High-frequency estimation of Itô semimartingale baseline for Hawkes processes

Yoann Potiron[1], **Olivier Scaillet**[2]
Vladimir Volkov[3] and **Seunghyeon Yu**[4]

[1] *Faculty of Business and Commerce, Keio University, e-mail:* potiron@fbc.keio.ac.jp

[2] *Swiss Finance Institute, University of Geneva, e-mail:* olivier.scaillet@unige.ch

[3] *School of Business and Economics, University of Tasmania, e-mail:* vladimir.volkov@utas.edu.au

[4] *Kellogg School of Management, Northwestern University, e-mail:* seunghyeon.yu@kellogg.northwestern.edu

Abstract: We consider Hawkes self-exciting processes with a baseline driven by an Itô semimartingale with possible jumps. Under in-fill asymptotics, we characterize feasible statistics induced by central limit theory for empirical average and variance of local Poisson estimates. As a by-product, we develop a test for the absence of a Hawkes component and a test for baseline constancy. Simulation studies corroborate asymptotic theory. An empirical application on high-frequency data of the E-mini S&P500 future contracts shows that the absence of a Hawkes component and baseline constancy is always rejected.

Keywords and phrases: Hawkes tests, in-fill asymptotics, high-frequency data, Itô semimartingale, self-exciting process, time-varying baseline.

JEL codes: C14, C22, C41, C58, G00, G13

1. Introduction

Point processes are widely used in econometrics to characterize event times. The main stylized fact in this strand of literature, the presence of event clustering in time, motivates the so-called Hawkes self-exciting processes (see Hawkes (1971a,b)). If we define N_t as the aggregated number of events up to time t and λ_t its corresponding intensity, a standard definition of a Hawkes self-exciting process is given by

$$\lambda_t = \mu + \int_0^t \phi(t-s)\, dN_s, \tag{1.1}$$

where $\mu > 0$ is the Poisson baseline and ϕ is the nonnegative exciting kernel. The particular case $\phi = 0$ corresponds to a classical Poisson process, so that we can view Hawkes processes as a natural extension of Poisson processes.

An early application of Hawkes processes evolves in seismology (see Rubin (1972), Vere-Jones (1978), Ozaki (1979), Vere-Jones and Ozaki (1982) and Ogata (1978)). More recently, Ikefuji et al. (2022) analyze the impact of earthquake risk based on marked Hawkes processes. There are also applications in financial econometrics (see Yu (2004), Bowsher (2007), Embrechts, Liniger and Lin (2011) and Aït-Sahalia, Laeven and Pelizzon (2014)), finance (see Large (2007), Aït-Sahalia, Cacho-Diaz and Laeven (2015) and Fulop, Li and Yu (2015)) and quantitative finance (see Chavez-Demoulin, Davison and McNeil (2005), Bacry et al. (2013a) and Jaisson and Rosenbaum (2015)). See also Liniger (2009) and Hawkes (2018) with the references therein. More recently, Corradi, Distaso and Fernandes (2020) develop a test for conditional independence in quadratic variation of jumps. A bootstrap approach is developed in Cavaliere et al. (2023). Christensen and Kolokolov (2024) propose an unbounded intensity model for point processes. Potiron and Volkov (2025) consider estimation of latency.

There already exists successful attempts to accommodate for time-varying parameter Hawkes processes. Chen and Hall (2013) allow for a nonrandom parametric time-varying baseline. Their in-fill asymptotics based on random observation times of order n within the time interval $[0, T]$ for a fixed horizon time T exploits a single boosting of the baseline, i.e., $\lambda_t^{(n)} = \alpha_n \mu_t + \int_0^t \phi(t-s)\, dN_s^{(n)}$, where $\alpha_n \to \infty$ is a scaling sequence when $n \to \infty$. They derive a central limit theorem (CLT) for the Maximum Likelihood Estimator (MLE) of parameters related to the baseline and kernel. Clinet and Potiron (2018) consider stochastic time-varying baseline and kernel parameters in the exponential kernel case, and exploits a joint boosting of the baseline and the kernel, i.e., $\lambda_t^{(n)} = n\mu_t + \int_0^t n a_t \exp(-n b_t (t-s))\, dN_s^{(n)}$ to derive CLTs on integrated baseline and parameters with local MLE. Kwan, Chen and Dunsmuir (2023) revisit Chen and Hall (2013) in the exponential kernel case and with the same in-fill asymptotics as in Clinet and Potiron (2018), i.e., $\lambda_t^{(n)} = n\mu_t + \int_0^t n a \exp(-n b (t-s))\, dN_s^{(n)}$. They advocate the use of in-fill asymptotics for statistical inference to better match high-frequency data widely used, for example, in financial applications (Aït-Sahalia and Jacod (2014)). Under large T asymptotics, Roueff, von Sachs and Sansonnet (2016) and Roueff and Von Sachs (2019) introduce a new class of locally stationary mutually exciting processes that permits to calculate approximations of first and second order moments based on local Bartlett nonparametric estimators. A nonparametric estimation approach based on locally fitting B-splines is given by Mammen and Müller (2023), while Omi, Hirata and Aihara (2017) study a Bayesian method with nonrandom parametric time-varying baseline. Recent works also include spectral parametric estimation for misobserved Hawkes processes, i.e., when the exact locations of points are unknown but only the number of points on each bin are known, by Cheysson and Lang (2022) with a setting also covering a time-varying baseline.

Empirical evidence suggests that the baseline is time-varying. Chen and Hall (2013) report in their empirical study (Section 5.2, pp. 7–10) that goodness-of-fit results are in favor of their time-varying baseline model compared to a group of alternatives. In Figure 2 (p. 20), they document the time-varying nonrandom function for both polynomial and exponential kernel. This nonrandom path is also visible in Figure 2 (p. 3488) from Clinet and Potiron (2018). Besides, the empirical findings in the two aforementioned papers suggest that there may be a remaining stochastic effect, and Rambaldi, Pennesi and Lillo (2015) document that there are frequent intensity bursts in the baseline.

In this paper, we consider Hawkes processes with a baseline driven by an Itô semimartingale with possible jumps, namely

$$\mu_t = \mu_0 + \int_0^t b_s\, ds + \int_0^t \sigma_s\, dW_s + (\delta \mathbb{1}_{\{|\delta|\leq 1\}}) \star (\underline{\mu} - \underline{\nu})_t + (\delta \mathbb{1}_{\{|\delta|>1\}}) \star \underline{\mu}_t.$$

The Itô semimartingale baseline is not allowed by any of the aforementioned work and this model suits the three aforementioned empirical facts for the baseline intensity: time-variation, stochasticity, and bursts. For inference purposes, we consider in-fill asymptotics with joint boosting of the baseline and the kernel. We assume that the kernel satisfies the short-range condition. It extends the asymptotic analysis of Clinet and Potiron (2018) and Kwan, Chen and Dunsmuir (2023) by not imposing an exponential kernel. Here, in-fill asymptotics are desirable because we can incorporate random features of the baseline into asymptotic variances in the CLT.

Our main result (Theorem 4.1) is a feasible joint CLT of suitably rescaled empirical average and variance of local Poisson estimates. This contribution extends the papers on time-varying parameter Hawkes processes, which are either concerned with log-likelihood function or nonparametric locally stationary methods to estimate the kernel. The key ingredient in deriving our feasible CLTs is a mix of (a) decomposing the estimation error into an error originating from the time-varying baseline and another related to the Hawkes structure, (b) studying the kernel resolvent and its relations with the Laplace transform of the kernel, and (c) using the martingale representation of the intensity based on the convolution of the resolvent kernel and the martingale. The representation in

(c) extends the machinery of Lemma 3 in Bacry et al. (2013a), which considers an invariant baseline and large T asymptotics, to the time-varying baseline and in-fill asymptotics. However, the study of (a) and (b) is new. With the decomposition in (a), a truncation to remove the jumps from the Itô semimartingale baseline allows us to use Theorem IX.7.28 (pp. 590–591) in Jacod and Shiryaev (2013).

As a by-product from our main result in a high-frequency setting, we provide five main corollaries. For estimation we demonstrate: (i) estimation of the integrated intensity $\int_0^T \lambda_t dt$; (ii) estimation of the integrated baseline $\int_0^T \mu_t dt$; (iii) estimation of the integrated volatility of the baseline $\int_0^T \sigma_t^2 dt$. For testing we show: (iv) hypothesis testing for the absence of a Hawkes component; (v) hypothesis testing for baseline constancy. If N_t is a nonhomogeneous Poisson process, i.e., the intensity λ_t is nonrandom, there is a body of literature (see, e.g., Leemis (1991)) for (i). It provides consistency and CLT in asymptotics exploiting independent realizations of N_t available over $[0, T]$. We accommodate for a larger class of processes, i.e., the intensity is random but in the absence of a Hawkes component, and use in-fill asymptotics instead. Two papers provide inference for (iii), but without a Hawkes component, namely Kimura and Yoshida (2016) and Stoltenberg, Mykland and Zhang (2022). Our strategy for (iv) differs from Dachian and Kutoyants (2006), who consider parametric and non parametric composite alternatives with large T asymptotics, and Türkmen and Cemgil (2018), who develop a Bayesian approach. Finally, (ii) and (v) are new to the literature.

The remainder of this paper is organized as follows. We provide the model in Section 2, and we introduce the estimation strategy in Section 3. We give our main feasible CLT result in Section 4. We investigate estimation problems (i)-(ii)-(iii) and testing problems (iv)-(v) in Section 5. In Section 6, we carry out a finite sample analysis, which corroborates the asymptotic theory. In Section 7, an empirical application on high-frequency data of the E-mini S&P500 future contracts is presented. Finally, we provide concluding remarks in Section 8. All the proofs are gathered in Appendix A.

2. Model

In this section, we introduce Hawkes self-exciting processes with a baseline driven by an Itô semimartingale with possible jumps when the horizon T is finite.

We define N_t as a simple point process on $[0, T]$, i.e., a family $\{N(C)\}_{C \in \mathcal{B}([0,T])}$

of random variables with values in $\mathbb{N}$ where $\mathcal{B}([0,T])$ is the Borel σ-algebra on the compact space $[0,T]$, $N(C) = \sum_{i\in\mathbb{N}} \mathbb{1}_C(T_i)$ and $\{T_i\}_{i\in\mathbb{N}}$ is a sequence of $\mathbb{R}^+$-valued random event times such that, a.s. $T_0 = 0 < T_1 < \ldots < T_{N_T} < T < T_{N_T+1}$. In informal terms, N_t counts the number of events on $[0,t]$. Let $\mathcal{B} = (\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P})$ be a filtered probability space which satisfies the usual conditions. For any $t \in [0,T]$, we denote the filtration generated by some stochastic process X as $\mathcal{F}_t^X = \sigma\{X_s : 0 \le s \le t\}$. We assume that, for any $t \in [0,T]$, the filtration generated by the point process N_t is included in the main filtration, i.e., $\mathcal{F}_t^N \subset \mathcal{F}_t$. Any nonnegative $\mathcal{F}_t$-progressively measurable process $\{\lambda_t\}_{t\in[0,T]}$ such that $\mathbb{E}[N((a,b]) \mid \mathcal{F}_a] = \mathbb{E}\big[\int_a^b \lambda_s ds \,\big|\, \mathcal{F}_a\big]$ a.s. for all intervals $(a,b]$, is called an $\mathcal{F}_t$-intensity of N_t. Intuitively, the intensity corresponds to the expected number of events given the past information, i.e.,

$$\lambda_t = \lim_{u\to 0} \mathbb{E}[\frac{N_{t+u} - N_t}{u} \mid \mathcal{F}_t] \text{ a.s..}$$

For background on point processes, the reader can consult Daley and Vere-Jones (2003, 2008) and Jacod and Shiryaev (2013).

The present work is concerned with simple point processes N_t admitting an $\mathcal{F}_t$-intensity of the form

$$\lambda_t = \mu_t + \int_0^t \phi(t-s)\, dN_s. \tag{2.1}$$

Here, we have that μ_t is the $\widetilde{\mathcal{F}}_t$-Itô semimartingale baseline process with $\widetilde{\mathcal{F}}_t \subset \mathcal{F}_t$ and ϕ is the nonnegative exciting kernel. Since μ_t follows an $\widetilde{\mathcal{F}}_t$-Itô semimartingale, then we can construct a filtered extension $\overline{\mathcal{B}} = (\overline{\Omega}, \overline{\mathcal{F}}, \{\overline{\mathcal{F}}_t\}_{t\in[0,T]}, \overline{\mathbb{P}})$ on which are defined a standard Brownian motion W and a Poisson random measure $\underline{\mu}$ on $\mathbb{R}_+ \times E$, which is compensated by $\underline{\nu}(dt,dx) = dt \otimes F_t(dx)$. Here, we assume that E is an auxiliary Polish space and that F_t is σ-finite, infinite, and optional measure, having no atom. Then, the baseline μ_t has the Grigelionis representation of the form

$$\mu_t = \mu_0 + \int_0^t b_s\, ds + \int_0^t \sigma_s\, dW_s + \int_0^t \int_E (\delta(s,z)\mathbb{1}_{\{|\delta(s,z)|\le 1\}})(\underline{\mu} - \underline{\nu})(ds,dz) + \int_0^t \int_E (\delta(s,z)\mathbb{1}_{\{|\delta(s,z)|>1\}})\underline{\mu}(ds,dz). \tag{2.2}$$

Here, we have that μ_0 is $\mathcal{F}_0$-measurable and, for any $t \in [0,T]$, b_t and σ_t are $\mathbb{R}$-valued predictable processes on $(\Omega, \mathcal{F}, \{\widetilde{\mathcal{F}}_t\}_{t\in[0,T]}, \mathbb{P})$ such that both integral defined in Eq. (2.2) are well-defined, and δ is a $\mathbb{R}$-valued predictable function on

$\Omega \times \mathbb{R}_+ \times E$ such that both integral defined in Eq. (2.2) are well-defined. Although we have extended the filtered space, in the sequel we keep the original space $\mathcal{B} = (\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P})$ pretending that the Grigelionis form above is defined on it, to avoid more complicated notation. For further details of definitions and notations, see Section 1.4.3 (pp. 47-49) in Aït-Sahalia and Jacod (2014). The baseline model Eq. (2.2) is general in the sense that it is a slightly restricted version of a semimartingale. The semimartingale class is the most general class of "stochastic integrator" (Jacod and Shiryaev, 2013).

3. Estimation

In this section, we introduce the in-fill asymptotics. We also introduce empirical average and variance of local Poisson estimates.

We prefer most of the time not to write explicitly the dependence on n, and any limit theorem refers to the convergence when $n \to \infty$. For inference purposes, we consider in-fill asymptotics with joint boosting of the baseline and the kernel, i.e.,

$$\lambda_t = n\mu_t + \int_0^t n\phi(n(t-s))\, dN_s. \tag{3.1}$$

In Eq. (3.1), in-fill asymptotics are based on random observation times of order n within the time interval $[0, T]$ for a finite horizon time T. They extend the asymptotic analysis of Clinet and Potiron (2018), Kwan, Chen and Dunsmuir (2023) and Potiron and Volkov (2025), also based on joint boosting, by not imposing an exponential or a mixture of generalized gamma kernel. Christensen and Kolokolov (2024) consider boosting of the baseline to detect intensity bursts without an Hawkes component. They are different from Chen and Hall (2013) in-fill asymptotics which considers no boosting of the kernel. Here, in-fill asymptotics are desirable because we can incorporate random features of the baseline into asymptotic variances in the CLT.

For a finite horizon T, we consider $M = \lfloor T/\Delta_n \rfloor$ intervals with equal length Δ_n such that $\bigcup_{i=1}^M [(i-1)\Delta_n, i\Delta_n) \subset [0, T)$, where $\lfloor \cdot \rfloor$ denotes the floor function. For $i = 1, \ldots, M$, we define an estimator for local Poisson estimates on the i-th interval $[(i-1)\Delta_n, i\Delta_n)$ as

$$\widehat{\lambda}_i = \frac{1}{\Delta_n}(N_{i\Delta_n -} - N_{(i-1)\Delta_n}).$$

Then, we propose an estimator for empirical average and two estimators for empirical variance of local Poisson estimates as

$$\widehat{\text{Mean}} = \Delta_n \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \widehat{\lambda}_i = N_T,$$

$$\widehat{\text{Var}}_1 = \sum_{i=2}^{\lfloor T/\Delta_n \rfloor} \big(\Delta_i \widehat{\lambda}\big)^2 \mathbb{1}_{\{|\Delta_i \widehat{\lambda}| \leq \alpha \Delta_n^{-\overline{\omega}}\}},$$

$$\widehat{\text{Var}}_2 = \sum_{i=2}^{\lfloor T/(2\Delta_n) \rfloor} \left(\frac{\Delta_{2i-2}\widehat{\lambda} + \Delta_{2i-1}\widehat{\lambda}}{2} \right)^2 \mathbb{1}_{\{|(\Delta_{2i-1}\widehat{\lambda} + \Delta_{2i}\widehat{\lambda})/2| \leq \alpha \Delta_n^{-\overline{\omega}}\}}.$$

Here, we have that $\Delta_i \widehat{\lambda} = \widehat{\lambda}_i - \widehat{\lambda}_{i-1}$. We also have that $\alpha > 0$ and $\overline{\omega}$ are truncation parameters. Since the intensity explodes asymptotically, the three aforementioned estimators also diverge to infinity. The two variance estimators with a different scale are requested for the applications in Section 5. We consider a truncation in our variance estimators since they would be contaminated by the jumps otherwise.

Moreover, we define the diverging target values as

$$\text{Mean} = n \frac{1}{1 - \|\phi\|_1} \int_0^T \mu_t \, dt, \tag{3.2}$$

$$\text{Var}_1 = n^2 \frac{1}{(1 - \|\phi\|_1)^2} \int_0^T \left(\frac{2}{3}\sigma_t^2 + \frac{1}{1 - \|\phi\|_1} \frac{2}{c} \mu_t \right) dt, \tag{3.3}$$

$$\text{Var}_2 = n^2 \frac{1}{(1 - \|\phi\|_1)^2} \int_0^T \left(\frac{2}{3}\sigma_t^2 + \frac{1}{1 - \|\phi\|_1} \frac{1}{2c} \mu_t \right) dt. \tag{3.4}$$

In practice, the order of observation number n is unknown. Thus, the length of intervals Δ_n cannot be chosen directly. Instead, we can estimate it as

$$\Delta_n = \frac{c}{\sqrt{N_T}}.$$

We use $c = 0.5$, which works the best in our numerical studies.

