k_n)_t$ is a Gaussian martingale conditional on $\mathcal{F}$ with conditional variance $U(G)_t$ where the function G is given by

$$\begin{aligned} G(x_1,x_2,y_1,y_2) = \frac{\alpha^\beta C_\beta(2)}{\big(C_\beta(1)\big)^2}\big(&y_1 H_3'(x_1,x_2,y_1,y_2)^2 \\ &+ y_2 H_4'(x_1,x_2,y_1,y_2)^2\big), \end{aligned} \tag{4.1}$$

and satisfies the requirements imposed on the function H in Theorem 1.

Hence, $U(G,k_n)_t$ is a consistent estimator of the limiting conditional variance. Therefore, upon noticing that

$$t_n := \sqrt{\frac{k_n\Delta_n}{\Delta_n^{\varpi\widehat{\beta}}}}\frac{U(H,k_n)_T}{\sqrt{U(G,k_n)_T}} \begin{cases} \xrightarrow{\mathbb{P}} -\infty & \omega \in \Omega_T^-, \\ \xrightarrow{\mathscr{L}_{st}} \mathcal{N}(0,1) & \omega \in \Omega_T^0, \\ \xrightarrow{\mathbb{P}} +\infty & \omega \in \Omega_T^+, \end{cases} \tag{4.2}$$

(with $\mathscr{L}_{st}$ denoting stable convergence in law and $\mathcal{N}(0,1)$ denoting a standard normal random variable) we can construct the following critical region:

$$C_n = \left\{ \sqrt{\frac{k_n\Delta_n}{\Delta_n^{\varpi\widehat{\beta}}}}\frac{U(H,k_n)_T}{\sqrt{U(G,k_n)_T}} > z_{\tilde{a}} \right\}, \tag{4.3}$$

where $z_{\tilde{a}}$ denotes the symmetric $\tilde{a}$-quantile of a standard normal random variable W, that is, $\mathbb{P}(W > z_{\tilde{a}}) = \tilde{a}$.

Corollary 4. *Suppose that the same assumptions as in Theorem 2 hold. Then, with a feasible choice of the function H, the critical region (4.3) has asymptotic level $\tilde{a}$ for testing the null hypothesis Ω_T^{no} (or Ω_T^0), and asymptotic power 1 for the alternative Ω_T^{self}.*

If $\mathbb{P}(\Omega_T^-) = 0$, we can choose functions with property (3.2), such as the function in (A.1), and rely on Theorem 3 to construct a meaningful critical region under the null $\Omega^{(0)}$. The limiting variances of $\overline{\mathcal{U}}$ and $\overline{\mathcal{U}}'$ (in (3.14)) are tedious to compute analytically. A more straightforward way is to employ the following Monte Carlo procedure: take a sequence $M_n \to \infty$ and simulate independent standard normal random variables $W_i^-(m)$ and

$W_i^+(m)$, for $m = 1, \cdots, M_n$. Then set

$$\overline{\mathcal{U}}(n,m) = \frac{\alpha^{\widehat{\beta}} C_{\widehat{\beta}}(2)}{2\big(C_{\widehat{\beta}}(1)\big)^2} \sum_{i=k_n+1}^{[T/\Delta_n]-k_n} \big(H''_{33}(X_{i-1}, X_i, \widehat{\lambda}_{i-k_n-1}, \widehat{\lambda}_i)\widehat{\lambda}_{i-k_n-1}(W_i^-(m))^2$$
$$+ H''_{44}(X_{i-1}, X_i, \widehat{\lambda}_{i-k_n-1}, \lambda_i)\widehat{\lambda}_i(W_i^+(m))^2$$
$$+ 2H''_{34}(X_{i-1}, X_i, \widehat{\lambda}_{i-k_n-1}, \lambda_i)\sqrt{\widehat{\lambda}_{i-k_n-1}\widehat{\lambda}_i} W_i^-(m) W_i^+(m)\big).$$