We define the non diverging error, which is also standardized by its convergence rate, as

$$X = \begin{bmatrix} \Delta_n^{-1} n^{-1} \big(\widehat{\text{Mean}} - \text{Mean}\big) \\ \Delta_n^{-\frac{1}{2}} n^{-2} \big(\widehat{\text{Var}}_1 - \text{Var}_1\big) \\ \Delta_n^{-\frac{1}{2}} n^{-2} \big(\widehat{\text{Var}}_2 - \text{Var}_2\big) \end{bmatrix}.$$

For any $t \in [0, T]$, we define $\breve{\vartheta}_t$ as $\breve{\vartheta}_t = \frac{\mu_t}{c(1-\|\phi\|_1)^3}$, $\breve{\sigma}_t$ as $\breve{\sigma}_t = \frac{\sigma_t}{1-\|\phi\|_1}$, and w_t as

$$w_t w_t^\top = \begin{bmatrix} \breve{\vartheta}_t & 0 & 0 \\ 0 & \breve{\sigma}_t^4 + 4\breve{\sigma}_t^2 \breve{\vartheta}_t + 12\breve{\vartheta}_t^2 & \frac{29}{24}\breve{\sigma}_t^4 + \frac{3}{2}\breve{\sigma}_t^2 \breve{\vartheta}_t + \frac{3}{2}\breve{\vartheta}_t^2 \\ 0 & \frac{29}{24}\breve{\sigma}_t^4 + \frac{3}{2}\breve{\sigma}_t^2 \breve{\vartheta}_t + \frac{3}{2}\breve{\vartheta}_t^2 & 2\breve{\sigma}_t^4 + 2\breve{\sigma}_t^2 \breve{\vartheta}_t + \frac{3}{2}\breve{\vartheta}_t^2 \end{bmatrix}.$$

We also define the diverging asymptotic variance as $\mathbf{\Sigma} = n^2 \int_0^T w_t w_t^\top \, dt$. Moreover, we define the estimator of the diverging asymptotic variance as

$$\widehat{\mathbf{\Sigma}} = \begin{bmatrix} \widehat{\Sigma}_{11} & 0 & 0 \\ 0 & \widehat{\Sigma}_{22} & \widehat{\Sigma}_{23} \\ 0 & \widehat{\Sigma}_{23} & \widehat{\Sigma}_{33} \end{bmatrix}.$$

Here, the components of the matrix are defined as $\widehat{\Sigma}_{11} = \frac{2}{3}\big(\widehat{\mathrm{Var}_1} - \widehat{\mathrm{Var}_2}\big)$, $\widehat{\Sigma}_{22} = \frac{3}{4}\widehat{\kappa}_{4,1} - 3\widehat{\eta}\widehat{\kappa}_{3,1} + 9\widehat{\eta}^2\widehat{\kappa}_{2,1}$, $\widehat{\Sigma}_{23} = \frac{29}{32}\widehat{\kappa}_{4,1} - \frac{69}{8}\widehat{\eta}\widehat{\kappa}_{3,1} + \frac{63}{8}\widehat{\eta}^2\widehat{\kappa}_{2,1}$, and $\widehat{\Sigma}_{33} = \frac{3}{2}\widehat{\kappa}_{4,2} - 6\widehat{\eta}\widehat{\kappa}_{3,2} + 18\widehat{\eta}^2\widehat{\kappa}_{2,2}$. We also define $\widehat{\kappa}_{2,1}$ as $\widehat{\kappa}_{2,1} = \Delta_n^{-3}\sum_{i=2}^{\lfloor T/\Delta_n \rfloor} \widehat{\lambda}_i^2 \mathbb{1}_{\{|\Delta_i\widehat{\lambda}| \leq n\varpi_i\}}$, and $\widehat{\kappa}_{2,2}$ as $\widehat{\kappa}_{2,2} = \Delta_n^{-3}\sum_{i=2}^{\lfloor T/(2\Delta_n)\rfloor}(\frac{\widehat{\lambda}_{i-1}+\widehat{\lambda}_i}{2})^2 \mathbb{1}_{\{|(\Delta_{2i-1}\widehat{\lambda}+\Delta_{2i}\widehat{\lambda})/2| \leq n\varpi_i\}}$. Additionally, we can define $\widehat{\kappa}_{3,1}$ as $\widehat{\kappa}_{3,1} = \Delta_n^{-2}\sum_{i=2}^{\lfloor T/\Delta_n\rfloor}\widehat{\lambda}_i(\Delta_i\widehat{\lambda})^2\mathbb{1}_{\{|\Delta_i\widehat{\lambda}| \leq n\varpi_i\}}$, as well as $\widehat{\kappa}_{3,2}$ is defined as

$$\widehat{\kappa}_{3,2} = \Delta_n^{-2}\sum_{i=2}^{\lfloor T/(2\Delta_n)\rfloor}\frac{\widehat{\lambda}_{i-1}+\widehat{\lambda}_i}{2}\Big(\frac{\Delta_{2i-2}\widehat{\lambda}+\Delta_{2i-1}\widehat{\lambda}}{2}\Big)^2\mathbb{1}_{\{|(\Delta_{2i-1}\widehat{\lambda}+\Delta_{2i}\widehat{\lambda})/2| \leq n\varpi_i\}}.$$

Finally, we define $\widehat{\kappa}_{4,1}$ as $\widehat{\kappa}_{4,1} = \Delta_n^{-1}\sum_{i=1}^{\lfloor T/\Delta_n\rfloor}(\Delta_i\widehat{\lambda})^4\mathbb{1}_{\{|\Delta_i\widehat{\lambda}| \leq n\varpi_i\}}$, but also $\widehat{\kappa}_{4,2}$ is defined as $\widehat{\kappa}_{4,2} = \Delta_n^{-1}\sum_{i=1}^{\lfloor T/(2\Delta_n)\rfloor}(\frac{\Delta_{2i-2}\widehat{\lambda}+\Delta_{2i-1}\widehat{\lambda}}{2})^4\mathbb{1}_{\{|(\Delta_{2i-1}\widehat{\lambda}+\Delta_{2i}\widehat{\lambda})/2| \leq n\varpi_i\}}$, and $\widehat{\eta}$ as $\widehat{\eta} = \frac{2}{3}\frac{\Delta_n^2(\widehat{\mathrm{Var}_1}-\widehat{\mathrm{Var}_2})}{\widehat{\mathrm{Mean}}}$.

4. Theory

In this section, we start with showing an existence result for Hawkes processes with a baseline driven by an Itô semimartingale. Then, our main result characterizes feasible statistics induced by central limit theory for empirical average and variance of local Poisson estimates.

As the Itô semimartingale is not bounded, it can in particular become nonpositive. This is incompatible with the constraint of a positive baseline for getting a well-defined Hawkes process. Let us introduce a set of assumptions required for the existence of Hawkes processes with a time-varying baseline driven by an Itô semimartingale.

Assumption 1.

(a) The baseline is positive a.e. a.s., i.e., $\mathbb{P}(\mu_t > 0 \forall t \in [0,T]) = 1$.
(b) The baseline is integrable a.s., i.e., $\mathbb{P}(\int_0^T \mu_s \, ds < \infty) = 1$.
(c) For any $0 \leq t \leq T$, we have $\mathcal{F}_t = \widetilde{\mathcal{F}}_t \vee \mathcal{F}_t^{\underline{N}}$, where the filtration $\widetilde{\mathcal{F}}_t$ is independent from the other filtration $\mathcal{F}_t^{\underline{N}}$. We also have $\underline{N}_t$ is a 2-dimensional $\mathcal{F}_t$-adapted Poisson process of intensity 1 that generates N_t, i.e., $N_t = \int_{[0,t]\times\mathbb{R}} \mathbb{1}_{[0,\lambda_s]}(x)\underline{N}(ds \times dx)$.

(d) The L^1 norm of the kernel is strictly less than one, i.e., $\|\phi\|_1 = \int_0^\infty \phi(t)dt < 1$.

Assumption 1(a) implies that the point process is well-defined and is weaker than the assumption from Clinet and Potiron (2018), who requires that the baseline belongs to a compact space. Assumption 1(b) is already an assumption in the simpler case of heterogeneous Poisson processes without a self-exciting kernel (see, e.g., Daley and Vere-Jones (2003)), and is also required to establish existence in Clinet and Potiron (2018) (see Assumption E (ii), p. 3476). Assumption 1(c) corresponds to Poisson imbedding (Brémaud and Massoulié (1996), Section 3, pp. 1571-1572) and assumes independence between $\widetilde{\mathcal{F}}_t$ and $\mathcal{F}_t^N$. It is already required in Clinet and Potiron (2018) (see the last sentence before Theorem 5.1, p. 3476). In particular, N_t is defined as the point process counting the points of $\underline{N}$ below the curve $t \to \lambda_t$. Finally, Assumption 1(d) is necessary to obtain a stationary intensity with finite first moment (see Lemma 1 (p. 495) in Hawkes and Oakes (1974) and Theorem 1 (p. 1567) in Brémaud and Massoulié (1996)) in the time-invariant classical problem where N_t starts from $-\infty$.

We provide now our existence result establishing a new theory for jump processes. It is obtained by extending the machinery of Poisson imbedding for time-invariant Hawkes processes (Brémaud and Massoulié (1996)) to the time-varying case. It also complements Theorem 5.1 (p. 3476) in Clinet and Potiron (2018) in which the kernel is exponential.

Proposition 4.1. *Under Assumption 1, there exists an $\mathcal{F}_t$-adapted simple point process N_t with an $\mathcal{F}_t$-intensity of the form Eq. (2.1).*

We define an alternative drift as $b_t' = b_t - \int_E \delta(t,z)\mathbb{1}_{\{|\delta(t,z)|\leq 1\}}F_t(dz)$ for any $t \in [0,T]$. Finally, we define $V_a^b(f)$ as the total variation of f from a to b.

Let us introduce a set of assumptions required for the CLT for empirical average and variances of local Poisson estimates.

Assumption 2.

(a) The kernel satisfies the *short-range condition*, i.e., $\int_0^\infty t\phi(t)dt < \infty$.
(b) There exists a $c > 0$ such that $n\Delta_n^2 \to^{\mathbb{P}} c$.
(c) There exists a $\beta \in [0,1)$ such that $\sup_{0\leq t\leq T} \int \min(|x|^r, 1)F_t(dx)$ is a.s. finite.
(d) The truncation level satisfies $\varpi \in (0, 1/(4-2\beta))$.
(e) For any $k \in \mathbb{N}_*$ and any $t \in [0,T]$, we have $\mathbb{E}|b_t|^k < \infty, \mathbb{E}|\sigma_t|^k < \infty$.
(f) We have that $\mathbb{E}\big[\exp\big(\frac{1}{2}\int_0^T \frac{(b_s')^2}{\sigma_s^2}ds\big)\big] < \infty$.

(g) The volatility process is a semimartingale, i.e., $\sigma_t^2 = A_t + M_t^{(\sigma)}$, where A_t is a $\mathcal{F}_t$-adapted cadlag process with finite variation and $M^{(\sigma)}$ is a square-integrable martingale. Moreover, $\mathbb{E}|V_0^T(A)|^k < \infty$ and $\mathbb{E}|\sigma_t - \sigma_s|^k \leq C(t-s)^{k\gamma}$ for a $\gamma > 0$ and any $k \in \mathbb{N}_*$.

Assumption 2(a) is required to obtain the asymptotic properties of the resolvent kernel and already appears in Exp. (8) (p. 125) from Brémaud and Massoulié (2001). Assumption 2(b) is natural for local estimation. Assumptions 2(c) and 2(d) are due to the presence of jumps, and can be found in Hypothesis (L-s) (p. 522) from Jacod (2008). Assumption 2(e) is used in the proof of jumps and also in the proof of the CLT. Assumption 2(f) is required to apply the Girsanov theorem in our proofs. Assumption 2(g) is used in the proof of the CLT.

We define T as a 3-dimensional standard normal vector. We denote $\xrightarrow{\mathcal{L}-s}$ as the $\mathcal{F}_t$-stable convergence for the Skorokhod topology on $\mathbb{D}([0,T], \mathbb{R}^3)$. We provide now the result of CLT for empirical average and variances of local Poisson estimates.

Theorem 4.1. *Under Assumptions 1 and 2, there is a canonical* 3*-dimensional standard Wiener extension of* $\mathcal{B}$*, with the canonical standard Wiener process* $\widetilde{W}_t$ *such that*

$$X \xrightarrow{\mathcal{L}-s} \int_0^T w_t d\widetilde{W}_t,$$

We also have the consistency of the estimator of non diverging asymptotic variance, i.e.,

$$n^{-2}(\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}) \to^{\mathbb{P}} 0. \tag{4.1}$$

Moreover, we show the normalized CLT with feasible variance, i.e.,

$$n^{-2}\widehat{\mathbf{\Sigma}}^{-1/2} X \xrightarrow{\mathcal{L}-s} \mathrm{T}. \tag{4.2}$$

5. Applications

In this section, we investigate estimation problems (i), (ii) and (iii). Testing problems (iv) and (v) are also investigated.

5.1. Estimation of integrated intensity

We start with the estimation of the diverging integrated intensity

$$\Lambda_T = n \int_0^T \lambda_t \, dt. \tag{5.1}$$

We consider the case when N_t is not a Hawkes process, i.e., $\phi(t) = 0$. We have applications in management science where the integrated intensity can be interpreted as the arrival rate in a queuing system. We also have applications in computer networks for the expected internet traffic, and seismology for the expected number of earthquakes. In finance, Hawkes process find applications, for example, in modeling financial contagion (Aït-Sahalia, Cacho-Diaz and Laeven (2015)) and microstructure noise (Bacry et al. (2013b)), managing risk (Aït-Sahalia and Laeven (2023)), and measuring order latency (Potiron and Volkov (2025)).

There are numerous work estimating Eq. (5.1), including consistency and CLT. These concern the case when N_t is a nonhomogeneous Poisson process, i.e., the intensity λ_t is nonrandom. They are based on asymptotics where several independent realizations of independent realizations of N_t are available over $[0, T]$, as opposed to the in-fill asymptotics of this paper. A pioneer work for nonparametric estimation of Eq. (5.1) is Leemis (1991). A different nonparametric approach based on kernel estimator is suggested by Lewis and Shedler (1976). A wavelet-based nonparametric method can be found in Kuhl and Bhairgond (2000). Parametric methods include and are not limited to Lee, Wilson and Crawford (1991), Kuhl, Wilson and Johnson (1997), Kuhl and Wilson (2000) and Kao and Chang (1988). Finally, a semiparametric framework is considered in Kuhl and Wilson (2001).

We define the asymptotic variance of the non diverging mean estimator $n^{-1}\widehat{\text{Mean}}$ as

$$\text{AVar}(n^{-1}\widehat{\text{Mean}}) = c^{-1}\int_0^T \lambda_t \, dt.$$

Moreover, we define the asymptotic variance estimator of $\widehat{\text{Mean}}$ as

$$\widehat{\text{AVar}}(\widehat{\text{Mean}}) = \frac{\widehat{\text{Mean}}}{n^2\Delta_n^2}.$$

The following corollary gives the CLT for the mean estimator.

Corollary 5.1. *Under Assumptions 1 and 2 and if we assume that $\phi(t) = 0$, we have*

$$\frac{\Delta_n^{-1} n^{-1}(\widehat{\text{Mean}} - \Lambda_T)}{\sqrt{\text{AVar}(n^{-1}\widehat{\text{Mean}})}} \xrightarrow{\mathcal{L}-s} \mathcal{N}(0, 1). \tag{5.2}$$

Moreover, we show the normalized CLT with feasible variance, i.e.,

$$\frac{\Delta_n^{-1}(\widehat{\text{Mean}} - \Lambda_T)}{\sqrt{\widehat{\text{AVar}}(\widehat{\text{Mean}})}} \xrightarrow{\mathcal{L}-s} \mathcal{N}(0, 1). \tag{5.3}$$

5.2. Estimation of integrated baseline

We continue with the estimation of the diverging integrated baseline

$$B_T = n \int_0^T \mu_t \, dt. \tag{5.4}$$

We note that when N_t is not a Hawkes process, the integrated baseline is equal to the integrated intensity. To the best of our knowledge, there is no method to estimate the integrated baseline in Eq. (5.4).

We first estimate the L^1 norm of the kernel as

$$\widehat{\|\phi\|_1} = 1 - \sqrt{\frac{3\widehat{\text{Mean}}}{2\Delta_n^2\left(\widehat{\text{Var}_1} - \widehat{\text{Var}_2}\right)}}. \tag{5.5}$$

In Eq. (5.5), we use the estimator from Hardiman and Bouchaud (2014) and we replace their variance by $\frac{2}{3}\left(\widehat{\text{Var}_1} - \widehat{\text{Var}_2}\right)$ since we have a time-varying baseline in this paper. Then, we can estimate the diverging integrated baseline as

$$\widehat{B}_T = (1 - \widehat{\|\phi\|_1})\widehat{\text{Mean}}. \tag{5.6}$$

We define the asymptotic variance of $\widehat{\|\phi\|_1}$ as

$$\text{AVar}(\widehat{\|\phi\|_1}) = \frac{\nabla f_1(\mathbf{x})^\top \mathbf{\Sigma} \nabla f_1(\mathbf{x})}{n^2},$$

where $f_1(\mathbf{x}) = 1 - \sqrt{\frac{3n\text{Mean}}{2c(x_2 - x_3)}}$, and $\mathbf{x} = [0, \text{Var}_1, \text{Var}_2]^\top$. Moreover, we define the estimator of the asymptotic variance as

$$\widehat{\text{AVar}}(\widehat{\|\phi\|_1}) = \frac{\nabla f_2(\widehat{\mathbf{x}})^\top \widehat{\mathbf{\Sigma}} \nabla f_2(\widehat{\mathbf{x}})}{n^2},$$

where $f_2(\mathbf{x}) = 1 - \sqrt{\frac{3\widehat{\text{Mean}}}{2\Delta_n^2(x_2 - x_3)}}$, and $\widehat{\mathbf{x}} = [0, \widehat{\text{Var}_1}, \widehat{\text{Var}_2}]^\top$. In the following corollary, we give the CLT for the integrated baseline.

Corollary 5.2. *Under Assumptions 1 and 2, we have*

$$\frac{\Delta_n^{-\frac{1}{2}} n^{-1}(\widehat{B}_T - B_T)}{\sqrt{\text{AVar}(\widehat{\|\phi\|_1}) n^{-2}\text{Mean}^2}} \xrightarrow{\mathcal{L}-s} \mathcal{N}(0,1). \tag{5.7}$$

Moreover, we show the normalized CLT with feasible variance, i.e.,

$$\frac{\Delta_n^{-\frac{1}{2}} (\widehat{B}_T - B_T)}{\sqrt{\widehat{\text{AVar}}(\widehat{\|\phi\|_1})\widehat{\text{Mean}}^2}} \xrightarrow{\mathcal{L}-s} \mathcal{N}(0,1). \tag{5.8}$$

5.3. Integrated volatility of the baseline

The third application is the estimation of the diverging integrated volatility of the baseline

$$\mathrm{IV} = n^2 \int_0^T \sigma_t^2 \, dt. \tag{5.9}$$

It can be seen as a measure of risk related to the integrated baseline. There are two general papers providing estimators of IV in the absence of a Hawkes component. Kimura and Yoshida (2016) provide in their Theorem 1 a general CLT result on estimation of correlation between two continuous Itô semimartingales. In their Theorem 2, the methodology is applied to the case where the intensity follows a continuous Itô semimartingale. Stoltenberg, Mykland and Zhang (2022) propose consistent estimators and CLT between one or more spot parameters associated with Itô semimartingales. Estimation of IV is treated in Section 4 (pp. 13–17). The use of Theorem 2.4 (p. 6) allows the authors to deduce consistency of their estimator in Corollary 4.3 (p. 16). There is no CLT, and they do not provide any asymptotic variance form. On the other hand, we do provide a feasible CLT, in a more general framework with the presence of a Hawkes component.

We estimate IV as

$$\widehat{\mathrm{IV}} = (1 - \widehat{\|\phi\|_1})^2 \big(2\widehat{\mathrm{Var}}_2 - \frac{1}{2}\widehat{\mathrm{Var}}_1\big).$$

We define the asymptotic variance of the non diverging integrated volatility $n^{-2}\widehat{\mathrm{IV}}$ as

$$\mathrm{AVar}(n^{-2}\widehat{\mathrm{IV}}) = \frac{\nabla g_1(\mathbf{x})^\top \mathbf{\Sigma} \nabla g_1(\mathbf{x})}{n^2},$$

where $g_1(\mathbf{x}) = \big(\frac{3n\mathrm{Mean}}{2c(x_2 - x_3)}\big)\big(2x_3 - \frac{1}{2}x_2\big)$. Moreover, we define the asymptotic variance estimator of $\widehat{\mathrm{IV}}$ as

$$\widehat{\mathrm{AVar}}(\widehat{\mathrm{IV}}) = \nabla g_2(\widehat{\mathbf{x}})^\top \widehat{\mathbf{\Sigma}} \nabla g_2(\widehat{\mathbf{x}}),$$

where $g_2(\mathbf{x}) = \big(\frac{3\widehat{\mathrm{Mean}}}{2\Delta_n^2(x_2 - x_3)}\big)\big(2x_3 - \frac{1}{2}x_2\big)$. The following corollary gives the CLT of the integrated volatility of the baseline.

Corollary 5.3. *Under Assumptions 1 and 2, we have*

$$\frac{\Delta_n^{-\frac{1}{2}} n^{-2} \big(\widehat{\mathrm{IV}} - \mathrm{IV}\big)}{\sqrt{\mathrm{AVar}(n^{-2}\widehat{\mathrm{IV}})}} \xrightarrow{\mathcal{L}-s} \mathcal{N}(0,1).$$

Moreover, we show the normalized CLT with feasible variance, i.e.,

$$\frac{\Delta_n^{-\frac{1}{2}} n^{-2}\big(\widehat{\mathrm{IV}} - \mathrm{IV}\big)}{\sqrt{\widehat{\mathrm{AVar}}(\widehat{\mathrm{IV}})}} \xrightarrow{\mathcal{L}-s} \mathcal{N}(0,1). \tag{5.10}$$

5.4. Test for the absence of a Hawkes component

In this part, we develop a test for the absence of a Hawkes component. We consider a Wald test, which is based on the estimation of the L^1 norm of the kernel. Two related papers are also providing similar tests, but our strategy is different. Dachian and Kutoyants (2006) propose a test for the absence of a Hawkes component based on parametric and non parametric composite alternatives under large T asymptotics. Their framework is simpler since they consider stationary Poisson process with known intensity under the null hypothesis. Türkmen and Cemgil (2018) give a Bayesian approach based on marginal likelihood estimation. They restrict to an homogeneous Poisson process under the null hypothesis, and a Hawkes process with exponential kernel under the alternative hypothesis.

We define respectively the null hypothesis and the alternative hypothesis as

$$\begin{aligned} H_0 &: \{\text{absence of a Hawkes component, i.e., } \|\phi\| = 0\}, \\ H_1 &: \{\text{presence of a Hawkes component, i.e., } \|\phi\| > 0\}. \end{aligned}$$

Let our test statistic be

$$S = \frac{\Delta_n^{-1}\widehat{\|\phi\|_1^2}}{\widehat{\mathrm{AVar}}(\widehat{\|\phi\|_1})}. \tag{5.11}$$

We define $q(u)$ as the quantile function of the chi-squared distribution with one degree of freedom. The following corollary gives the limit theory of the test for the absence of a Hawkes component.

Corollary 5.4. *We assume that Assumptions 1 and 2 hold. Then, the test statistic S converges in distribution to a chi-squared random variable with one degree of freedom under the null hypothesis H_0 and is consistent under the alternative hypothesis H_1, i.e., for any $0 < \alpha < 1$, we have*

$$\begin{aligned} \mathbb{P}(S > q(1-\alpha) \mid H_0) &\to \alpha, \\ \mathbb{P}(S > q(1-\alpha) \mid H_1) &\to 1. \end{aligned}$$

5.5. Test for baseline constancy

Finally, we introduce a test for baseline constancy. To the best of our knowledge, there is no related test. We consider a test based on the Hausman principle (Hausman (1978)), which is based on the difference between two estimators, one that is efficient but not robust to the deviation being tested, and one that is robust but not as efficient (Aït-Sahalia and Xiu (2019), Clinet and Potiron (2019)).

We define respectively the null hypothesis and the alternative hypothesis as

$$\begin{aligned} H_0' &: \{\text{The baseline } \mu_t \text{ is constant on } [0,T]\}, \\ H_1' &: \{\text{The baseline } \mu_t \text{ is not constant on } [0,T]\}. \end{aligned}$$

We first define a variance estimator in case of baseline constancy as

$$\widehat{\mathrm{Var}} = \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \left(\widehat{\lambda}_i - \frac{\widehat{\mathrm{Mean}}}{T} \right)^2. \tag{5.12}$$

Following Hardiman and Bouchaud (2014), we then estimate the L^1 norm of the kernel in case of baseline constancy as

$$\widehat{\|\phi\|_1^H} = 1 - \sqrt{\frac{\widehat{\mathrm{Mean}}}{\Delta_n^2 \widehat{\mathrm{Var}}}}. \tag{5.13}$$

We also define $\widehat{\vartheta}$ as $\widehat{\vartheta} = \widehat{\mathrm{Var}}_1/(2T)$, and $\widehat{\mathbf{\Sigma}}'$ as

$$\widehat{\mathbf{\Sigma}}' = \begin{bmatrix} \widehat{\vartheta} T & 0 & 0 & 0 \\ 0 & 2\widehat{\vartheta}^2 T & 4\widehat{\vartheta}^2 T & \widehat{\vartheta}^2 T \\ 0 & 4\widehat{\vartheta}^2 T & 12\widehat{\vartheta}^2 T & \frac{3}{2}\widehat{\vartheta}^2 T \\ 0 & \widehat{\vartheta}^2 T & \frac{3}{2}\widehat{\vartheta}^2 T & \frac{3}{2}\widehat{\vartheta}^2 T \end{bmatrix}.$$

Finally, we define the asymptotic variance of $\widehat{\|\phi\|_1} - \widehat{\|\phi\|_1^H}$ as

$$\widehat{\mathrm{AVar}}(\widehat{\|\phi\|_1} - \widehat{\|\phi\|_1^H}) = \big(\nabla f(\widehat{\mathbf{x}}_1) - \nabla g(\widehat{\mathbf{x}}_2)\big)^\top \widehat{\mathbf{\Sigma}}_n' \big(\nabla f(\widehat{\mathbf{x}}_1) - \nabla g(\widehat{\mathbf{x}}_2)\big).$$

Here, we have that $f(\mathbf{x}) = 1 - \sqrt{\frac{3}{2} \frac{\widehat{\mathrm{Mean}}}{\Delta_n^2 (x_3 - x_4)}}$, $\widehat{\mathbf{x}}_1 = \begin{bmatrix} 0, & 0, & \widehat{\mathrm{Var}}_1, & \widehat{\mathrm{Var}}_2 \end{bmatrix}^\top$, $g(\mathbf{x}) = 1 - \sqrt{\frac{\widehat{\mathrm{Mean}}}{\Delta_n^2 x_2}}$ and $\widehat{\mathbf{x}}_2 = \begin{bmatrix} 0, & \widehat{\mathrm{Var}}, & 0, & 0 \end{bmatrix}^\top$.

Let our test statistic be

$$S' = \frac{\Delta_n^{-1} (\widehat{\|\phi\|_1} - \widehat{\|\phi\|_1^H})^2}{\widehat{\mathrm{AVar}}(\widehat{\|\phi\|_1} - \widehat{\|\phi\|_1^H})}.$$

We first give the following CLT.

Proposition 5.1. *Under Assumptions 1 and 2 and H_0', we have*

$$(\widehat{\mathbf{\Sigma}}_n')^{-\frac{1}{2}} \begin{bmatrix} \Delta_n^{-1}\big(\widehat{\text{Mean}} - \text{Mean}\big) \\ \Delta_n^{-\frac{1}{2}}\big(\widehat{\text{Var}} - \text{Var}\big) \\ \Delta_n^{-\frac{1}{2}}\big(\widehat{\text{Var}_1} - \text{Var}_1\big) \\ \Delta_n^{-\frac{1}{2}}\big(\widehat{\text{Var}_2} - \text{Var}_2\big) \end{bmatrix} \xrightarrow{\mathcal{L}-s} \mathcal{MN}\left(\mathbf{0}, \mathbf{I}\right). \tag{5.14}$$

It implies that

$$\frac{\Delta_n^{-\frac{1}{2}}(\widehat{\|\phi\|_1} - \widehat{\|\phi\|_1^H})}{\sqrt{\widehat{\text{AVar}}(\widehat{\|\phi\|_1} - \widehat{\|\phi\|_1^H})}} \xrightarrow{\mathcal{L}-s} \mathcal{N}(0,1). \tag{5.15}$$

The following corollary gives the limit theory of the test for baseline constancy.

Corollary 5.5. *We assume that Assumptions 1 and 2 hold. Then, the test statistic S' converges in distribution to a chi-squared random variable with one degree of freedom under the null hypothesis H_0' and is consistent under the alternative hypothesis H_1', i.e., for any $0 < \alpha < 1$, we have*

$$\mathbb{P}(S' > q(1-\alpha) \mid H_0') \to \alpha,$$
$$\mathbb{P}(S' > q(1-\alpha) \mid H_1') \to 1.$$

6. Simulation studies

In this section, we conduct simulation studies to document how the estimators and tests from Section 5 behave.

6.1. Simulation design

We consider the following simulation design to be as close as possible from the data application in finance. We set $T = 1$, i.e., 6.5 hour long day of trading. The order of the observation number n varies from 50,000 to 1,000,000. With these realistic values, the simulation design allows for both less traded and highly traded stocks. The number of replications is equal to 1,000. We use the python package `tick` (Bacry et al., 2017) for the generation of the point process.

We define the intensity process as

$$\lambda_t \;=\; n(1 - \|\phi\|_1)\big(\mu_t^{\text{C}} + \mu_t^{\text{B}}\big) + \int_0^t n\phi(n(t-s))dN_s. \tag{6.1}$$

Here, the component of the baseline μ_t^{C} satisfies a square root process (SRP)

$$d\mu_t^{\mathrm{C}} \;=\; 30(b_t - \mu_t^{\mathrm{C}})dt + 3\sqrt{\mu_t^{\mathrm{C}}}dW_t. \tag{6.2}$$

Here, b_t is a solution of the ordinary differential $dr_t = 30(b_t - r_t)dt$ with inverse J-shape r_t defined as

$$r_t = 20\left((t-0.53)^4 + \frac{1}{24}\right),$$

and $\mu_0^{\mathrm{C}} = r_0$. We have that the drift term in Eq. (6.2) ensures mean reversion of μ_t^{C} to the process b_t. Moreover, b_t pushes μ_t^{C} to follow the inverse J-shape nonrandom term r_t. In Eq. (6.2), the diffusion term $\sqrt{\mu_t^{\mathrm{C}}}dW_t$ is the random fluctuation. The Feller condition (Feller, 1951) is satisfied with $30 \times b_t \geq 3^2$ for any $t \in [0, T]$, thus μ_t^{C} is positive.

In Eq. (6.2), μ_t^{B} are the intensity bursts (Rambaldi, Filimonov and Lillo (2018)). They are defined as a sudden occurrence of a big number of exogenous points for a short period of time, i.e., around one second. The arrival time of bursts z_i is sampled from an homogeneous Poisson process with rate $2/T$. The size of the bursts Z_i are drawn from $\max(\mathcal{N}(200n, (50n)^2), 50n)$. The intensity bursts have the form

$$\mu_t^{\mathrm{B}} = \sum_{z_i \leq t} Z_i \mathbb{1}_{\{(t-z_i)\in[0,1/(3600\times 6.5)]\}}. \tag{6.3}$$

The parameter values are taken from our empirical application and the results from Rambaldi, Filimonov and Lillo (2018) (p. 6), where the authors report an average number of bursts between 1.95–3.25 for a 6.5-hour period.

In Eq. (6.1), we consider an exponential kernel defined as $\phi(t) = 1.6e^{-2t}$ and a power kernel defined as $\phi(t) = 1.6(1+t)^{-3}$. With these kernel values, the L^1 norm is equal to $\|\phi\|_1 = 0.8$, which is the average value that we obtain in our own empirical application and in the results of Filimonov and Sornette (2012). Finally, we set the truncation level as

$$\varpi = \Delta_n^{-\frac{1}{4}} \sqrt{\frac{1}{\lfloor T/\Delta_n \rfloor} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} (\Delta_i \widehat{\lambda})^2}.$$

We consider the following variety of models to disentangle the effects. First, we set Model 1 as a null kernel and a constant baseline, i.e., $\lambda_t = n$. Second, we set Model 2 as a null kernel and a J-shape baseline, i.e., $\lambda_t = 20\big((t-0.53)^4 + \frac{1}{24}\big)n$. Third, we set Model 3 as a null kernel and a J-shape + SRP + burst baseline, i.e., $\lambda_t = n(\mu_t^{\mathrm{C}} + \mu_t^{\mathrm{B}})$. Then, we set Model 4 as an exponential kernel and a constant baseline, i.e., $\lambda_t = n + \int_0^t n\phi(n(t-s))dN_s$. We also set Model 5 as an exponential kernel and a J-shape baseline, i.e., $\lambda_t = n\mu_t + \int_0^t n\phi(n(t-s))dN_s$ where $\mu_t = 20(1 - \|\phi\|_1)\big((t-0.53)^4 + \frac{1}{24}\big)$. We set Model 6 as an exponential kernel and a J-shape + SRP + burst baseline, i.e., $\lambda_t = n(1 - \|\phi\|_1)(\mu_t^{\mathrm{C}} + \mu_t^{\mathrm{B}}) + \int_0^t n\phi(n(t-s))dN_s$. We set Model 7 as a power kernel, and a constant baseline as $\lambda_t = n + \int_0^t n\phi(n(t-s))dN_s$. We set Model 8 as a Power kernel and a J-shape baseline, i.e., $\lambda_t = n\mu_t + \int_0^t n\phi(n(t-s))dN_s, \mu_t = 20(1 - \|\phi\|_1)\big((t-0.53)^4 + \frac{1}{24}\big)$. Finally, we set Model 9 as a power kernel and a J-shape + SRP + burst baseline, i.e., $\lambda_t = n(1 - \|\phi\|_1)(\mu_t^{\mathrm{C}} + \mu_t^{\mathrm{B}}) + \int_0^t n\phi(n(t-s))dN_s$. These models are summarized in Table 1.

TABLE 1
Summary of models.

	Baseline Model (μ_t)		
Kernel	Constant	J-shape	J-shape + SRP + burst
Null	Model 1	Model 2	Model 3
Exponential	Model 4	Model 5	Model 6
Power	Model 7	Model 8	Model 9

In general, the intensity bursts μ^{B} follow Eq. (6.3), but in the case of power kernel, we first generate points without the burst and then add points whose intensity follows $(1 - \|\phi\|_1)^{-1}\mu^{\mathrm{B}}$. It is due to the implemented function in the package `tick` taking over a day to generate points when there is a burst. However, it does not give any significant differences in the results.

Figure 1 provides a comparison between simulated intensity with Model 9 (left panel) and intensity based on AAPL (Apple) data on April 1st 2016 (right panel). The intensity is obtained from one-minute intervals. The simulated process captures the U-shaped pattern and intensity burst well; it also exhibits some random fluctuation of the baseline intensity. These patterns can also be seen in the data that justify our simulation design being realistic.

6.2. Asymptotic approximation

Table 2 and Figure 2 report the summary statistics and the histogram for the integrated baseline with Models 1-3. The order of the observation number n is 150,000 and 1,000,000. The absolute value of the mean ranges from 1% to 15%, with an average of 5%. It has an average of 3% for the statistics with unfeasible variance, and an average of 7% for the statistics with feasible variance. Overall, the mean is adequate, especially when the variance is feasible and when n increases. The variance ranges from 101% to 107%, with an average of 104%.

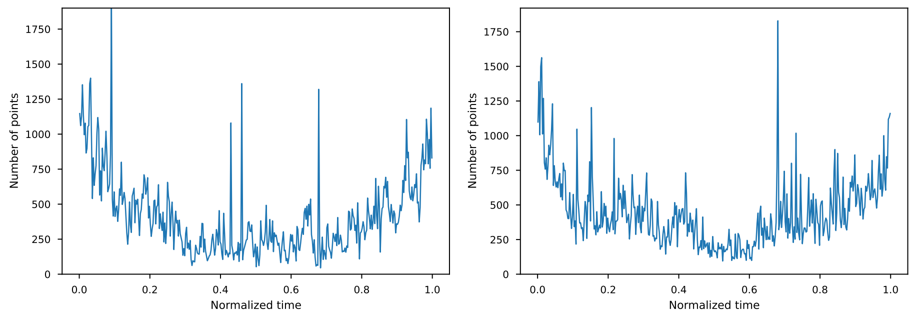

Fig 1: Comparison between simulated intensity with Model 9 (left panel) and intensity based on AAPL data on April 1st 2016 (right panel).

It has an average of 102% for the statistics with unfeasible variance, and an average of 106% for the statistics with feasible variance. Overall, the variance is close to unity.

Table 2

Summary statistics for the integrated intensity with Models 1-3. The order of the observation number n is 150,000 and 1,000,000, and the number of replications is 1,000.

n	150,000				1,000,000			
Variance	Unfeasible		Feasible		Unfeasible		Feasible	
Model	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
Model 1	-0.0117	1.0398	-0.0215	1.0665	0.0318	1.0086	0.0842	1.0192
Model 2	0.0329	1.0183	0.0640	1.0406	-0.0176	1.0331	-0.0297	1.0575
Model 3	-0.0679	1.0435	-0.1541	1.0625	-0.0360	1.0105	0.0723	1.0119

Table 3 and Figure 3 report the summary statistics and the histogram for the integrated baseline with Models 4-9. The order of the observation number n is 150,000 and 1,000,000. The absolute value of the mean ranges from 2% to 28%, with an average of 10%. It has an average of 5% for the statistics with unfeasible variance, and an average of 15% for the statistics with feasible variance. Overall, the statistics are slightly biased, especially when the variance is unfeasible. However, the bias gets smaller when n increases. The variance ranges from 98% to 109%, with an average of 103%. It has an average of 101% for the statistics with unfeasible variance, and an average of 105% for the statistics with feasible variance. Overall, the variance is close to unity.