Next, arrange the sequence of these simulated variables in descending order, $\overline{\mathcal{U}}(n)_{(1)} \geq \overline{\mathcal{U}}(n)_{(2)} \geq \cdots \geq \overline{\mathcal{U}}(n)_{(N_n)}$, and the critical region is given by

$$C_n = \left\{ \frac{k_n \Delta_n}{\Delta_n^{\varpi\widehat{\beta}}} |U(H, k_n)_T| > \overline{\mathcal{U}}(n)_{([\tilde{a}N_n])} \right\}. \tag{4.4}$$

Corollary 5. *Suppose that the set Ω_T^- is of zero probability and the same assumptions as in Theorem 3 hold with the function H satisfying (3.2). Then the critical region (4.4) has asymptotic level $\tilde{a}$ for testing the null hypothesis Ω_T^0, and asymptotic power 1 for the alternative Ω_T^{self}.*

4.2 Testing the null hypothesis of "self-excitation"

In this section, the null hypothesis becomes self-excitation, or $\omega \in \Omega_T^{\text{self}}$. Again, if we do not assume that Ω_T^- is of zero probability, we can choose the function H given in (A.2) or the like. Then, using the same notation as in the previous subsection, we get a critical region as follows:

$$C_n = \left\{ \sqrt{\frac{k_n \Delta_n}{\Delta_n^{\varpi\widehat{\beta}}}} \frac{U(H, k_n)_T}{\sqrt{U(G, k_n)_T}} \leq -z_{\tilde{a}} \right\}. \tag{4.5}$$

One may note this critical region is trivially the "opposite" of (4.3). Yet, if we do not assume $\mathbb{P}(\Omega_T^-) = 0$, this is probably the best that can be achieved.

If we impose the assumption that $\mathbb{P}(\Omega_T^-) = 0$, then there is only one possible alternative left, $\omega \in \Omega_T^0$. Hence, we have more appropriate choices for the function H, as long as (3.2) is satisfied, and the limiting processes in Theorem 3 do not vanish on $\omega \in \Omega_T^0$.

However, an important difficulty still remains. Unlike in the previous situation in which we knew $U(H)(\omega)_T = 0$ under the null $\omega \in \Omega_T^0$, now we have no knowledge about it, other than that it should be positive on Ω_T^+. Therefore, under the null hypothesis $\omega \in \Omega_T^{\text{self}}$, $U(H, k_n)_T$ can not be regarded as the finite sample estimation error any more. Thus, it is unclear how to provide an "absolute" error based critical region other than (4.5).

Therefore, we turn to the *relative* estimation error given by

$$R_n = \frac{U(H, wk_n)_T - U(H, k_n)_T}{U(H, k_n)_T}, \tag{4.6}$$

for an integer $w \geq 2$. Upon combining Theorems 1 and 3, we obtain the following asymptotic behavior of the ratio R_n:

$$\begin{cases} R_n \xrightarrow{\mathbb{P}} 0, & \text{on the set } \Omega_T^{\text{self}}, \\ R_n \xrightarrow{\mathscr{L}_{st}} \dfrac{\overline{\mathcal{U}}'_T}{\overline{\mathcal{U}}_T} - 1 \neq 0, & \text{on the set } \Omega_T^0. \end{cases}$$

In words, the ratio R_n has a degenerate distribution located at the point zero on the set Ω_T^{self}, while it has $\mathcal{F}$-conditionally a density on Ω_T^0, hence is different from zero with probability one. Such a distinction enables us to construct a test based on how much the relative estimation error deviates from the point zero.

First note that

$$R_n = \frac{(U(H, wk_n)_T - U(H)_T) - (U(H, k_n)_T - U(H)_T)}{U(H, k_n)_T},$$

where the numerator on the right-hand side converges stably in law to a Gaussian random variable, and the denominator converges in probability to $U(H)_T$, in restriction to the set Ω_T^{self}. On the set Ω_T^{self}, we do not have to worry about the degeneracy problem. Hence, we can make use of Theorem 2 to compute the asymptotic variance. Then, by the continuous mapping theorem, we obtain

$$\sqrt{\frac{k_n \Delta_n}{\Delta_n^{\varpi \widehat{\beta}}}}\, R_n \xrightarrow{\mathscr{L}_{st}} \frac{\sqrt{w-1}\, \mathcal{U}'_T - (w-1)\, \mathcal{U}_T}{w\, U(H)_T},$$

and the critical region is

$$C_n = \{|R_n| > z_{\tilde{a}} \sqrt{V_n}\}, \tag{4.7}$$

with $V_n = \frac{\Delta_n^{\varpi \widehat{\beta}}}{k_n \Delta_n} \frac{(w-1) U(G, k_n)_T}{w (U(H, k_n)_T)^2}$, where the function G is the same as in (4.1).