Table 4 and Figure 4 report the summary statistics and the histogram for the integrated volatility of the baseline with Models 1-9. The order of the observation number n is 150,000 and 1,000,000. The absolute value of the mean ranges

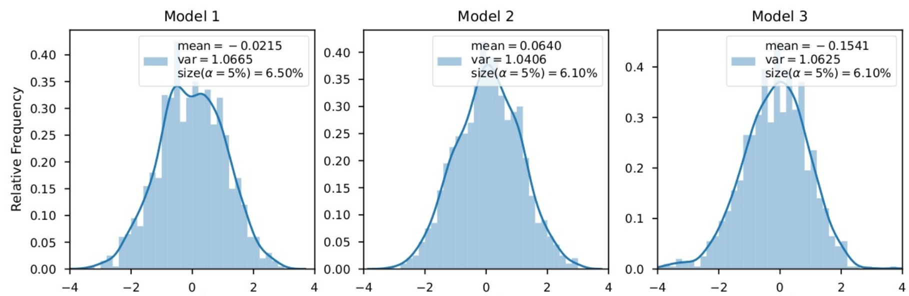

Fig 2: Histogram of the normalized CLT with feasible variance (5.3) for the integrated intensity with Models 1-3. The order of the observation number n is 150,000, and the number of replications is 1,000.

Table 3

Summary statistics for the integrated baseline with Models 4-9. The order of the observation number is n=150,000 or 1,000,000, and the number of replications is 1,000.

n	150,000				1,000,000			
Variance	Unfeasible		Feasible		Unfeasible		Feasible	
Model	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
Model 4	0.0664	1.0085	0.1582	1.0142	-0.0281	1.0558	0.0579	1.0753
Model 5	0.0589	1.0440	0.1302	1.0869	0.0638	1.0108	0.1141	1.0195
Model 6	0.0401	1.0186	0.0874	1.0421	0.0183	1.0048	0.0400	1.0098
Model 7	0.1612	0.9826	0.2838	0.9927	0.0593	1.0046	0.1215	1.0089
Model 8	0.1189	1.0176	0.2644	1.0179	0.0657	1.0113	0.1173	1.0220
Model 9	0.1032	1.0563	0.2056	1.0866	0.0332	1.0394	0.8122	1.0723

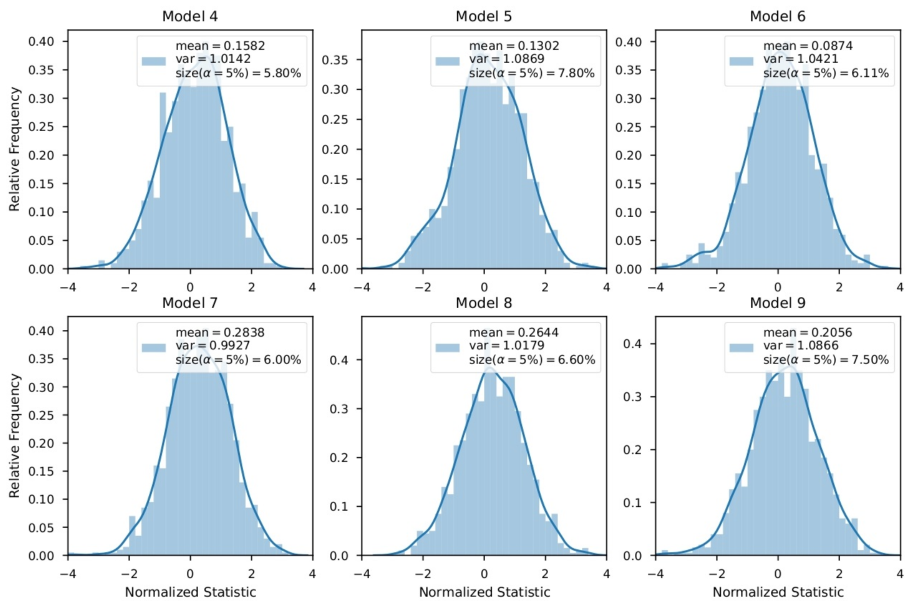

Fig 3: Histogram of the normalized CLT with feasible variance (5.8) for the integrated baseline with Models 4-9. The order of the observation number n is 150,000, and the number of replications is 1,000.

from 1% to 49%, with an average of 11%. It has an average of 7% for the statistics with unfeasible variance, and an average of 14% for the statistics with feasible variance. Overall, the statistics are biased, especially when the variance is unfeasible. However, the bias is smaller when n increases. The variance ranges from 98% to 120%, with an average of 110%. It has an average of 107% for the statistics with unfeasible variance, and an average of 114% for the statistics with feasible variance. Overall, the variance is reasonably close to unity.

Table 4
Summary statistics for the integrated volatility of the baseline with Models 1-9. The order of the observation number n is 150,000 and 1,000,000, and the number of replications is 1,000.

n	150,000				1,000,000			
Variance	Unfeasible		Feasible		Unfeasible		Feasible	
Model	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
Model 1	-0.0292	1.0393	-0.0620	1.0634	0.0488	1.0020	0.0985	1.0035
Model 2	0.0256	1.0447	0.0369	1.0617	-0.0276	1.0449	-0.0447	1.0563
Model 3	-0.1596	1.0428	-0.4900	1.0765	-0.1204	0.9850	-0.2589	0.9772
Model 4	-0.0258	1.0810	0.0004	1.1992	0.0213	1.0859	0.0302	1.1512
Model 5	0.0189	1.1989	-0.0276	1.4319	0.0344	1.0356	0.0507	1.0591
Model 6	-0.1904	1.1476	-0.3378	1.3072	-0.1422	1.1550	-0.2834	1.1390
Model 7	0.0587	1.1671	0.0807	1.2958	0.0326	1.0954	0.0540	1.1422
Model 8	0.0183	1.0567	0.0315	1.0900	0.0055	1.0328	0.0172	1.0542
Model 9	-0.1573	1.2021	-0.2987	1.3761	-0.0871	1.0992	-0.1460	1.2037

6.3. Hypothesis testing

Table 5 reports the percentage of rejections at the 5% level of the null hypothesis for the two tests with Models 1-9. The order of the observation number n is 50,000, 150,000 and 1,000,000. The size ranges from 4.2% to 6.4%, with an average of 5.4%. It has an average of 5.8% with the test for the absence of a Hawkes component, and an average of 5.0% with the test for baseline constancy. Overall, the test for the absence of a Hawkes component is slightly oversized while the size of the test for baseline constancy is adequate. The power is always equal to 100%, and thus is also adequate.

Table 6 reports the percentage of rejections at the 10% level of the null hypothesis for the two tests with Models 1-9. The order of the observation number n is 50,000, 150,000 and 1,000,000. The size ranges from 9.3% to 12.2%, with an average of 10.6%. It has an average of 11.0% with the test for the absence of a Hawkes component, and an average of 10.3% with the test for baseline constancy. Overall, the test for the absence of a Hawkes component is slightly oversized while the size of the test for baseline constancy is adequate.

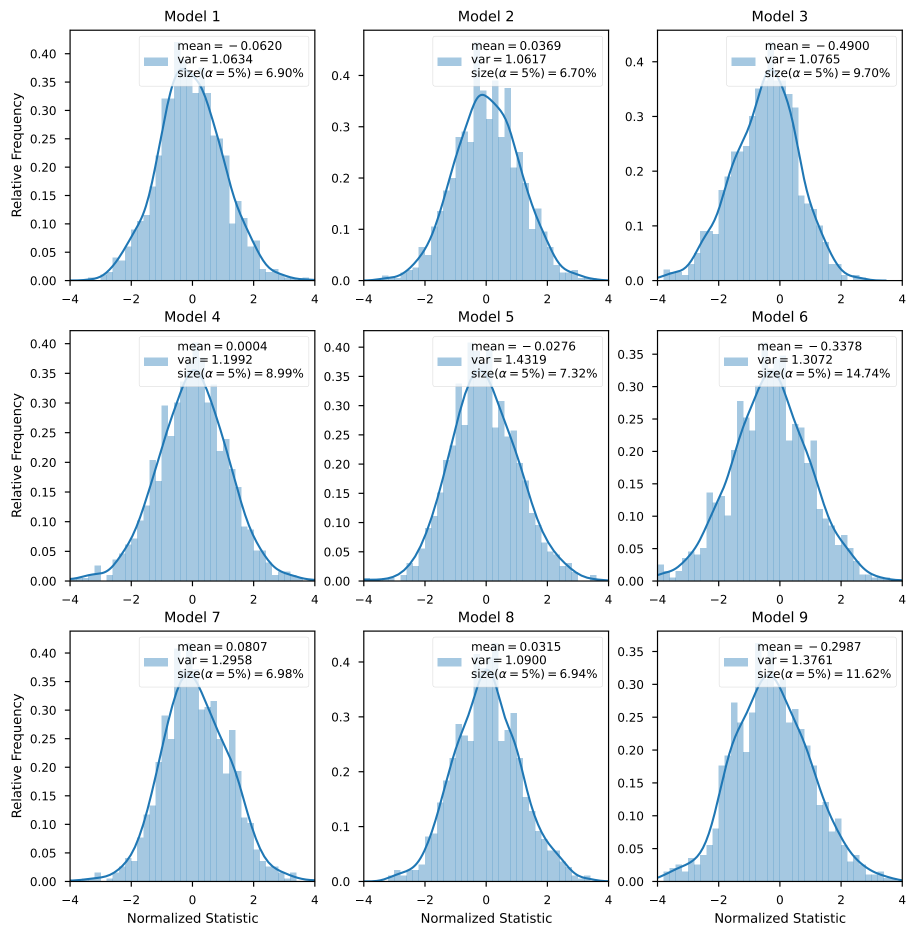

Fig 4: Histogram of the normalized CLT with feasible variance (5.10) for the integrated volatility of the baseline with Models 1-9. The order of the observation number n is 150,000, and the number of replications is 1,000.

Table 5
Percentage of rejections at the 5% level of the null hypothesis for the two tests with Models 1-9. The order of the observation number n is 50,000, 150,000 and 1,000,000, and the number of replications is 1,000.

	Test for the absence of a Hawkes component								
	Size			Power					
n	1	2	3	4	5	6	7	8	9
50,000	5.6	5.3	6.4	100	100	100	100	100	99.9
150,000	6.0	6.0	5.9	100	100	100	100	100	100
1,000,000	5.5	6.2	5.3	100	100	100	100	100	100
	Test for baseline constancy								
	Size			Power					
n	1	4	7	2	3	5	6	8	9
50,000	4.2	4.7	4.9	100	1.000	100	100	100	99.9
150,000	5.4	5.2	4.3	100	100	100	100	100	100
1,000,000	5.2	6.0	4.6	100	100	100	100	100	100

The power is always equal to 100%, and thus is also adequate.

Table 6
Percentage of rejections at the 10% level of the null hypothesis for the two tests with Models 1-9. The order of the observation number n is 50,000, 150,000 and 1,000,000, and the number of replications is 1,000.

	Test for the absence of a Hawkes component								
	Size			Power					
n	1	2	3	4	5	6	7	8	9
50,000	10.3	11.5	11.1	100	100	100	100	100	99.9
150,000	12.2	11.6	11.3	100	100	100	100	100	100
1,000,000	10.2	10.9	9.7	100	100	100	100	100	100
	Test for baseline constancy								
	Size			Power					
n	1	4	7	2	3	5	6	8	9
50,000	9.3	9.9	11.5	100	100	100	99.9	100	99.9
150,000	10.1	10.0	9.5	100	100	100	100	1.000	100
1,000,000	10.0	11.7	10.5	100	100	100	100	100	100

7. Empirical application

Our empirical application focuses on the S&P500 E-mini futures. They are liquid contracts traded on the Chicago Mercantile Exchange. We obtain the mid-quote price, i.e., the average price between best bid and ask prices, and time stamps from the consolidated trade history in the transaction Tickdatamarket database.

The data set covers the period from January 2020 to December 2021. All index quotes are considered during normal trading hours.

In Figure 5, we plot the estimated intensity for the whole sample by averaging the intraday estimates $\widehat{\lambda}_i$ across days, each day with a normalized time of [0,1]. The intensity reported shows the U-shaped pattern captured by μ_t^C and the intensity bursts captured by μ_t^B in our simulation design based on (6.1). The most pronounced bursts occur at the beginning of the trading session and just before closing.

Now, we turn to testing the hypotheses formulated in Sections 5.4 and 5.5, namely absence of a Hawkes component and baseline constancy. For each day in the sample, we perform the tests following Corollaries 5.4 and 5.5. Figure 6 shows corresponding test statistics revealing rejection of the null hypothesis in both cases. Namely, we confirm the presence of Hawkes component (blue line) and the varying baseline (orange line) for all days.

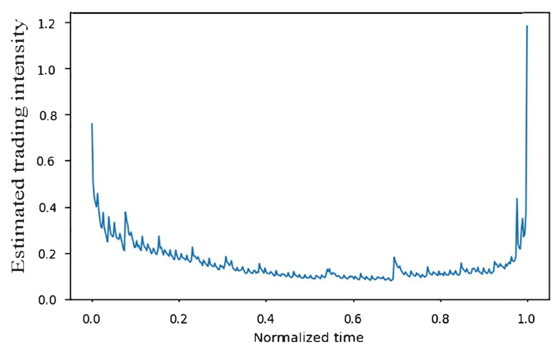

Fig 5: Estimated intensity of the S&P500 E-mini futures quotes. The number of the mid-quote price changes in millions per minute is shown.

To verify that our test results are not distorted due to a multiple statistical inference problem, we implement the sequential Bonferroni procedure of Holm (1979) for all p-values. The adjusted p-values computed at the 1% level provide identical conclusions about all hypotheses, confirming the statistical robustness of our results. Another robustness check of our test results is conducted following Bajgrowicz, Scaillet and Treccani (2016) and the results are in agreement with the Bonferroni corrected tests.

In summary, the empirical findings are in favor of Hawkes processes with time-varying baseline.

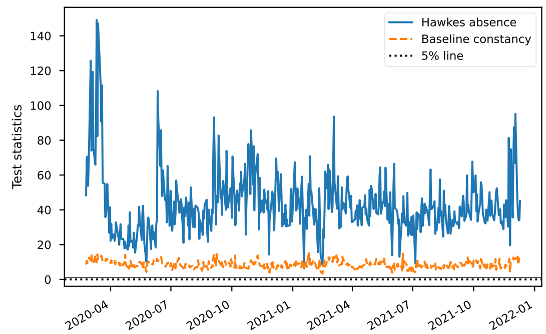

Fig 6: Test statistics for the null hypothesis in the two tests from Sections 5.4 and 5.5 with the 5% critical value.

8. Conclusion

In this paper, we have considered Hawkes processes with Itô semimartingale baseline. This time-varying baseline can accommodate for time-variation, stochasticity, and bursts. We have derived CLT for empirical average and variance of local Poisson estimates. For the applications, we have studied the integrated intensity, the integrated baseline, and the integrated volatility of the baseline. We have also developed a test for the absence of a Hawkes component and a test for baseline constancy. The simulation study corroborates the asymptotic theory. The empirical application shows that the absence of Hawkes component and baseline constancy is always rejected.

The code is available at `https://github.com/SeunghyeonTonyYu/TSRV2Hawkes`.

Acknowledgments

The authors would like to thank participants of Statistical Methods in Finance 2022, Le Mans University seminar, Capital Fund Management, 33rd $(EC)^2$ Conference and the session entitled "Hawkes processes in finance" in CFE-CMStatistics 2022 for their constructive comments that improved the quality of this paper.

Funding

The first author was supported in part by Japanese Society for the Promotion of Science Grants-in-Aid for Scientific Research (B) 23H00807 and Early-Career Scientists 20K13470.

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. (2014). *High-frequency financial econometrics*. Princeton University Press.

Aït-Sahalia, Y., Laeven, R. and Pelizzon, L. (2014). Mutual excitation in eurozone sovereign CDS. *Journal of Econometrics* **183** 151-167.

Aït-Sahalia, Y. and Laeven, R. J. (2023). Saddlepoint Approximations for Hawkes Jump-Diffusion Processes with an Application to Risk Management.

Aït-Sahalia, Y. and Xiu, D. (2019). A Hausman test for the presence of market microstructure noise in high frequency data. *Journal of Econometrics* **211** 176–205.

Bacry, E., Delattre, S., Hoffmann, M. and Muzy, J.-F. (2013a). Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes and their Applications* **123** 2475–2499.

Bacry, E., Delattre, S., Hoffmann, M. and Muzy, J.-F. (2013b). Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance* **13** 65–77.

Bacry, E., Bompaire, M., Deegan, P., Gaïffas, S. and Poulsen, S. V. (2017). tick: A Python library for statistical learning, with an emphasis on Hawkes processes and time-dependent models. *The Journal of Machine Learning Research* **18** 7937–7941.

Bajgrowicz, P., Scaillet, O. and Treccani, A. (2016). Jumps in high-frequency data: Spurious detections, dynamics, and news. *Management Science* **62** 2198–2217.

Barndorff-Nielsen, O. E., Graversen, S. E., Jacod, J., Podolskij, M. and Shephard, N. (2006). A central limit theorem for realised power and bipower variations of continuous semimartingales. In *From stochastic calculus to mathematical finance* 33–68. Springer.

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

Brémaud, P. and Massoulié, L. (1996). Stability of nonlinear Hawkes processes. *Annals of Probability* 1563–1588.

Brémaud, P. and Massoulié, L. (2001). Hawkes branching point processes without ancestors. *Journal of Applied Probability* **38** 122–135.

Cavaliere, G., Lu, Y., Rahbek, A. and Stærk-Østergaard, J. (2023). Bootstrap inference for Hawkes and general point processes. *Journal of Econometrics* **235** 133–165.

Chavez-Demoulin, V., Davison, A. C. and McNeil, A. J. (2005). Estimating value-at-risk: a point process approach. *Quantitative Finance* **5** 227–234.

Chen, F. and Hall, P. (2013). Inference for a nonstationary self-exciting point process with an application in ultra-high frequency financial data modeling. *Journal of Applied Probability* **50** 1006–1024.

Cheysson, F. and Lang, G. (2022). Spectral estimation of Hawkes processes from count data. *The Annals of Statistics* **50** 1722–1746.

Christensen, K. and Kolokolov, A. (2024). An unbounded intensity model for point processes. *Journal of Econometrics* **244** 105840.

Clinet, S. and Potiron, Y. (2018). Statistical inference for the doubly stochastic self-exciting process. *Bernoulli* **24** 3469–3493.

Clinet, S. and Potiron, Y. (2019). Testing if the market microstructure noise is fully explained by the informational content of some variables from the limit order book. *Journal of Econometrics* **209** 289–337.

Corradi, V., Distaso, W. and Fernandes, M. (2020). Testing for jump spillovers without testing for jumps. *Journal of the American Statistical Association* **115** 1214–1226.

Dachian, S. and Kutoyants, Y. A. (2006). Hypotheses Testing: Poisson Versus Self-exciting. *Scandinavian Journal of Statistics* **33** 391–408.

Daley, D. J. and Vere-Jones, D. (2003). *An introduction to the theory of point processes: Elementary theory and methods* **1**. Springer Verlag.

Daley, D. J. and Vere-Jones, D. (2008). *An introduction to the theory of point processes: General theory and structure* **2**. Springer Verlag.

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

Feller, W. (1951). Two singular diffusion problems. *Annals of Mathematics* 173–182.

Filimonov, V. and Sornette, D. (2012). Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review E* **85** 056108.

Fulop, A., Li, J. and Yu, J. (2015). Self-exciting jumps, learning, and asset pricing implications. *The Review of Financial Studies* **28** 876–912.

Girsanov, I. V. (1960). On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. *Theory of Probability & Its Applications* **5** 285–301.

Hardiman, S. J. and Bouchaud, J.-P. (2014). Branching-ratio approximation for the self-exciting Hawkes process. *Physical Review E* **90** 062807.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica* 1251–1271.

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

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

Hawkes, A. G. (2018). Hawkes processes and their applications to finance: a review. *Quantitative Finance* **18** 193–198.

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

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6** 65-70.

Ikefuji, M., Laeven, R., Magnus, J. and Yue, Y. (2022). Earthquake risk embedded in property prices: Evidence from five Japanese cities. *Journal of the American Statistical Association* **117** 82–93.

Jacod, J. (2008). Asymptotic properties of realized power variations and related functionals of semimartingales. *Stochastic processes and their applications* **118** 517–559.

Jacod, J. and Protter, P. (2011). *Discretization of processes* **67**. Springer Science & Business Media.

Jacod, J. and Shiryaev, A. (2013). *Limit theorems for stochastic processes* **288**. Springer Science & Business Media.

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

Kao, E. P. and Chang, S.-L. (1988). Modeling time-dependent arrivals to service systems: A case in using a piecewise-polynomial rate function in a nonhomogeneous Poisson process. *Management Science* **34** 1367–1379.

Kimura, A. and Yoshida, N. (2016). Estimation of Correlation Between Latent Processes. In *Advanced Modelling in Mathematical Finance* 131–146. Springer.

Kuhl, M. E. and Bhairgond, P. S. (2000). Nonparametric estimation of nonhomogeneous Poisson processes using wavelets. In *2000 Winter Simulation Conference Proceedings (Cat. No. 00CH37165)* **1** 562–571. IEEE.

Kuhl, M. E., Wilson, J. R. and Johnson, M. A. (1997). Estimating and simulating Poisson processes having trends or multiple periodicities. *IIE trans-*

actions **29** 201–211.

Kuhl, M. E. and Wilson, J. R. (2000). Least squares estimation of nonhomogeneous Poisson processes. *Journal of Statistical Computation and Simulation* **67** 699–712.

Kuhl, M. E. and Wilson, J. R. (2001). Modeling and simulating Poisson processes having trends or nontrigonometric cyclic effects. *European Journal of Operational Research* **133** 566–582.

Kwan, T.-K. J., Chen, F. and Dunsmuir, W. T. (2023). Alternative asymptotic inference theory for a nonstationary Hawkes process. *Journal of Statistical Planning and Inference* **227** 75–90.

Large, J. (2007). Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets* **10** 1-25.

Lee, S., Wilson, J. R. and Crawford, M. M. (1991). Modeling and simulation of a nonhomogeneous Poisson process having cyclic behavior. *Communications in Statistics-Simulation and Computation* **20** 777–809.

Leemis, L. M. (1991). Nonparametric estimation of the cumulative intensity function for a nonhomogeneous Poisson process. *Management Science* **37** 886–900.

Lewis, P. A. and Shedler, G. S. (1976). Statistical analysis of non-stationary series of events in a data base system. *IBM Journal of Research and Development* **20** 465–482.

Liniger, T. (2009). Multivariate Hawkes processes, PhD thesis, ETH Zurich.

Mammen, E. and Müller, M. (2023). Nonparametric estimation of locally stationary Hawkes processes. *Bernoulli* **29** 2062–2083.

Novikov, A. A. (1972). On an identity for stochastic integrals. *Teoriya Veroyatnostei i ee Primeneniya* **17** 761–765.

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

Omi, T., Hirata, Y. and Aihara, K. (2017). Hawkes process model with a time-dependent background rate and its application to high-frequency financial data. *Physical Review E* **96** 012303.

Ozaki, T. (1979). Maximum likelihood estimation of Hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics* **31** 145–155.

Potiron, Y. and Volkov, V. (2025). Mutually exciting point processes with latency. *To appear in Journal of the American Statistical Association.*

Rambaldi, M., Filimonov, V. and Lillo, F. (2018). Detection of intensity bursts using Hawkes processes: An application to high-frequency financial data. *Physical Review E* **97** 032318.

Rambaldi, M., Pennesi, P. and Lillo, F. (2015). Modeling foreign exchange market activity around macroeconomic news: Hawkes-process approach. *Physical Review E* **91** 012819.

Roueff, F., von Sachs, R. and Sansonnet, L. (2016). Locally stationary Hawkes processes. *Stochastic Processes and their Applications* **126** 1710–1743.

Roueff, F. and Von Sachs, R. (2019). Time-frequency analysis of locally stationary Hawkes processes. *Bernoulli* **25** 1355–1385.

Rubin, I. (1972). Regular point processes and their detection. *IEEE Transactions on Information Theory* **18** 547–557.

Stoltenberg, E. A., Mykland, P. A. and Zhang, L. (2022). A CLT for second difference estimators with an application to volatility and intensity. *The Annals of Statistics* **50** 2072–2095.

Todorov, V. and Tauchen, G. (2011). Limit theorems for power variations of pure-jump processes with application to activity estimation. *Annals of Applied Probability* **21** 546–588.

Türkmen, A. C. and Cemgil, A. T. (2018). Testing for Self-excitation in Financial Events: A Bayesian Approach. In *ECML PKDD 2018 Workshops* 95–102. Springer.

Vere-Jones, D. (1978). Earthquake prediction-a statistician's view. *Journal of Physics of the Earth* **26** 129–146.

Vere-Jones, D. and Ozaki, T. (1982). Some examples of statistical estimation applied to earthquake data: I. Cyclic Poisson and self-exciting models. *Annals of the Institute of Statistical Mathematics* **34** 189–207.

Yu, J. (2004). Empirical characteristic function estimation and its applications. *Econometric Reviews* **23** 93–123.

Appendix A: Proofs

This appendix provides the detailed proofs for the theoretical results, namely Proposition 4.1, Theorem 4.1, Proposition 5.1 and the five corollaries of Section 5. They rely on theory developed in Bacry et al. (2013a), Barndorff-Nielsen et al. (2006), Brémaud and Massoulié (1996), Clinet and Potiron (2018) , Girsanov (1960), Jacod and Shiryaev (2013), Jacod and Protter (2011), Novikov (1972), and Todorov and Tauchen (2011).

A.1. Notations and definitions

To start with, we introduce some notations and definitions. In what follows, for any $i = 1, \dots, M$, we use $\mathbb{E}_{i-1}$, $\mathbb{V}_{i-1}$ and Cov_{i-1} in place of $\mathbb{E}[\cdot \,|\, \mathcal{F}_{(i-1)\Delta_n}]$, $\mathbb{V}[\cdot \,|\, \mathcal{F}_{(i-1)\Delta_n}]$ and $\mathrm{Cov}[\cdot \,|\, \mathcal{F}_{(i-1)\Delta_n}]$. We also make use of X_i instead of $X_{i\Delta_n}$ and ΔX_i instead of $X_i - X_{i-1}$. We define $\overline{X}_i$ as the average of X_t on the i-th block, i.e., $\overline{X}_i = \Delta_n^{-1} \int_{(i-1)\Delta_n}^{i\Delta_n} X_t \, dt$. We denote the L^p-norm of X as $\|X\|_p$. We denote the big O in probability and the big O in L^p-norm by $O_{\mathbb{P}}$ and O_{L^p}. They are defined through $X_n = O_{\mathbb{P}}(\alpha_n) \iff \frac{X_n}{\alpha_n}$ is stochastically bounded and $X_n = O_{L^p}(\alpha_n) \iff \|X_n\|_p = O(\alpha_n)$. We denote a uniform big O by $\underline{O}$. It is defined through $f(n,t) = \underline{O}\big(g(n,t)\big) \iff |f(n,t)| \leq Cg(n,t)$, for all $n \in \mathbb{N}$, $t \in [0,T]$, and some $C \in \mathbb{R}_+$ which does not depend on n and t. If g has no dependency on t, then $f(n,t) = \underline{O}\big(g(n)\big)$ implies uniformly boundedness by $g(n)$, i.e., $\sup_{t\in[0,T]} |f(n,t)| \leq Cg(n)$ for all $n \in \mathbb{N}$, so we can most of the time freely exchange the order of the limit and the integral/sum. We denote the strict big O in L^k-norm by $\underline{O}_{L^k}$. It is defined through $f(n,i) = \underline{O}_{L^k}\big(g(n,i)\big) \iff \mathbb{E}\big[\big(f(n,i)\big)^k\big]^{\frac{1}{k}} = \underline{O}\big(g(n,i)\big)$. No dependency of g on i means that $f(n,i)$ can be bounded uniformly in i with g. In the proofs, f, g, and h are temporary functions which may vary, and C denotes a generic constant that does not depend on n and may differ.

Let us begin with the proof of the existence of Hawkes processes with a time-varying baseline driven by an Itô semimartingale. It extends the proof of Theorem 4 (pp. 1574-1575) in Brémaud and Massoulié (1996) to the time-varying baseline case and the proof of Theorem 5.1 (pp. 3-4) in the supplement of Clinet and Potiron (2018) in which the kernel is exponential to the general kernel case.

A.2. Proof of Proposition 4.1

The strategy of the proof consists in defining a suitable sequence of simple point processes and intensity $(N_t^k, \lambda_t^k)_{k\geq 0}$ such that their limit defined as $(N_t, \lambda_t) = \lim_{k\to\infty}(N_t^k, \lambda_t^k)$ exists and N_t admits λ_t as $\mathcal{F}_t$-intensity given by Eq. (2.1). We first define for $t \in [0,T]$ $\lambda^0(t) = \mu_t$ and N_t^0 the simple point process counting the points of $\underline{N}$ below the curve $t \to \lambda_t^0$ as $N_t^0 = \int_{[0,t]\times\mathbb{R}} \mathbb{1}_{[0,\lambda_s^0]}(x)\underline{N}(ds \times dx)$. We then define recursively the sequence of $(N_t^k, \lambda_t^k)_{k\geq 1}$ as

$$\begin{aligned} \lambda_t^{k+1} &= \mu_t + \int_0^{t-} \phi(t-s) dN_s^k \\ N_t^{k+1} &= \int_{[0,t]\times\mathbb{R}} \mathbb{1}_{[0,\lambda_s^{k+1}]}(x)\underline{N}(ds \times dx). \end{aligned} \tag{A.1}$$

First, we have that λ_t^k is a.s. positive as an application of Assumption 1(a) so that λ_t^k is a well-defined intensity. Then, an extension to the time-varying case of the arguments from Lemma 3 and Example 4 (pp. 1571-1572) in Brémaud and Massoulié (1996) yields that N_t^k is $\mathcal{F}_t$-adapted, λ_t^k is $\mathcal{F}_t$-predictable and an $\mathcal{F}_t$-intensity of N_t^k. Moreover, nonnegative ϕ implies that (N_t^k, λ_t^k) is componentwise increasing with k and thus converges to some limit (N_t, λ_t) a.s. for any $t \in [0,T]$. We now introduce the sequence of processes ρ_t^k defined as $\rho_t^k = \mathbb{E}[\lambda_t^k - \lambda_t^{k-1} | \widetilde{\mathcal{F}}_T]$. Then $\rho_t^{k+1} = \mathbb{E}\Big[\int_0^t \phi(t-s)(\lambda_s^k - \lambda_s^{k-1})ds\Big|\widetilde{\mathcal{F}}_T\Big] = \int_0^t \phi(t-s)\rho_s^k ds$, where the first equality is obtained by Lemma 10.1 (p. 2) in the supplement of Clinet and Potiron (2018) when $\mathcal{G} = \widetilde{\mathcal{F}}_T$ along with Eq. (A.1), and the second equality by Tonelli's theorem and the definition of ρ_t^k. If we define Φ_t^k as $\Phi_t^k = \int_0^t \rho_s^k ds$ a.s., we have by another application of Tonelli's theorem that

$$\Phi_t^{k+1} = \int_0^t \Big(\int_0^{t-s} \phi(u)du\Big)\rho_s^n ds. \tag{A.2}$$

By definition of the L^1 norm, we deduce that $\int_0^{t-s} \phi(u)du \leq \|\phi\|_1$. Thus, an application of the definition of Φ_t^k along with Eq. (A.2) implies that $\Phi_t^{k+1} \leq \|\phi\|_1\Phi_t^k$. Then, since Assumption 1(d) states that $\|\phi\|_1 < 1$, we can deduce that $F : \Phi_t^k \to \Phi_t^{k+1}$ is a.s. a contraction function. It turns out that the limit of the telescopic series $(\sum_{l=0}^k \Phi_t^l)_{k\geq 1}$ exists by arguments used in Banach fixed-point theorem. Working with the telescopic series and applying the monotone convergence theorem to the series yields

$$\mathbb{E}\Big[\int_0^t \lambda_s ds\Big|\widetilde{\mathcal{F}}_T\Big] \leq \int_0^t \mu_s ds + \|\phi\|_1\mathbb{E}\Big[\int_0^t \lambda_s ds\Big|\widetilde{\mathcal{F}}_T\Big]. \tag{A.3}$$

By rearranging the terms in Eq. (A.3), we get that

$$\mathbb{E}\Big[\int_0^t \lambda_s ds\Big|\widetilde{\mathcal{F}}_T\Big] \leq (1 - \|\phi\|_1)^{-1}\int_0^t \mu_s ds. \tag{A.4}$$

Given Assumption 1(b), the expression in the left side of Eq. (A.4) is finite a.s.. Given that its conditional expectation is finite, $\int_0^t \lambda_s ds$ is finite a.s.. Moreover, λ_t is $\mathcal{F}_t$-predictable as a limit of such processes. N_t counts the points of $\underline{N}$ under the curve $t \mapsto \lambda_t$ by an application of the monotone convergence theorem. N_t therefore admits λ_t as an $\mathcal{F}_t$-intensity by an extension to the time-varying case of the arguments from Lemma 3 (p. 1571) in Brémaud and Massoulié (1996). It implies that N_t is finite a.s.. Finally, it remains to show that λ_t is of the form Eq. (2.1). The monotonicity properties of N_t^k and λ_t^k ensure that, for all $k \geq 0$ and all $t \in [0,T]$, $\lambda_t^k \leq \mu_t + \int_0^t \phi(t-s)\, dN_s$ and $\lambda_t \geq \mu_t + \int_0^t \phi(t-s)\, dN_s^k$, which gives Eq. (2.1) by taking the limit $k \to +\infty$ in both inequalities.

A.3. Preliminary results

Let us define $\phi^n(t)$ as $\phi^n(t) = n\phi(nt)$ and the Laplace transform of the kernel as $\widehat{\phi}(s) = \int_0^\infty e^{-st}\phi(t)\, dt$. For f and g two integrable functions, we define the convolution of f and g as $f * g_t = \int_{-\infty}^\infty f(t-s)g(s)\, ds$. For an integrable function f and a stochastic process X, we define the convolution of f and X as $f * dX_t \coloneqq \int_{-\infty}^\infty f(t-s)\, dX_s$. Let $\psi\colon \mathbb{R}_+ \to \mathbb{R}_+$ be the resolvent kernel of ϕ which satisfies $\phi + \phi * \psi = \psi$, i.e., $\psi(t) = \phi(t) + \phi * \psi(t)$. Similarly, let ψ^n be the resolvent kernel of ϕ^n, i.e., $\phi^n + \phi^n * \psi^n = \psi^n$. Finally, we define the integral of ψ^n as $\Psi^n(t) = \int_0^t \psi^n(s)\, ds$ and the integral between $(i-1)\Delta_n$ and $i\Delta_n$ as $\Delta_i \Psi^n(-t) = \int_{(i-1)\Delta_n}^{i\Delta_n} \psi^n(s-t)\, ds$.

The first lemma gives the asymptotic properties of the resolvent kernel, which can be expressed as a Laplace transform of the kernel.

Lemma A.1. *Under Assumptions 1 and 2(a), we have*

$$\psi(t) = \begin{cases} \geq 0 & \text{for any } t \geq 0 \\ 0 & \text{for any } t < 0 \end{cases}, \tag{A.5}$$

$$\psi^n(t) = n\psi(nt), \tag{A.6}$$

$$\Psi^n(t) = \widehat{\psi}(0) + \underline{O}\left(1 \wedge \frac{1}{nt}\right), \tag{A.7}$$

$$\Delta_i \Psi^n(-t) = \underline{O}\left(1 \wedge \frac{1}{n((i-1)\Delta_n - t)}\right), \tag{A.8}$$

Proof of Lemma A.1. Since $\|\phi\|_1 < 1$ by Assumption 1(d), the transform $T\colon f \mapsto (\mu + \phi * f)$ is a contraction map. Thus, we can apply Banach fixed-point theorem to get a fixed-point $\psi = f_\infty$ with recursion $f_n = T(f_{n-1})$ and we obtain

$$f_n = T(f_{n-1}) = \mu + \phi * f_{n-1} = \mu + \phi * (Tf_{n-2}) = \mu + \phi * (\mu + \phi * f_{n-2})$$

$$= \dots = \mu + \phi * \mu + \phi^{*2} * \mu + \cdots + \phi^{*(n-2)} * \mu + \phi^{*(n-1)} * f_1.$$

For the initial value, we can choose $f_1 = 0$. Then, for all $n > 1$, $f_n(t)$ is nonnegative and $f_n(t) = 0$, for $t < 0$, since the kernel ϕ is nonnegative and equal to zero for $t < 0$. So we have Eq. (A.5). By the scaling property of the Laplace transform, we have that $\widehat{\phi^n}(s) = \widehat{\phi}(s/n)$, and hence Eq. (A.6). For Eq. (A.7), since ψ and ϕ are nonnegative, we have

$$\begin{aligned}\widehat{\psi}(0) - \widehat{\psi}\Big(\frac{1}{nt}\Big) &= \int_0^\infty \big(1 - e^{-\frac{s}{nt}}\big)\psi(s)\,ds \geq \int_{nt}^\infty \big(1 - e^{-\frac{s}{nt}}\big)\psi(s)\,ds \\ &\geq (1 - e^{-1}) \int_{nt}^\infty \psi(s)\,ds. \end{aligned} \tag{A.9}$$

From Assumption 2(a), we have that $\widehat{\phi}'(0) = \int_0^\infty t\phi(t)dt < \infty$ and then we can apply Taylor's theorem for $\widehat{\psi}(s) = \widehat{\phi}(s)/(1 - \widehat{\phi}(s))$ with the Peano's form of the remainder $\widehat{\psi}(s) = \widehat{\psi}(0) + \widehat{\psi}'(0)s + h(s)s$, where $\lim_{s\to 0} h(s) = 0$. The function $\widehat{\psi}$ is decreasing such that $\widehat{\psi}(s) \geq \widehat{\psi}(t)$ for $s < t$, and $\widehat{\psi}(0) < \infty$, so we have

$$\begin{aligned} 0 \leq \widehat{\psi}(0) - \widehat{\psi}\Big(\frac{1}{nt}\Big) &= -\Big(\widehat{\psi}'(0) + h\Big(\frac{1}{nt}\Big)\Big)\frac{1}{nt} \\ &\leq \begin{cases} \big(|\widehat{\psi}'(0)| + \big|h\big(\frac{1}{nt}\big)\big|\big)\frac{1}{nt} & \text{if } nt \geq 1, \\ \widehat{\psi}(0) & \text{if } nt < 1. \end{cases} \end{aligned}$$

Since $\sup_{x\in[0,1]} |h(x)| < \infty$, we obtain

$$\Big|\widehat{\psi}(0) - \widehat{\psi}\Big(\frac{1}{nt}\Big)\Big| \leq C\Big(1 \wedge \frac{1}{nt}\Big). \tag{A.10}$$

It implies that

$$\begin{aligned} \Psi^n(t) &= \int_0^t \psi^n(s)\,ds = \int_0^t n\psi(ns)\,ds = \int_0^{nt} \psi(s)\,ds \\ &= \widehat{\psi}(0) - \int_{nt}^\infty \psi(s)\,ds \leq \widehat{\psi}(0) + \frac{1}{1 - e^{-1}}\Big(\widehat{\psi}(0) - \widehat{\psi}\Big(\frac{1}{nt}\Big)\Big) \\ &= \widehat{\psi}(0) + C\Big(1 \wedge \frac{1}{nt}\Big), \end{aligned}$$

for sufficiently large n, where we have used successively Eq. (A.6), the change of variable $ns \to s$, Eq. (A.9), and Eq. (A.10). With the same arguments, we can show the result for Eq. (A.8). □

For any $t \in [0, T]$, we define the sum of the baseline and the convolution of the resolvent kernel and baseline as $\nu_t^n = \mu_t + \psi^n * \mu_t$ and $\mathcal{F}_t$-martingale of compensated N_t as $M_t = N_t - \int_0^t \lambda_u du$. We also define the limit of ν_t^n as $\nu_t = (1 + \widehat{\psi}(0))\mu_t$.