One important remark should be made about the asymptotic power of this critical region. On the set Ω_T^0, the variable $|R_n|/\sqrt{V_n}$ converges stably in law to a limit $\mathcal{V}$, which is *a.s.* non-vanishing. In other words, $|R_n|/\sqrt{V_n}$ does not diverge to infinity under the alternative hypothesis. The asymptotic power of our test is

$$\tilde{b} = \inf\left(\mathbb{P}\big(|R_n| > z_{\tilde{a}}\sqrt{V_n} \mid A\big) : A \in \mathcal{F}, A \subset \Omega_T^{\text{no}}, \mathbb{P}(A) > 0\right),$$

and there is no guarantee that this quantity is close to 1. To solve this problem, we employ the same idea as in Jacod and Todorov (2010) to truncate the estimated variance V_n: choose a sequence of positive numbers v_n such that

$$v_n \to 0, \quad \text{and} \quad \frac{k_n v_n \Delta_n}{\Delta_n^{\varpi\widehat{\beta}}} \to \infty,$$

and define $V_n' = V_n \wedge v_n$. Now the critical region becomes

$$C_n' = \{|R_n| > z_{\tilde{a}}\sqrt{V_n'}\}. \tag{4.8}$$

The rationale is as follows. Observe that

$$\frac{k_n \Delta_n}{\Delta_n^{\varpi\widehat{\beta}}} V_n' = \frac{(w-1)U(G,k_n)_T}{w(U(H,k_n)_T)^2} \wedge \frac{k_n v_n \Delta_n}{\Delta_n^{\varpi\widehat{\beta}}}.$$

Then, restricted to the set Ω_T^+, the first term on the right-hand side, equivalent to $k_n \Delta_n V_n / \Delta_n^{\varpi\widehat{\beta}}$, converges to a positive finite limit. Consequently, we have $\mathbb{P}(V_n = V_n') \to 1$. In other words, this truncation does not have an asymptotic effect on the size of this test. But it will increase its power against the alternative.

Corollary 6. *Suppose that the set Ω_T^- is of zero probability and the same assumptions as in Theorem 3 hold with the function H satisfying (3.2).*

Then the critical regions (4.7) and (4.8) have asymptotic level $\tilde{a}$ for testing the null hypothesis Ω_T^{self}. In addition, (4.8) has asymptotic power 1 for the alternative Ω_T^0.

Remark 11. *Based on the analysis above, one could also perform a two stage test, without assuming $\mathbb{P}(\Omega_T^-) = 0$. First, one employs a function H satisfying (3.2) to test $\omega \in \Omega_T^+ \cup \Omega_T^-$ against $\omega \in \Omega_T^0$, since the ratio R_n converges in probability to 1 on the set Ω_T^-, too. Next, if one does not reject the null, one uses the critical region (4.5) (with an appropriate choice of H) to test $\omega \in \Omega_T^+$ against $\omega \in \Omega_T^-$.*

5 Monte Carlo Study

Throughout this section, we adopt a sampling scheme with a total observation length of one year (i.e., $T = 1$) and a sampling frequency of 10 minutes so that $\Delta_n = 1/(252*39)$, assuming 252 trading days per annum and 6.5 trading hours per day. For comparison, we also report the results for one week ($T = 5/252$) and a sampling frequency of 5 seconds ($\Delta_n = (5/252)/23{,}400$). Furthermore, we take $\varpi = 1/3$, $\rho = 0.6$ and $k_n = [\Delta_n^{-\rho}]$ (cf. conditions (3.7) and (3.10)). Finally, we take $\beta = 1.25$, in view of the fact that the estimated values of β in Aït-Sahalia and Jacod (2009a) and Jing et al. (2012b) are all larger than 1.[8] For all the experiments in this section, we conduct 5,000 simulations.