The following lemma exhibits an $\mathcal{F}_t$-martingale representation of the $\mathcal{F}_t$-intensity λ_t. It is based on the convolution of the resolvent kernel and the martingale. It extends Lemma 3 in Bacry et al. (2013a), which considers an invariant baseline and large T asymptotics, to the time-varying baseline and in-fill asymptotics case.

Lemma A.2. *Under Assumptions 1 and 2(a), we have for any* $t \in [0, T]$

$$\lambda_t = n\nu_t^n + \psi^n * dM_t. \tag{A.11}$$

Moreover, we have

$$\nu_t^n - \nu_t = \underline{O}_{L^k}(1/\sqrt{n}). \tag{A.12}$$

Proof of Lemma A.2. By Lemma 3 in Bacry et al. (2013a), the solution of the equation $f(t) = h(t) + \phi^n * f(t)$ with measurable locally bounded function h_t is

$$f(t) = h(t) + \psi^n * h(t).$$

In our case, we have to solve λ_t which satisfies

$$\lambda_t = n\mu_t + \phi^n * dN_t = n\mu_t + \phi^n * (\lambda_t + dM_t) = \big(n\mu_t + \phi^n * dM_t\big) + \phi^n * \lambda_t.$$

Applying Lemma 3 in Bacry et al. (2013a) with the function $h(t)$ defined as $h(t) = n\mu_t + \phi^n * dM_t$, we have

$$\begin{aligned} \lambda_t &= h(t) + \phi^n * \lambda_t \\ &= h(t) + \psi^n * h(t) \\ &= n\mu_t + \phi^n * dM_t + \psi^n * (n\mu_t + \phi^n * dM_t) \\ &= n(\mu_t + \psi^n * \mu_t) + (\phi^n + \psi^n * \phi^n) * dM_t \\ &= n(\mu_t + \psi^n * \mu_t) + \psi^n * dM_t. \end{aligned}$$

Thus, we can obtain Eq. (A.11). We show now Eq. (A.12). Since $\psi^n * \mu_t = \int_{-\infty}^t \psi^n(t-s)\mu_s\, ds = \int_0^\infty \psi^n(s)\mu_{t-s}\, ds = \int_0^\infty \psi(s)\mu_{t-\frac{s}{n}}\, ds$ by Eq. (A.6) from Lemma A.1 along with Assumptions 1 and 2(a), we obtain that

$$\begin{aligned} \psi^n * \mu_t - \widehat{\psi}(0)\mu_t &= \int_0^\infty \psi(s)\mu_{t-\frac{s}{n}} ds - \widehat{\psi}(0)\mu_t \\ &= \int_0^\infty \psi(s)\big(\mu_{t-\frac{s}{n}} - \mu_t\big) ds. \end{aligned} \tag{A.13}$$

Then the local boundedness of μ_t gives

$$\left|\psi(s)\big(\mu_{t-\frac{s}{n}} - \mu_t\big)\right| \leq 2\psi(s) \sup_{0\leq s\leq t} \mu_s < C\psi(s),$$

so we can apply the dominated convergence theorem which yields

$$\lim_{n\to\infty} \int_0^\infty \psi(s)\big(\mu_{t-\frac{s}{n}} - \mu_t\big)ds = \int_0^\infty \psi(s) \lim_{n\to\infty} \big(\mu_{t-\frac{s}{n}} - \mu_t\big)ds,$$

for all $t \in [0,T]$. An extension of these arguments yields Eq. (A.12). □

The following lemma extends Lemma 10.3 from Clinet and Potiron (2018) (pp. 4-6 in the supplement) in which the kernel is exponential.

Lemma A.3. *Under Assumption 1, for any $k > 0$, there exists C such that $\sup_t \mathbb{E}\lambda_t^k \leq Cn^k$, for any $t \in [0,T]$.*

Proof of Lemma A.3. We can use the same arguments as in the proof of Lemma 10.3 in Clinet and Potiron (2018) along with Assumption 1(d). □

For any $i = 1,\ldots,M$, we define the estimator of rescaled spot intensity as $\widehat{\bar{\nu}}_i^n = \frac{\widehat{\lambda}_i}{n}$. We also define the rescaled increment of the martingale as $\varepsilon_i = \frac{1}{n\Delta_n}\int_{(i-1)\Delta_n}^{i\Delta_n} dM_t$ and ϵ_i as $\epsilon_i = \frac{1}{n\Delta_n}\big\{\int_0^{(i-1)\Delta_n} \Delta_i\Psi^n(-t)\, dM_t + \int_{(i-1)\Delta_n}^{i\Delta_n} \Psi^n(i\Delta_n - t)\, dM_t\big\}$. Finally, we define the sum of ε_i and ϵ_i as $u_i = \varepsilon_i + \epsilon_i$.

The following lemma is a decomposition of the estimation error u_i as the sum of the error originating from the time-varying baseline ε_i and another related to the Hawkes structure ϵ_i.

Lemma A.4. *Under Assumptions 1 and 2(a), we have for any $i = 1,\ldots,M$, that*

$$\widehat{\bar{\nu}}_i^n = \bar{\nu}_i^n + u_i. \tag{A.14}$$

Proof of Lemma A.4. It is obtained by Lemma A.2 along with Fubini's theorem, Assumptions 1 and 2(a). □

For any $t \in [0,T]$, we define ϑ_t^n as $\vartheta_t^n = (1+\widehat{\psi}(0))^2\nu_t^n$. The following lemma provides moments of u_i.

Lemma A.5. *Under Assumptions 1 and 2(a), for any $k \in \mathbb{N}_*$, we have*

$$\mathbb{E}[|u_i|^k] \leq \frac{C}{(n\Delta_n)^{k/2}},$$

$$\mathbb{E}_{i-1}[u_i \mid \mathcal{F}^\mu] = \underline{O}_{L^k}\left(\frac{\log n}{n\Delta_n}\right),$$

$$\mathbb{E}_{i-1}[u_i u_j \,|\, \mathcal{F}^\mu] = \underline{O}_{L^k}\left(\frac{\log n}{(n\Delta_n)^2}\right), \qquad \textit{for any } i < j,$$
$$\mathbb{E}_{i-1}[u_i^2 \,|\, \mathcal{F}^\mu] = \frac{1}{n\Delta_n}\bar{\vartheta}_i^n + \underline{O}_{L^k}\left(\frac{\log n}{(n\Delta_n)^2}\right),$$
$$\mathbb{E}_{i-1}[u_i^3 \,|\, \mathcal{F}^\mu] = \frac{(1+3\widehat{\psi}(0))(1+\widehat{\psi}(0))}{(n\Delta_n)^2}\bar{\vartheta}_i^n + \underline{O}_{L^k}\left(\frac{(\log n)^2}{(n\Delta_n)^3}\right),$$
$$\mathbb{E}_{i-1}[u_i^4 \,|\, \mathcal{F}^\mu] = \frac{3}{(n\Delta_n)^2}(\bar{\vartheta}_i^n)^2 + \underline{O}_{L^k}\left(\frac{1}{(n\Delta_n)^3} + \frac{(\log n)^3}{(n\Delta_n)^4}\right). \tag{A.15}$$

Proof of Lemma A.5. Without loss of generality and for convenience of notation, we assume that μ_t and thus ν_t are nonrandom throughout this proof. We first calculate the moments of u_i. For the first moment, it is sufficient to consider ϵ_i because $\mathbb{E}_{i-1}[\varepsilon_i] = 0$. We have that $\mathbb{E}_{i-1}[\epsilon_i] = \underline{O}_{L^k}\big(\frac{\log n}{n\Delta_n}\big)$ holds for $k=1$ and $k=2$ since $\mathbb{E}_{i-1}[\epsilon_i] = (n\Delta_n)^{-1}\int_0^{(i-1)\Delta_n}\Delta_i\Psi^n(-t)\,dM_t$ by Lemma A.4, Itô isometry for point processes, Lemma A.3 along with Assumptions 1 and 2(a). We thus obtain that $\mathbb{E}\Big[\big(\mathbb{E}_{i-1}[\epsilon_i]\big)^2\Big] \le \frac{Cn}{(n\Delta_n)^2}\bigg(\int_0^{(i-1)\Delta_n - n^{-1}}\frac{1}{n((i-1)\Delta_n - t)}dt + \int_{(i-1)\Delta_n - n^{-1}}^{(i-1)\Delta_n}dt\bigg) \le \frac{Cn}{(n\Delta_n)^2}\bigg(\frac{\log n}{n} + \frac{1}{n}\bigg)$. For $k>2$, by Lemma 2.1.5 in Jacod and Protter (2011), Lemma A.3 along with Assumption 1, and Hölder's inequality, we have that $(n\Delta_n)^k\mathbb{E}\Big[\Big(\mathbb{E}_{i-1}[\epsilon_i]\Big)^k\Big] \le \underline{O}\big((\log n)^k\big)$. Similar arguments yield that $\mathbb{E}_{i-1}[u_i u_j] = \underline{O}_{L^k}\big(\frac{\log n}{(n\Delta_n)^2}\big)$, for $i<j$. To calculate the moments of ε_i, we can use the same arguments and Lemma A.2 along with Assumptions 1 and 2(a). Finally, we can calculate the moments of ϵ_i and the cross moments of ε_i and ϵ_i with similar arguments. □

For any $t \in [0,T]$, we define $\nearrow_t$ as $\nearrow_t = \frac{t}{\Delta_n}$, $\searrow_t$ as $\searrow_t = 1 - \frac{t}{\Delta_n}$, and $\wedge_t$ as $\wedge_t = \frac{t \wedge (2\Delta_n - t)}{\Delta_n}$. The next lemma greatly simplifies notations for the proofs.

Lemma A.6. *We have* $\bar{\nu}_i - \nu_{i-1} = \int_0^{\Delta_n}\searrow_t d\nu_{i-1+t}, \nu_i - \bar{\nu}_i = \int_0^{\Delta_n}\nearrow_t d\nu_{i-1+t}$, *and* $\Delta_i\bar{\nu} = \int_0^{2\Delta_n}\wedge_t d\nu_{i-2+t}$.

Proof of Lemma A.6. We have

$$\begin{aligned}
\bar{\nu}_i - \nu_{i-1} &= \frac{1}{\Delta_n}\left(\int_{(i-1)\Delta_n}^{i\Delta_n}\big(\nu_{i-1} + \int_{(i-1)\Delta_n}^{t} d\nu_s\big)dt - \nu_{i-1}\Delta_n\right) \\
&= \frac{1}{\Delta_n}\left(\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{(i-1)\Delta_n}^{t} d\nu_s\, dt\right) \\
&= \frac{1}{\Delta_n}\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{s}^{i\Delta_n} dt\, d\nu_s \\
&= \frac{1}{\Delta_n}\int_{(i-1)\Delta_n}^{i\Delta_n}(i\Delta_n - s)\, d\nu_s
\end{aligned}$$

$$= \int_0^{\Delta_n} \diagdown_s \, d\nu_{i-1+s},$$

and symmetry yields $\nu_i - \bar{\nu}_i = \int_0^{\Delta_n} \diagup_t \, d\nu_{i-1+t}$. We also have that $\bar{\nu}_i - \bar{\nu}_{i-1} = \bar{\nu}_i - \nu_{i-1} + \nu_{i-1} - \bar{\nu}_{i-1} = \int_{\Delta_n}^{2\Delta_n} \diagdown_t \, d\nu_{i-2+t} + \int_0^{\Delta_n} \diagup_t \, d\nu_{i-2+t} = \int_0^{2\Delta_n} \wedge_t \, d\nu_{i-2+t}$. We can show the third assertion by using similar arguments. □

Without loss of generality and with an abuse of notation, we rewrite ν_t itself as an Itô semimartingale with Grigelionis representation of the form (2.2). Assumption 2(c) implies that a.s. $\sum_{s\leq t} |\Delta \nu_s| < \infty$ for any $t \in [0, T]$ and we obtain

$$\nu_t = \int_0^t b'_s \, ds + \int_0^t \sigma_s \, dW_s + \sum_{s\leq t} \Delta \nu_s. \tag{A.16}$$

Thus, we can define the continuous part of the process ν_t as $\nu_t^{(c)} = \int_0^t b'_s \, ds + \int_0^t \sigma_s \, dW_s$ and the discontinuous part of the process ν_t as $\nu_t^{(d)} = \sum_{s\leq t} \Delta \nu_s$. For any $i = 1, \dots, M$, we also define $\widehat{\bar{\nu}}_i^{(c)}$ as $\widehat{\bar{\nu}}_i^{(c)} = \bar{\nu}_i^{(c)} + u_i$. The next lemma shows that we can remove the discontinuous part of ν_t in the remaining of the proofs.

Lemma A.7. *Under Assumptions 1 to 2(e), we have*

$$\Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \left(\Delta_i \widehat{\bar{\nu}}^n\right)^2 \mathbb{1}_{\{|\Delta_i \widehat{\bar{\nu}}^n| \leq \alpha \Delta_n^{\varpi}\}} = \Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \left(\Delta_i \widehat{\bar{\nu}}^{(c)}\right)^2 + o_{\mathbb{P}}(1).$$

Proof of Lemma A.7. Step 1 This step shows that

$$\Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \left(\Delta_i \widehat{\bar{\nu}}\right)^2 \mathbb{1}_{\{|\Delta_i \widehat{\bar{\nu}}| \leq \alpha \Delta_n^{\varpi}\}} = \Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \left(\Delta_i \widehat{\bar{\nu}}^{(c)}\right)^2 + o_{\mathbb{P}}(1).$$

From Eq. (A.16) along with Assumption 2(c), we can deduce that $4(\Delta_i \widehat{\bar{\nu}})^2 = (\Delta_i \widehat{\bar{\nu}}^{(c)})^2 + 2(\Delta_i \widehat{\bar{\nu}}^{(c)})(\Delta_i \bar{\nu}^{(d)}) + (\Delta_i \bar{\nu}^{(d)})^2$ and thus

$$\begin{aligned}
\Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} (\Delta_i \widehat{\bar{\nu}})^2 \mathbb{1}_{\{|\Delta_i \widehat{\bar{\nu}}| \leq \alpha \Delta_n^{\varpi}\}} &= \Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} (\Delta_i \widehat{\bar{\nu}}^{(c)})^2 \mathbb{1}_{\{|\Delta_i \widehat{\bar{\nu}}| \leq \alpha \Delta_n^{\varpi}\}} \\
&\quad + \Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} 2(\Delta_i \widehat{\bar{\nu}}^{(c)})(\Delta_i \bar{\nu}^{(d)}) \mathbb{1}_{\{|\Delta_i \widehat{\bar{\nu}}| \leq \alpha \Delta_n^{\varpi}\}} \\
&\quad + \Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} (\Delta_i \bar{\nu}^{(d)})^2 \mathbb{1}_{\{|\Delta_i \widehat{\bar{\nu}}| \leq \alpha \Delta_n^{\varpi}\}} \\
&= (\mathrm{I}) + (\mathrm{II}) + (\mathrm{III}).
\end{aligned}$$

In what follows, we first show that (I) $= \Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \big(\Delta_i \widehat{\bar{\nu}}^{(c)}\big)^2 + o_{\mathbb{P}}(1)$. The domination $\mathbb{1}_{\{|x|>a\}} \leq 2^k |x|^k / a^k$ along with Lemma A.6 and Eq. (A.15) for any $k > 0$ gives

$$\Big|(\mathrm{I}) - \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \big(\Delta_i \widehat{\bar{\nu}}^{(c)}\big)^2\Big| = O_{L^1}\bigg(\Delta_n^{-\frac{1}{2}} \Delta_n^{-1} \frac{\Delta_n^{(k+2)/2} + (n\Delta_n)^{-(k+2)/2}}{\varpi_i^k}\bigg)$$
$$+ O_{\mathbb{P}}\bigg(\frac{\Delta_n^{-\frac{1}{2}}}{\varpi_i} \Delta_n^{-\frac{1}{p}} \big(\Delta_n + (n\Delta_n)^{-1}\big)\bigg). \tag{A.17}$$

By choosing sufficiently large k and p and Assumption 2(d), we obtain (I) $= \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \big(\Delta_i \widehat{\bar{\nu}}^{(c)}\big)^2 + o_{\mathbb{P}}(1)$. For (II), we have by Hölder's inequality and Assumption 2(c) that

$$|(\mathrm{II})| = O_{\mathbb{P}}\bigg(\frac{\Delta_n^{-\frac{1}{2}}}{\varpi_i^k} \Delta_n^{-\frac{1}{p}} \big(\Delta_n^{(k+1)/2} + (n\Delta_n)^{-(k+1)/2}\big)\bigg) + O_{\mathbb{P}}\big(\Delta_n^{-\frac{1}{2}} \varpi_i^{2-\beta}\big).$$

Thus, (II) $= o_{\mathbb{P}}(1)$ by Assumption 2(d) and with a sufficiently large k. Finally, we can show that (III) $= o_{\mathbb{P}}(1)$ with the same arguments as for (II).