To assess the finite sample testing power and size, we consider two models, one with and one without self-excitation in jumps. The basic setting is

[8]The results for $\beta = 1.5$ or $\beta = 1.75$ are similar.

given by the following dynamics (additional Monte Carlo results are given in the online appendix):

$$\begin{cases} dX_t &= \sigma_t dW_t + \lambda_{t-} dY_t, \\ d\sigma_t^2 &= \kappa(\theta - \sigma_t^2)dt + \eta\sigma_t d\tilde{B}_t, \\ d\lambda_t &= \kappa_\lambda(\lambda_\infty - \lambda_t)dt + \eta_\lambda d\tilde{B}'_t + \xi\, 1_{\{|\Delta X_t| > \epsilon\}}. \end{cases} \tag{5.1}$$

Here, W and $\tilde{B}$ are two standard Brownian motions with $\mathbb{E}(dW_t d\tilde{B}_t) = \varphi dt$, $\tilde{B}'$ is a standard Brownian motion independent of $\tilde{B}$ and W, and Y is a β-stable process. As in Aït-Sahalia and Jacod (2009a) and Jing et al. (2012b), we take $\kappa = 5$, $\theta = 1/16$, $\eta = 0.5$ and $\varphi = -0.5$. The novelty in this setting, and the key interest in this paper, appears in the third equation, in which we allow increments of X with absolute value larger than some threshold ϵ to excite the jump intensity, as long as $\xi > 0$. As a benchmark parameter specification, we set $\kappa_\lambda = 1{,}400$, $\eta_\lambda = 200$ and $\epsilon = 120\sqrt{\theta}\Delta_n^{\varpi}$. The mean reversion level λ_∞ is calibrated to deliver a pre-specified value of the tail probability (specifically, we use 0.25%; see the online appendix for details).

We study test statistics based on the following form of the function H (adopting the notation of Assumption 3 and Example A.2):

$$H(p,q) = \begin{cases} g_0(x_2 - x_1) \cdot (y_2 - y_1)^q \cdot 1_{\{|x_2 - x_1| \geq \epsilon\}}, & p = 0; \\ g_{2,p}(x_2 - x_1) \cdot (y_2 - y_1)^q \cdot 1_{\{|x_2 - x_1| \geq \epsilon\}}, & p > 2; \end{cases}$$

where $\epsilon > 0$. To distinguish between the three sets Ω_T^-, Ω_T^0 and Ω_T^+, one can consider, for example, $H(6,1)$ and $H(0,1)$, which do not satisfy the degeneracy condition (3.12). With such choices, Theorem 2 yields the result displayed in (4.2). Recall that we constructed the critical region (4.3) based on the asymptotic normality of t_n under the null $\omega \in \Omega^{(0)}$. Figure 1 illustrates kernel density estimates of t_n on the set $\Omega^{(0)}$, for different sampling frequencies, and different choices of the function H, and of the function g used for spot jump intensity estimation. Throughout this section, we use Jing et al. (2012b) to estimate β. The results indicate that the behavior of t_n becomes closer to that of a standard normal random variable when we increase the sampling frequency and, in this case, when we use $H = H(0,1)$ and $g = g_0$.

Figure 2 shows the size and power of the test for the null hypothesis of no self-excitation based on t_n. When $\xi = 0$, in which case all outcomes necessarily belong to $\Omega^{(0)}$, the percentages of rejection (which happens when $t_n > z_{\tilde{a}}$) are all close to their corresponding nominal level, indicating a well-behaved size. The power is also quite good. Note that the testing power depends on the speed of $\sqrt{k_n}\Delta_n^{(1-\varpi\beta)/2}$ diverging to infinity, which is moderate, a fact reflecting the inherent difficulties in recovering the latent jump intensity. If we increase (decrease) ξ or decrease (increase) κ_λ, implying a larger (smaller) or longer (shorter) effect, respectively, the testing power will be stronger (weaker), as we will see below. Note also that a potentially somewhat limited testing power of the no self-excitation test, especially applicable in the 10-min sampling frequency case, means that it is unlikely to yield spurious detection of self-excitation. Therefore, the empirical evidence of self-excitation that we find in our empirical analysis is only a conservative estimate, making it even more convincing.