Step 2 We show that

$$\Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \big(\Delta_i \widehat{\bar{\nu}}^n\big)^2 \mathbb{1}_{\{|\Delta_i \widehat{\bar{\nu}}^n| \leq \alpha \Delta_n^{\varpi}\}} = \Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \big(\Delta_i \widehat{\bar{\nu}}\big)^2 \mathbb{1}_{\{|\Delta_i \widehat{\bar{\nu}}| \leq \alpha \Delta_n^{\varpi}\}} + o_{\mathbb{P}}(1).$$

We first show that $\big(\mathbb{E}\big|\Delta_i(\bar{\nu}^{n(c)} - \bar{\nu}^{(c)})\big|^k\big)^{\frac{1}{k}} \leq C n^{-\frac{5}{8}}$ for all $k > 0$ by Eq. (A.13) in Lemma A.2 and Assumptions 1 and 2(a), Burkholder-Davis-Gundy inequality along with Assumptions 2(b) and 2(e). By similar arguments, we show that $\sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \big|\Delta_i(\bar{\nu}^{n(d)} - \bar{\nu}^{(d)})\big| = O_{\mathbb{P}}\big(n^{-\frac{3}{8}}\big)$. If we define (IV) as (IV) $= \Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \big(\Delta_i \widehat{\bar{\nu}}^n\big)^2 \Big(\mathbb{1}_{\{|\Delta_i \widehat{\bar{\nu}}^n| \leq \alpha \Delta_n^{\varpi}\}} - \mathbb{1}_{\{|\Delta_i \widehat{\bar{\nu}}| \leq \alpha \Delta_n^{\varpi}\}}\Big)$ and (V) as

$$(\mathrm{V}) = \Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \big(\Delta_i \widehat{\bar{\nu}}^n\big)^2 \big(\mathbb{1}_{\{|\Delta_i \widehat{\bar{\nu}}^n| \leq \alpha \Delta_n^{\varpi}\}} - \mathbb{1}_{\{|\Delta_i \widehat{\bar{\nu}}| \leq \alpha \Delta_n^{\varpi}\}}\big),$$

it is sufficient to show that (IV) $= o_{\mathbb{P}}(1)$ and (V) $= o_{\mathbb{P}}(1)$. For (IV), we have

$$\begin{aligned} &\mathbb{1}_{\{|\Delta_i \widehat{\bar{\nu}}^n| \leq \alpha \Delta_n^{\varpi}\}} - \mathbb{1}_{\{|\Delta_i \widehat{\bar{\nu}}| \leq \alpha \Delta_n^{\varpi}\}} \\ &= \mathbb{1}_{\{|\Delta_i \widehat{\bar{\nu}}| > \alpha \Delta_n^{\varpi}, |\Delta_i \widehat{\bar{\nu}}^n| \leq \alpha \Delta_n^{\varpi}\}} - \mathbb{1}_{\{|\Delta_i \widehat{\bar{\nu}}| \leq \alpha \Delta_n^{\varpi}, |\Delta_i \widehat{\bar{\nu}}^n| > \alpha \Delta_n^{\varpi}\}} \end{aligned}$$

and that $\big\{|\Delta_i \widehat{\bar{\nu}}| > \alpha \Delta_n^{\varpi}, |\Delta_i \widehat{\bar{\nu}}^n| \leq \alpha \Delta_n^{\varpi}\big\} \subset \big\{|\Delta_i(\widehat{\bar{\nu}}^n - \widehat{\bar{\nu}})| > \alpha \Delta_n^{\varpi}, |\Delta_i \widehat{\bar{\nu}}^n| \leq \alpha \Delta_n^{\varpi}\big\} \cup \big\{|\Delta_i(\widehat{\bar{\nu}}^n - \widehat{\bar{\nu}})| \leq \alpha \Delta_n^{\varpi}, |\Delta_i \widehat{\bar{\nu}}| \in (\alpha \Delta_n^{\varpi}, 2\alpha \Delta_n^{\varpi}]\big\}$, $\big\{|\Delta_i \widehat{\bar{\nu}}| \leq \alpha \Delta_n^{\varpi}, |\Delta_i \widehat{\bar{\nu}}^n| > \alpha \Delta_n^{\varpi}\big\} \subset \big\{|\Delta_i(\widehat{\bar{\nu}}^n - \widehat{\bar{\nu}})| > \alpha \Delta_n^{\varpi}, |\Delta_i \widehat{\bar{\nu}}| \leq \alpha \Delta_n^{\varpi}\big\} \cup \big\{|\Delta_i(\widehat{\bar{\nu}}^n - \widehat{\bar{\nu}})| \leq \alpha \Delta_n^{\varpi}, |\Delta_i \widehat{\bar{\nu}}^n| \in$

$(\alpha\Delta_n^{\varpi}, 2\alpha\Delta_n^{\varpi}]\big\}$. With similar arguments as for the proof of Eq. (A.17), (II) and (III), it leads to

$$\Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \big(\Delta_i \widehat{\nu}^n\big)^2 \mathbb{1}_{\{|\Delta_i \widehat{\nu}| > \alpha\Delta_n^{\varpi}, |\Delta_i \widehat{\nu}^n| \leq \alpha\Delta_n^{\varpi}\}} = o_{\mathbb{P}}(1).$$

We can also show $\Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \big(\Delta_i \widehat{\nu}^n\big)^2 \mathbb{1}_{\{|\Delta_i \widehat{\nu}| \leq \alpha\Delta_n^{\varpi}, |\Delta_i \widehat{\nu}^n| > \alpha\Delta_n^{\varpi}\}} = o_{\mathbb{P}}(1)$, so we can deduce that (IV) $\xrightarrow{\mathbb{P}} 0$. Finally, we have with similar arguments that (V) $= o_{\mathbb{P}}(1)$. □

The next lemma shows that we can remove the drift from ν_t if we assume Novikov's condition Assumption 2(f) (see Novikov (1972)), which is required to apply Girsanov theorem (see Girsanov (1960)). We consider an equivalent probability measure $\mathbb{P}^*$ under which ν_t is a local martingale, i.e., $\nu_t = \nu_0 + \int_0^t \sigma_s dW_s^*$, where W_t^* is a standard Wiener process under $\mathbb{P}^*$.

Lemma A.8. *Under Assumption 2(f) and if we assume that the statement of Theorem 4.1 holds under $\mathbb{P}^*$, the same statement holds under $\mathbb{P}$.*

Proof of Lemma A.8. We define $\overline{M}_t$ as $\overline{M}_t = \exp\big(\int_0^t \frac{b_s'}{\sigma_s} dW_s - \frac{1}{2}\int_0^t \frac{(b_s')^2}{\sigma_s^2} ds\big)$ for any $0 \leq t \leq T$, which by Assumption 2(f) satisfies Novikov's condition and thus is a positive martingale. By Girsanov theorem, we can thus consider an equivalent probability distribution $\mathbb{P}^*$. Then, we have that the Radon-Nikodym derivative is defined as $\frac{d\mathbb{P}^*}{d\mathbb{P}}|\mathcal{F}_T = \overline{M}_T$ and $W_t^* = W_t + \int_0^t \frac{b_s'}{\sigma_s} dW_s$ is a standard Wiener process under $\mathbb{P}^*$. To show that the statement of Theorem 4.1 holds under $\mathbb{P}$, it is sufficient to prove that

$$\mathbb{E}_{\mathbb{P}}\Big[h(X)\mathbb{1}_E\Big] \to \mathbb{E}_{\mathbb{P}}\Big[h\Big(\int_0^T w_t d\widetilde{W}_t\Big)\mathbb{1}_E\Big], \tag{A.18}$$

for any $E \in \mathcal{F}_T$ and any measurable function h. By a change of probability in the expectation, we obtain

$$\mathbb{E}_{\mathbb{P}}\Big[h(X)\mathbb{1}_E\Big] = \mathbb{E}_{\mathbb{P}^*}\Big[h(X)\mathbb{1}_E \overline{M}_T^{-1}\Big].$$

Since $\overline{M}_T^{-1} \in \mathcal{F}_T$ and the statement of Theorem 4.1 holds under $\mathbb{P}^*$, we can deduce that

$$\mathbb{E}_{\mathbb{P}^*}\Big[h(X)\mathbb{1}_E \overline{M}_T^{-1}\Big] \to \mathbb{E}_{\mathbb{P}^*}\Big[h\Big(\int_0^T w_t d\widetilde{W}_t\Big)\mathbb{1}_E \overline{M}_T^{-1}\Big].$$

Finally, we obtain $\mathbb{E}_{\mathbb{P}^*}\Big[h\Big(\int_0^T w_t d\widetilde{W}_t\Big)\mathbb{1}_E \overline{M}_T^{-1}\Big] = \mathbb{E}_{\mathbb{P}}\Big[h\Big(\int_0^T w_t d\widetilde{W}_t\Big)\mathbb{1}_E\Big]$, by another change of probability in the expectation. Thus, we have shown Appendix A.3. □

A.4. Proof of Theorem 4.1

By Lemmas A.7 and A.8 along with Assumptions 1 to 2(f), we can assume that ν_t is continuous with no drift. Let us write the $\mathcal{F}_t$-martingale X_t as

$$X_t = \begin{bmatrix} \Delta_n^{-1} n^{-1}\big(\widehat{\text{Mean}} - \text{Mean}\big)_t \\ \Delta_n^{-\frac{1}{2}} n^{-2}\big(\widehat{\text{Var}_1} - \text{Var}_1\big)_t \\ \Delta_n^{-\frac{1}{2}} n^{-2}\big(\widehat{\text{Var}_2} - \text{Var}_2\big)_t \end{bmatrix} = \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \big(\boldsymbol{\xi}_i - \mathbb{E}_{i-1}[\boldsymbol{\xi}_i]\big).$$

Here, $\boldsymbol{\xi}_i = [\boldsymbol{\xi}_{i,1}; \boldsymbol{\xi}_{i,2}; \boldsymbol{\xi}_{i,3}]$ is a 3-dimensional vector defined by

$$\boldsymbol{\xi}_i = \begin{bmatrix} \Delta_n^{-1}\big(\widehat{\nu}_i \Delta_n - \int_{(i-1)\Delta_n}^{i\Delta_n} \nu_t\, dt\big) \\ \Delta_n^{-\frac{1}{2}}\Big((\Delta_i \widehat{\nu})^2 - \mathbb{E}_{i-1}\big[(\Delta_i\widehat{\nu})^2\big] + \mathbb{E}_i\big[(\Delta_{i+1}\widehat{\nu})^2\big] - \int_{(i-1)\Delta_n}^{i\Delta_n} \big(\tfrac{2}{3}\sigma_t^2 + 2\breve{\vartheta}_t\big)dt\Big) \\ \cdots \end{bmatrix},$$

and $\mathbf{F}_t$ is the discretized filtration, i.e., $\mathbf{F}_t = \mathcal{F}_{\Delta_n \lfloor t/\Delta_n \rfloor}$, for any $t \in [0, T]$. In $\boldsymbol{\xi}_i$, we do not explicit the third component, which is similar to the second component. We now verify that all the conditions, namely the five Conditions (7.27) to (7.31), from Theorem IX.7.28 (pp. 590-591) in Jacod and Shiryaev (2013) are satisfied. We set $Z_t = 0$, which is obviously a square-integrable $\mathcal{F}_t$-martingale. Thus, Condition (7.29) is directly satisfied. We also have that each $\boldsymbol{\xi}_i$ is componentwise square-integrable, because $\widehat{\nu}_i$, ν_t, and σ_t have bounded 4th moments by Lemma A.3 along with Assumption 1, and Assumption 2(g). We show that Condition (7.27) holds with $B_t = 0$ in the following proposition.

Proposition A.1. *Under Assumptions 1 to 2(b) and 2(g), we have for $j = 1, 2, 3$ that*

$$\sup_{0 \le t \le T} \left| \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}_{i-1}[\boldsymbol{\xi}_{i,j}] \right| \xrightarrow{\mathbb{P}} \mathbf{0}. \tag{A.19}$$

Proof of Proposition A.1. By Eq. (A.14) from Lemma A.5 along with Assumptions 1 and 2(a), we have

$$\sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \xi_{i,1} = \frac{1}{n\Delta_n} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \int_0^{(i-1)\Delta_n} \Delta_i \Psi(-t) dM_t + \int_{(i-1)\Delta_n}^{i\Delta_n} \Psi(i\Delta_n - t) dM_t + o_{\mathbb{P}}(1).$$

Since it is a martingale, we deduce Eq. (A.19) from the martingale definition. By Lemma A.6, we obtain that

$$\mathbb{E}_{i-1}[\xi_{i,2}] = \Delta_n^{-\frac{1}{2}} \int_{i-1}^{i+1} \Big(\wedge_t^2 - \frac{1}{3}\Big)\mathbb{E}_{i-1}[\sigma_t^2]\, dt + \Delta_n^{-\frac{1}{2}} \mathbb{E}_{i-1}\left[u_{i+1}^2 - \frac{\bar{\vartheta}_{i+1}}{n\Delta_n} + u_i^2 - \frac{\bar{\vartheta}_i}{n\Delta_n}\right]$$

$$+ \underline{O}_{L^2}\left(\Delta_n^{\frac{1}{2}} \frac{\log n}{n\Delta_n}\right) = \text{(I)} + \text{(II)} + \underline{O}_{L^2}\left(\Delta_n^{\frac{1}{2}} \frac{\log n}{n\Delta_n}\right). \tag{A.20}$$

Since $\int_{i-1}^{i+1} \big(\frown_t^2 - \frac{1}{3}\big)\sigma_{i-1}^2 \, dt = 0$ and by Assumption 2(g), we obtain

$$\sup_{0\le t\le T} \big|\text{(I)}\big| = o_{\mathbb{P}}(1).$$

By Lemma A.5 with Assumptions 1 and 2(a), we have $\big|\sup_{0\le t\le T}\text{(II)}\big| = o_{L^1}(1)$, whenever $\big(\frac{\log n}{n^2}\big)^{\frac{2}{7}} \prec \Delta_n$ which holds by Assumption 2(b). Thus, we can deduce that $\sup_{0\le t\le T}\big|\sum_{i=1}^{\lfloor t/\Delta_n\rfloor} \mathbb{E}_{i-1}[\xi_{i,2}]\big| \xrightarrow{\mathbb{P}} 0$. The proof of the case $j=3$ follows with the same arguments. □

We show that Condition (7.28) holds in the following proposition.

Proposition A.2. *Under Assumptions 1 to 2(b), 2(e) and 2(g), we have for any $0\le t\le T$ that*

$$\sum_{i=1}^{\lfloor t/\Delta_n\rfloor} \mathbb{V}_{i-1}[\boldsymbol{\xi}_i] \xrightarrow{\mathbb{P}} \int_0^t w_u w_u^\top \, du. \tag{A.21}$$

Proof of Proposition A.2. Since $\xi_{i,1} = u_i$ and $\mathbb{E}_{i-1}[u_i] = \underline{O}_{L^2}\big(\frac{\log n}{n\Delta_n}\big)$, we have $\sum_{i=1}^{\lfloor t/\Delta_n\rfloor}\big(\mathbb{E}_{i-1}[\xi_{i,1}]\big)^2 = \sum_{i=1}^{\lfloor t/\Delta_n\rfloor} \underline{O}_{L^1}\big(\frac{(\log n)^2}{(n\Delta_n)^2}\big) = o_{L^1}(1)$ by Assumption 2(b). Thus, Riemann integrability and Assumption 2(b) yield $\sum_{i=1}^{\lfloor t/\Delta_n\rfloor} \mathbb{V}_{i-1}[\xi_{i,1}] \xrightarrow{\mathbb{P}} \int_0^t \breve{\vartheta}_u du$. From Eq. (A.20) and Assumptions 2(e) and 2(g), we obtain

$$\sum_{i=1}^{\lfloor t/\Delta_n\rfloor} \mathbb{V}_{i-1}[\xi_{i,2}] = \sum_{i=1}^{\lfloor t/\Delta_n\rfloor} \mathbb{E}_{i-1}\xi_{i,2}^2 + o_{\mathbb{P}}(1).$$

By Eq. (A.15) from Lemma A.5 along with Assumptions 1 and 2(a), and by Assumptions 2(b), 2(e) and 2(g), we obtain

$$\sum_{i=1}^{\lfloor t/\Delta_n\rfloor} \mathbb{E}_{i-1}[\xi_{i,2}^2] \xrightarrow{\mathbb{P}} \int_0^t \breve{\sigma}_u^4 + 4\breve{\sigma}_u^2\breve{\vartheta}_u + 12\breve{\vartheta}_u^2 du.$$

The proof of $\sum_{i=1}^{\lfloor t/\Delta_n\rfloor} \mathbb{V}_{i-1}[\xi_{i,3}]$ follows with the same arguments if we replace Δ_n by $2\Delta_n$. Since ν_t is a martingale,

$$\sum_{i=1}^{\lfloor t/\Delta_n\rfloor} \text{Cov}_{i-1}[\xi_{i,1},\xi_{i,2}] = \sum_{i=1}^{\lfloor t/\Delta_n\rfloor} \mathbb{E}_{i-1}[\xi_{i,1}\xi_{i,2}] + o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1)$$

and $\sum_{i=1}^{\lfloor t/\Delta_n\rfloor} \text{Cov}_{i-1}[\xi_{i,1},\xi_{i,3}] = o_{\mathbb{P}}(1)$. Finally, we obtain with the same arguments that $\sum_{i=1}^{\lfloor t/(2\Delta_n)\rfloor} \text{Cov}_{i-1}[\xi_{i,2},\xi_{i,3}] \xrightarrow{\mathbb{P}} \frac{1}{2}\int_0^t \frac{29}{24}\breve{\sigma}_u^4 + \frac{3}{2}\breve{\sigma}_u^2\breve{\vartheta}_u + \frac{3}{2}\breve{\vartheta}_u^2 du.$ □

We show that Condition (7.30) holds in the following proposition.

Proposition A.3. *Under Assumptions 1 to 2(b), 2(e) and 2(g), we have for any* $0 \le t \le T$ *that*

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}_{i-1}[\|\boldsymbol{\xi}_i\|^2 \mathbb{1}_{\{\|\boldsymbol{\xi}_i\| > \varepsilon\}}] \xrightarrow{\mathbb{P}} 0, \tag{A.22}$$

Proof of Proposition A.3. By Hölder's inequality, we have

$$\mathbb{E}\Big|\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}_{i-1}[\|\boldsymbol{\xi}_i\|^2 \mathbb{1}_{\{\|\boldsymbol{\xi}_i\| > \varepsilon\}}\Big| \le C\sqrt{\frac{\mathbb{E}\|\boldsymbol{\xi}_i\|^4}{\Delta_n^2}}\sqrt{\mathbb{P}\Big\{\frac{\|\boldsymbol{\xi}_i\|}{\Delta_n^{1/2}} > \frac{\varepsilon}{\Delta_n^{1/2}}\Big\}},$$

thus it is sufficient to have $\mathbb{E}\|\boldsymbol{\xi}_i\|^4 = O(\Delta_n^2)$, whose proof follows from similar arguments as in the proof of Lemma A.2 along with Assumptions 1 to 2(b), 2(e) and 2(g). □

We show that Condition (7.31) holds in the following proposition.