Next, we proceed to the test for the null hypothesis of self-excitation based on the relative estimation error R_n presented in Section 4.2. This test relies on the behavior of $U(H, k_n)_t$ and $U(H, wk_n)_t$, where we take $w = 2$. In finite samples, the local observations within ranges k_n and wk_n can be quite different. At the same time, the spot jump intensity estimation is sensitive to the realized jump sizes. Nevertheless, we found that e.g., the choice of $H = H(0, 2)$ yields desirable results, illustrated in Figure 3. Although when the null hypothesis is true, the rejection rates differ somewhat from their corresponding nominal levels at low levels, these deviations remain relatively small and acceptable. Furthermore, the power is simply excellent.

Finally, we present simulation results for the various alternative sets of parameter values displayed in Table 1. The sets are obtained by letting only one of the parameters deviate from the benchmark parameter specification. (We recall that the benchmark parameter specification is given by κ_λ = 1,400, $\eta_\lambda = 200$, $\xi = 50$, and $\epsilon_\lambda = 120$, with $\epsilon = \epsilon_\lambda \sqrt{\theta} \Delta_n^{\varpi}$, which corresponds to parameter set 1 in the table.) The results are in Table A.2 in the online appendix.

Table 1: Alternative sets of parameter values

	1	2	3	4	5	6	7	8	9
κ_λ	1,400	1,000	1,800	1,400	1,400	1,400	1,400	1,400	1,400
η_λ	200	200	200	100	300	200	200	200	200
ξ	50	50	50	50	50	40	60	50	50
ϵ_λ	120	120	120	120	120	120	120	100	140

Note: This table displays 9 sets of parameter values, obtained by letting only one of the parameters deviate from the benchmark parameter specification, given by parameter set 1.

The parameter κ_λ controls the speed of mean reversion of the jump intensity process. The faster the jump intensity reverts back to its longrun level after the occurrence of an intensity jump, the lower the post-jump local average jump intensity, and the shorter (hence, less pronounced) the self-excitation effect. Consequently, a larger (smaller) value of κ_λ tends to decrease (increase) the power of the no self-excitation test. The size of the self-excitation test is not much affected by changes in κ_λ. Changes in the parameter η_λ have little impact on both the testing size and the testing power, essentially because, once a jump occurs, the diffusive increments are dominated by the jump component. The parameter ξ controls the magnitude by which the jump intensity process changes due to the self-excitation effect. Consequently, a larger (smaller) value of ξ leads to a larger (smaller) effect and increases (decreases) the testing power. At the same time, we observe that a larger (smaller) value of ξ also leads to slightly higher (lower) rejection rates for the self-excitation test, which, depending on the function g used for spot jump intensity estimation, leads to an improvement or deterioration of the testing size. Finally, a larger value of ϵ_λ means that fewer price jumps can excite the jump intensity process, hence decreases the effective sample size. Consequently, a larger (smaller) value of ϵ_λ tends to decrease (increase) the power of the no self-excitation test. As for the self-excitation test, we observe that the rejection rates slightly decrease (increase) as ϵ_λ increases (decreases).

In the online appendix, we also analyze the performance of our tests when (i) the data generating process is, in fact, a tempered stable process and (ii) when volatility jumps are present. The results show that the testing functions used in our empirical analysis don't generate false detection of self-excitation in these cases. In sum, the Monte Carlo results show that our tests have good power and size properties under reasonable sets of parameter values.

6 Conclusions

This paper first extends the concept of self-excitation in jumps to a rich class of continuous time semimartingale models accommodating infinite-activity jumps. Self-excitation in jumps relaxes the commonly adopted assumption that jump increments in continuous time financial models are serially independent, by allowing realized jumps to exert a positive feedback effect onto their own stochastic jump intensity processes. Models that allow for self-excitation in jumps are more capable of reproducing the jump clustering phenomenon observed during financial crises.

Next, we propose functionals of the log-price process and its stochastic jump intensity, that behave distinctly on the disjoint outcome sets of self-excitation in jumps and of no self-excitation in jumps, construct estimators of these functionals and derive their intricate asymptotic properties. Then, we design statistical tests under two (complementary) null hypotheses to control for both type I and type II errors. We analyze the finite sample performance of these tests in Monte Carlo simulations.

All proofs, some examples, additional Monte Carlo experiments, and the description of our empirical analysis are collected in an online appendix, see Boswijk et al. (2017).

Acknowledgements

We are very grateful to an Editor, an Associate Editor and two referees for comments and suggestions that significantly improved the paper. We are also grateful to Yacine Aït-Sahalia, Torben G. Andersen, George Tauchen, Viktor Todorov and conference and seminar participants at the North American Summer Meeting of the Econometric Society in Minneapolis, the China Meeting of the Econometric Society in Xiamen, the Econometric Society Australasian Meeting in Hobart, the Econometric Society European Meeting in Toulouse, the Quantitative Economics Doctorate (QED) Conference in Bielefeld, the Tinbergen Institute, and the University of Amsterdam for their comments and suggestions. This research was funded in part by the Netherlands Organization for Scientific Research (NWO) under grant VIDI-2009 (Laeven).

References

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

Aït-Sahalia, Y. and Jacod, J. (2009a). Estimating the degree of activity of jumps in high frequency data. *The Annals of Statistics*, 37:2202–2244.

Aït-Sahalia, Y. and Jacod, J. (2009b). Testing for jumps in a discretely observed process. *The Annals of Statistics*, 37:184–222.

Aït-Sahalia, Y. and Jacod, J. (2011). Testing whether jumps have finite or infinite activity. *The Annals of Statistics*, 39:1689–1719.

Aït-Sahalia, Y. and Jacod, J. (2012). Identifying the successive Blumenthal-Getoor indices of a discretely observed process. *The Annals of Statistics*, 40:1430–1464.

Aït-Sahalia, Y., Jacod, J., and Li, J. (2012). Testing for jumps in noisy high frequency data. *Journal of Econometrics*, 168(2):207–222.

Andersen, T. G., Fusari, N., and Todorov, V. (2015). Parametric inference and dynamic state recovery from option panels. *Econometrica*, 83:1081–1145.

Bertoin, J. (1996). *Lévy Processes*. Cambridge University Press.

Bollerslev, T. and Todorov, V. (2011). Estimation of jump tails. *Econometrica*, 79(6):1727–1783.

Bollerslev, T. and Todorov, V. (2014). Time-varying jump tails. *Journal of Econometrics*.

Boswijk, H. P., Laeven, R. J. A., and Yang, X. (2017). Supplement to "Testing for self-excitation in jumps".

Bowsher, C. G. (2007). Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141:876–912.

Brillinger, D. R. (1988). Some statistical methods for random process data from seismology and neurophysiology. *The Annals of Statistics*, 16:1–54.

Bücher, A. and Vetter, M. (2013). Nonparametric inference on Lévy measures and copulas. *The Annals of Statistics*, 41(3):1485–1515.

Christoffersen, P., Jacobs, K., and Ornthanalai, C. (2012). Dynamic jump intensities and risk premiums: Evidence from S&P500 returns and options. *Journal of Financial Economics*, 106(3):447–472.

Cont, R. and Tankov, P. (2003). *Financial modelling with jump processes*. Chapman and Hall/ CRC Press.

Corradi, V., Silvapulle, M. J., and Swanson, N. R. (2014). Consistent pretesting for jumps. Working paper.

Dungey, M., Erdemlioglu, D., Matei, M., and Yang, X. (2015). Financial flights, stock market lingkages and jump excitation. Working paper.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.

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

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

Jacod, J. and Protter, P. (2011). *Discretization of Processes*. Springer.

Jacod, J. and Shiryaev, A. (2003). *Limit Theorems for Stochastic Processes*. Springer-Verlag: Berlin.

Jacod, J. and Todorov, V. (2009). Testing for common arrivals of jumps for discretely observed multidimensional processes. *The Annals of Statistics*, 37:1792–1838.

Jacod, J. and Todorov, V. (2010). Do price and volatility jump together? *Annals of Applied Probability*, 20(4):1425–1469.

Jing, B.-Y., Kong, X.-B., and Liu, Z. (2012a). Modeling high frequency financial data by pure jump models. *The Annals of Statistics*, 40(2):759–784.

Jing, B.-Y., Kong, X.-B., Liu, Z., and Mykland, P. (2012b). On the jump activity index for semimartingales. *Journal of Econometrics*, 166(2):213–223.

Kong, X.-B., Liu, Z., and Jing, B.-Y. (2015). Testing for pure-jump processes for high-frequency data. *The Annals of Statistics*, 43(2):847–877.

Lee, S. S. and Mykland, P. A. (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies*, 21(6):2535–2563.

Li, G. and Zhang, C. (2013). Conditional jump intensity, conditional expected jump size, and relative stock price level. Working paper.

Mancini, C. (2001). Disentangling the jumps of the diffusion in a geometric jumping Brownian motion. *Giornale dell'Istituto Italiano degli Attuari*, LXIV:19–41.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3:125–144.

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., and Tita, G. E. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493):100–108.

Pan, J. (2002). The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of Financial Economics*, 63:3–50.

Reynaud-Bouret, P. and Schbath, S. (2010). Adaptive estimation for Hawkes processes; Application to genome analysis. *The Annals of Statistics*, 38:2781–2822.

Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press.

Todorov, V. (2010). Variance risk-premium dynamics: The role of jumps. *Review of Financial Studies*, 23(1):345–383.

Todorov, V. (2015). Jump activity estimation for pure-jump semimartingales via self-normalized statistics. *The Annals of Statistics*, 43(4):1831–1864.

Todorov, V. and Tauchen, G. (2010). Activity signature functions for high-frequency data analysis. *Journal of Econometrics*, 154(2):125–138.

Figures

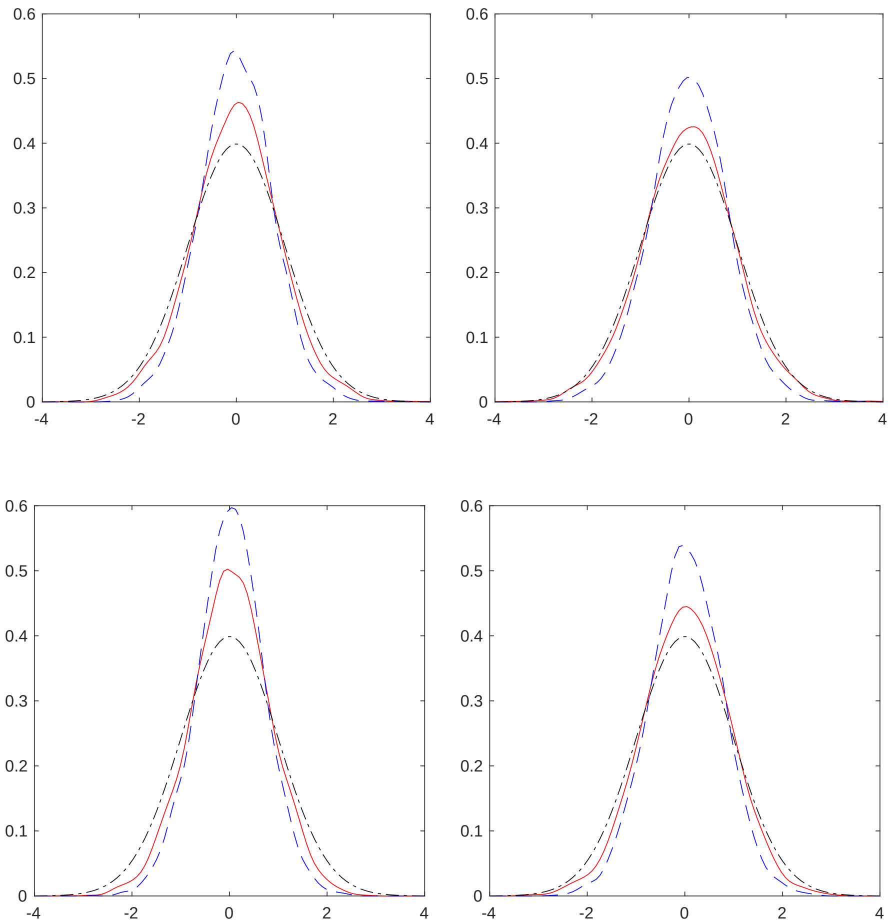

Figure 1: Standard normal probability density (black dash dotted curve) and kernel density estimates of t_n on the set $\Omega^{(0)}$ obtained from Monte Carlo simulations, with a 5-sec sampling frequency and 1-week time horizon in the upper panel and a 10-min sampling frequency and 1-year time horizon in the lower panel, and with $H = H(6,1)$ in the left panel and $H = H(0,1)$ in the right panel. The red solid curve corresponds to using $g = g_0$ in spot jump intensity estimation (see Assumption 3) and the blue dashed curve corresponds to using $g = g_{2,6}$ (see Example A.2).

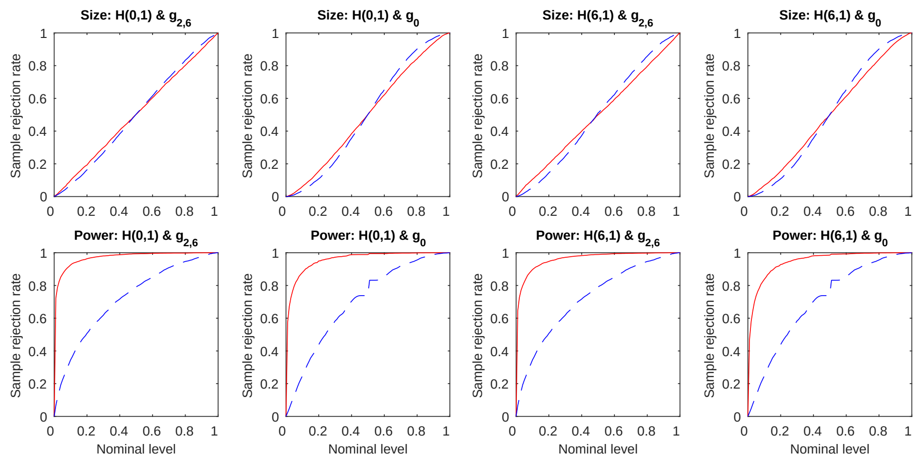

Figure 2: Size and power of the test for no self-excitation based on t_n. The x-axis shows the nominal level $\tilde{a}$ of the test, while the y-axis shows the percentages of rejection (which happens when $t_n > z_{\tilde{a}}$) in the Monte Carlo simulated sample. The red solid lines correspond to a 5-sec sampling frequency (1-week horizon), while the blue dashed lines correspond to a 10-min sampling frequency (1-year horizon).

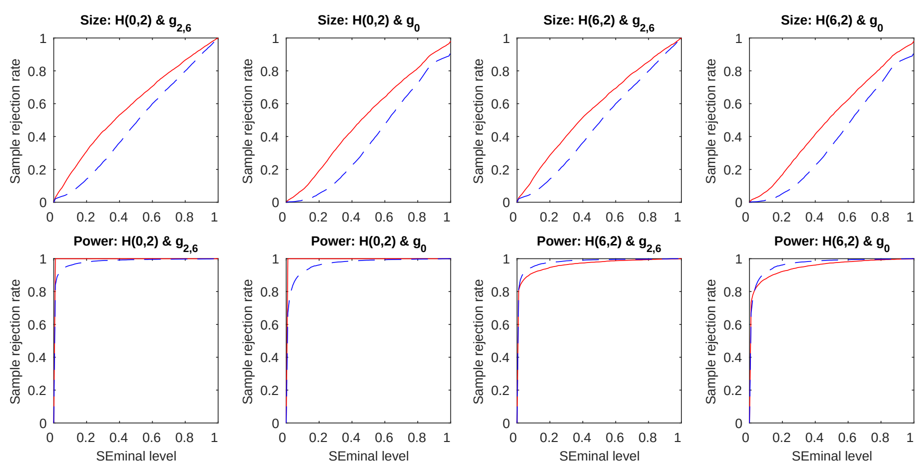

Figure 3: Size and power of the test for self-excitation based on t_n. The x-axis shows the nominal level $\tilde{a}$ of the test, while the y-axis shows the percentages of rejection (which happens when $t_n > z_{\tilde{a}}$) in the Monte Carlo simulated sample. The red solid lines correspond to a 5-sec sampling frequency (1-week horizon), while the blue dashed lines correspond to a 10-min sampling frequency (1-year horizon).