Proposition A.4. *Under Assumptions 1 to 2(b), 2(e) and 2(g), we have for any* $0 \le t \le T$ *and for any bounded 3-dimensional* $\mathcal{F}_t$*-martingale* $\mathbf{M}'$ *that*

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}_{i-1}[\boldsymbol{\xi}_i^\top \Delta_i \mathbf{M}'] \xrightarrow{\mathbb{P}} 0. \tag{A.23}$$

Proof of Proposition A.4. When $\mathbf{M}'_t = W_t$, from the proof of Proposition A.2 along with Assumptions 1 to 2(b), 2(e) and 2(g), we have $\mathbb{E}_{i-1}[\xi_{i,1}\Delta_i W] = \mathbb{E}_{i-1}[u_i \Delta_i W] = \underline{O}_{L^1}\big(\frac{\sqrt{\log n}}{n\sqrt{\Delta_n}}\big)$, and $\mathbb{E}\big(\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}_{i-1}[\xi_{i,2}\Delta_i W]\big)^2 = o(1)$. When $\mathbf{M}'_t$ is a continuous martingale orthogonal to W_t, since $\mathbb{E}|\sigma_t - \sigma_{i-1}|^k \le C\Delta_n^{k\gamma}$, we can approximate locally and replace σ_t^2 in $\boldsymbol{\xi}_i$ by σ_{i-1}^2 by using similar arguments as in the proof of Proposition 4.1 (pp. 15-16) in Barndorff-Nielsen et al. (2006). We denote the local approximation as $\boldsymbol{\xi}'_i$, and its conditional expectation as $\boldsymbol{\xi}_t^{(M)} = \mathbb{E}_t[\boldsymbol{\xi}_i]$. By Theorem III.4.34 (p. 189) in Jacod and Shiryaev (2013), we can express $\boldsymbol{\xi}_t^{(M)}$ into a stochastic integration of W_t and M_t. Thus, the orthogonality of $\mathbf{M}'_t$ implies that $(d\boldsymbol{\xi}_t^{(M)})^\top(d\mathbf{M}'_t) = 0$, so we deduce that

$$\mathbb{E}_{i-1}[\boldsymbol{\xi}_i'^\top \Delta_i \mathbf{M}'_t] = \mathbb{E}_{i-1}\bigg[\int_{(i-1)\Delta_n}^{i\Delta_n} (d\boldsymbol{\xi}_t^{(M)})^\top (d\mathbf{M}'_t)\bigg] = 0.$$

With the same arguments as in the proof of Eq. (6.10) in Todorov and Tauchen (2011), if $\boldsymbol{\xi}_i$ is C-tight and $\mathbf{M}'_t$ is a discontinuous martingale orthogonal to $M_t = N_t - \int_0^t \lambda_s\, ds$, then $\sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \boldsymbol{\xi}_i$ and $\sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \Delta_i \mathbf{M}'$ are jointly tight by Corollary VI.3.33 (p. 353) in Jacod and Shiryaev (2013), and the left-hand side of Eq. (A.23) converges to the predictable quadratic variation of the limit of $\sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \boldsymbol{\xi}_i$ and $\sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \Delta_i \mathbf{M}'$, which is obviously zero due to the orthogonality of continuous and discontinuous martingales. The C-tightness of $\boldsymbol{\xi}_i$ is implied by $\sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \mathbb{E}_{i-1}|\xi_{i,j}|^k \xrightarrow{\mathbb{P}} 0$ for some $k > 2$, and this can be derived from local boundedness of σ_t and $\mathbb{E}u_i^k \leq C(n\Delta_n)^{-k/2}$. With similar arguments, we can show the case when $\mathbf{M}'_t$ is $M_t \Delta_n$. □

The next proposition is useful to prove the normalized CLT with feasible variance. It is based on the continuous mapping theorem along with Slutsky's theorem.

Proposition A.5. *Under Assumptions 1 and 2, we have*

$$\frac{1}{\Delta_n} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} (\Delta_i \widehat{\bar{\nu}})^4 \xrightarrow{\mathbb{P}} \int_0^T \left(\frac{4}{3}\sigma_t^4 + 8\sigma_t^2 \breve{\vartheta}_t + 12\breve{\vartheta}_t^2 \right) dt, \tag{A.24}$$

$$\sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \widehat{\bar{\nu}}_i (\Delta_i \widehat{\bar{\nu}})^2 \xrightarrow{\mathbb{P}} \int_0^T \left(\frac{2}{3}\sigma_t^2 \nu_t + 2\nu_t \breve{\vartheta}_t \right) dt, \tag{A.25}$$

$$\Delta_n \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \widehat{\bar{\nu}}_i^2 \xrightarrow{\mathbb{P}} \int_0^T \nu_t^2 \, dt. \tag{A.26}$$

Proof of Proposition A.5. We have that Lemma A.7 also holds for power greater than 2. Thus, we can a consider continuous ν_t without truncation. For Eq. (A.24), we can show that

$$\begin{aligned}
&\Delta_n^{-1} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} (\Delta_i \widehat{\bar{\nu}})^4 \\
&= \Delta_n^{-1} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \left\{ (\Delta_i \bar{\nu})^4 + 6(\Delta_i \bar{\nu})^2 (\Delta_i u)^2 + (\Delta_i u)^4 \right\} + o_{\mathbb{P}}(1) \\
&= \Delta_n^{-1} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \left\{ 6 \int_{i-2}^{i} (\wedge_t^{(\nu)})^2 (d\wedge_t^{(\nu)})^2 + 6\Big(\int_{i-2}^{i} \wedge_t^2 \sigma_t^2 \, dt \Big) \Big(\mathbb{E}_{i-1} u_i^2 + \mathbb{E}_{i-2} u_{i-1}^2 \Big) \right. \\
&\qquad \left. + \Big(\mathbb{E}_{i-1} u_i^4 + \mathbb{E}_{i-2} u_{i-1}^4 + 6\mathbb{E}_{i-1} u_i^2 \mathbb{E}_{i-2} u_{i-1}^2 \Big) \right\} + o_{\mathbb{P}}(1) \\
&= \Delta_n^{-1} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \left\{ 6\frac{1}{2} \Big(\int_{i-2}^{i} \wedge_t^2 \sigma_t^2 \, dt \Big)^2 \right. \\
&+ 6\Big(\sigma_{i-2}^2 \int_{i-2}^{i} \wedge_t^2 dt \Big) \Big(\frac{\vartheta_{i-2}}{n\Delta_n} + \frac{\vartheta_{i-2}}{n\Delta_n} \Big) \\
&\qquad \left. + \left(3\Big(\frac{\vartheta_{i-2}}{n\Delta_n} \Big)^2 + 3\Big(\frac{\vartheta_{i-2}}{n\Delta_n} \Big)^2 + 6\Big(\frac{\vartheta_{i-2}}{n\Delta_n} \Big)^2 \right) \right\} + o_{\mathbb{P}}(1)
\end{aligned}$$

$$\xrightarrow{\mathbb{P}} \int_0^T \left\{ \frac{4}{3}\sigma_t^4 + 8\sigma_t^2 \frac{\vartheta_t}{n\Delta_n^2} + 12\left(\frac{\vartheta_t}{n\Delta_n^2}\right)^2 \right\} dt.$$

For Eq. (A.25), we have

$$\begin{aligned}
\sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \widehat{\nu}_i (\Delta_i \widehat{\nu})^2 &= \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \bar{\nu}_i \Big((\Delta_i \bar{\nu})^2 + (\Delta_i u)^2 \Big) + o_{\mathbb{P}}(1) \\
&= \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \bar{\nu}_i \Big(\int_{i-2}^{i} \frown_t^2 \sigma_t^2 \, dt + \mathbb{E}_{i-1} u_i^2 + \mathbb{E}_{i-2} u_{i-1}^2 \Big) + o_{\mathbb{P}}(1) \\
&= \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \nu_{i-2} \Big(\sigma_{i-2}^2 \int_{i-2}^{i} \frown_t^2 \, dt + \frac{\vartheta_{i-2}}{n\Delta_n} + \frac{\vartheta_{i-2}}{n\Delta_n} \Big) + o_{\mathbb{P}}(1) \\
&\xrightarrow{\mathbb{P}} \int_0^T \left\{ \frac{2}{3}\sigma_t^2 \nu_t + 2\nu_t \frac{\vartheta_t}{n\Delta_n^2} \right\} dt.
\end{aligned}$$

For Eq. (A.26), we have

$$\Delta_n \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \widehat{\nu}_i^2 = \Delta_n \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \bar{\nu}_i^2 + o_{\mathbb{P}}(1) \xrightarrow{\mathbb{P}} \int_0^T \nu_t^2 \, dt.$$

□

In what follows, we show the consistency of the estimator of non diverging asymptotic variance, and the normalized CLT with feasible variance.

Proof of Eqs. (4.1) and (4.2). A linear combination of Eqs. (A.24) to (A.26) yields Eqs. (4.1) and (4.2). First, $n^{-4}\widehat{\Sigma}_{22}^n$ converges in probability to $\int_0^T \big(\sigma_t^4 + 4\sigma_t^2 \breve{\vartheta}_t + 12\breve{\vartheta}_t^2\big)\, dt$ by Theorem 4.1. Then, we obtain that

$$4\frac{1}{\Delta_n} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} (\Delta_i \widehat{\nu})^4 \xrightarrow{\mathbb{P}} \int_0^T \big(\sigma_t^4 + 6\sigma_t^2 \breve{\vartheta}_t + 9\breve{\vartheta}_t^2\big)\, dt.$$

By subtracting $\frac{3}{1-\|\phi\|_1} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \widehat{\nu}_i (\Delta_i \widehat{\nu})^2$ to it, we obtain

$$4\frac{1}{\Delta_n} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} (\Delta_i \widehat{\nu})^4 - \frac{3}{1-\|\phi\|_1} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \widehat{\nu}_i (\Delta_i \widehat{\nu})^2 \xrightarrow{\mathbb{P}} \int_0^T \big(\sigma_t^4 + 4\sigma_t^2 \breve{\vartheta}_t + 3\breve{\vartheta}_t^2\big)\, dt.$$

If we add $\frac{9}{(1-\|\phi\|_1)^2} \Delta_n \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \widehat{\nu}_i^2$ to it, we obtain

$$4\frac{1}{\Delta_n} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} (\Delta_i \widehat{\nu})^4 - \frac{3}{1-\|\phi\|_1} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \widehat{\nu}_i (\Delta_i \widehat{\nu})^2 + \frac{9}{(1-\|\phi\|_1)^2} \Delta_n \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \widehat{\nu}_i^2$$

$$\xrightarrow{\mathbb{P}} \int_0^T \left(\sigma_t^4 + 4\sigma_t^2 \breve{\vartheta}_t + 12\breve{\vartheta}_t^2\right) dt.$$

Since $\frac{2}{3}\frac{\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2}{\widehat{\mathrm{Mean}}}$ converges to $\frac{1}{(1-\|\phi\|_1)^2}$ in probability, we have

$$\begin{aligned}\frac{1}{n^4}\widehat{\Sigma}_{22}^n &= \frac{3}{4}\frac{\widehat{\kappa}_4(\Delta_n)}{n^4} - (3\widehat{\eta})\frac{\widehat{\kappa}_3(\Delta_n)}{n^4} + (3\widehat{\eta})^2\frac{\widehat{\kappa}_4(\Delta_n)}{n^4} \\ &\xrightarrow{\mathbb{P}} \int_0^T \left(\sigma_t^4 + 4\sigma_t^2 \breve{\vartheta}_t + 12\breve{\vartheta}_t^2\right) dt.\end{aligned}$$

We can show the other cases with similar arguments. □

A.5. Proofs from Section 5

Proof of Corollary 5.1. A direct application of the Delta method to Theorem 4.1 with $\|\phi\|_1 = 0$ yields the result. □

Proof of Corollary 5.2. A direct application of the Delta method to Theorem 4.1 yields the result. □

Proof of Corollary 5.3. A direct application of the Delta method to Theorem 4.1 yields the result. □

Proof of Corollary 5.4. A direct application of the Delta method to Theorem 4.1 yields $S \xrightarrow{\mathcal{L}-s} \chi_1^2$ under H_0. Under H_1, we can show that $S \to \infty$ since $\widehat{\mathrm{AVar}}(\widehat{\|\phi\|_1}) = O_{\mathbb{P}}(1)$. It results from $\frac{\widehat{\mathrm{Mean}}}{\Delta_n^2(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)} = O_{\mathbb{P}}(1)$, $\Delta_n^{-2}\widehat{\mathrm{Mean}} = O_{\mathbb{P}}(1)$, and $\Delta_n^{-\frac{1}{2}}\widehat{\|\phi\|_1} \succeq \Delta_n^{-\frac{1}{2}}$. □

Proof of Proposition 5.1. By Theorem 4.1, it is sufficient to consider only the components of $\widehat{\mathrm{Var}}$. We have

$$\begin{aligned}\frac{1}{n^2}\widehat{\mathrm{Var}} &= \sum_{i=1}^{\lfloor T/\Delta_n \rfloor}\left(\widehat{\overline{\nu}}_i - \frac{\widehat{\mathrm{Mean}}}{nT}\right)^2 \\ &= \sum_{i=1}^{\lfloor T/\Delta_n \rfloor}\left(\nu + u_i - \frac{\Delta_n}{T}\sum_{j=1}^{\lfloor T/\Delta_n \rfloor}(\nu + u_j)\right)^2 \\ &= \sum_{i=1}^{\lfloor T/\Delta_n \rfloor}\left(u_i - \frac{\Delta_n}{T}\sum_{j=1}^{\lfloor T/\Delta_n \rfloor}u_j\right)^2 \\ &= \sum_{i=1}^{\lfloor T/\Delta_n \rfloor}u_i^2 - \frac{\Delta_n}{T}\left(\sum_{i=1}^{\lfloor T/\Delta_n \rfloor}u_i\right)^2.\end{aligned}$$

For the second term, we can deduce that $\frac{\Delta_n}{T}\left(\sum_{i=1}^{\lfloor T/\Delta_n \rfloor} u_i\right)^2 = o_{\mathbb{P}}(1)$. For the first term, we have by Lemma A.5 that $\mathbb{E}_{i-1}[u_i^2 - \Delta_n \breve{\vartheta}] = \underline{O}_{L^1}\left(\frac{\log n}{(n\Delta_n)^2}\right)$. Thus, we obtain that

$$\sup_{0 \le t \le T} \Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \mathbb{E}_{i-1}[u_i^2 - \Delta_n \breve{\vartheta}] \xrightarrow{\mathbb{P}} 0.$$

We also have that

$$\begin{aligned}
\Delta_n^{-1} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \mathbb{E}_{i-1}(u_i^2 - \Delta_n \breve{\vartheta})^2 &= \Delta_n^{-1} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \mathbb{E}_{i-1}\left[u_i^4 - 2\breve{\vartheta}\Delta_n u_i^2 + (\breve{\vartheta}\Delta_n)^2\right] \\
&= \Delta_n^{-1} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} 2\left(\frac{\vartheta}{n\Delta_n}\right)^2 + o_{\mathbb{P}}(1) \\
&= 2\frac{\vartheta}{n\Delta_n^2} T + o_{\mathbb{P}}(1).
\end{aligned}$$

Thus, we obtain $\mathrm{AVar}(\widehat{\mathrm{Var}}) = 2\frac{\vartheta}{c}T$ by Assumption 2(b). For $\mathrm{ACov}\left(\widehat{\mathrm{Var}}, \widehat{\mathrm{Var}_1}\right)$, we first have that $(\Delta_i \widehat{\nu})^2 = (\Delta_i u)^2$. Then, we obtain that

$$n^{-2}\widehat{\mathrm{Var}_1} = \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} 2(u_i - u_{i-1})u_i + o_{\mathbb{P}}(1).$$

Thus, we have

$$\begin{aligned}
&\Delta_n^{-1} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \mathbb{E}_{i-1}\Big[(u_i^2 - \Delta_n \breve{\vartheta}) 2(u_i^2 - u_{i-1}u_i - \Delta_n \breve{\vartheta})\Big] \\
&= 2\Delta_n^{-1} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \mathbb{E}_{i-1}\left[u_i^4 - u_{i-1}u_i^3 - 2\frac{\vartheta}{n\Delta_n}u_i^2 + u_{i-1}u_i\frac{\vartheta}{n\Delta_n} + \left(\frac{\vartheta}{n\Delta_n}\right)^2\right] \\
&= 4\Delta_n^{-1} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \left(\frac{\vartheta}{n\Delta_n}\right)^2 + o_{\mathbb{P}}(1) \\
&\xrightarrow{\mathbb{P}} 4\left(\frac{\vartheta}{c}\right)^2 T.
\end{aligned}$$

Here, we use Assumption 2(b) in the convergence. We obtain $\mathrm{ACov}\left(\widehat{\mathrm{Var}}, \widehat{\mathrm{Var}_2}\right) = \frac{1}{2}\left(\frac{\vartheta}{c}\right)^2 T$ with the same arguments. We can apply Theorem IX.7.28 (pp. 590-591) in Jacod and Shiryaev (2013) since we can show that all the remaining conditions are met with the same arguments as in the proof of Theorem 4.1. We have $n^{-2}\widehat{\mathrm{Var}_1}/(2T) \xrightarrow{\mathbb{P}} \breve{\vartheta}$, and $n^{-2}\widehat{\mathrm{Var}}/T \xrightarrow{\mathbb{P}} \breve{\vartheta}$. Thus, the continuous mapping theorem and Slutsky's theorem yield Eq. (5.14). Finally, a direct application of the delta method gives Eq. (5.15). □

Proof of Corollary 5.5. Proposition 5.1 yields $S' \xrightarrow{\mathcal{L}-s} \chi_1^2$ under H_0'. Under H_1', we have

$$\begin{aligned}
\Delta_n^2 \widehat{\mathrm{Var}} &= \Delta_n^2 \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \left(\widehat{\lambda}_i - \frac{\widehat{\mathrm{Mean}}}{T} \right)^2 \\
&= \Delta_n^2 n^2 \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \left(\widehat{\bar{\nu}}_i - \frac{\Delta_n}{T} \sum_{j=1}^{\lfloor T/\Delta_n \rfloor} \widehat{\bar{\nu}}_j \right)^2 \\
&= \Delta_n^2 n^2 \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \left(\bar{\nu}_i - \frac{\Delta_n}{T} \sum_{j=1}^{\lfloor T/\Delta_n \rfloor} \bar{\nu}_j \right)^2 + O_{\mathbb{P}}(1) \\
&= \Delta_n n^2 \int_0^T (\nu_t - \bar{\nu})^2 \, dt + o_{\mathbb{P}}(1).
\end{aligned}$$

It implies $\widehat{\mathrm{Mean}}/(\Delta_n^2 \widehat{\mathrm{Var}}) \xrightarrow{\mathbb{P}} 0$ because $\mathbb{P}(\int_0^T (\nu_t - \bar{\nu})^2 \, dt > 0) = 1$. Thus, we obtain $\widehat{\|\phi\|_1^H} \xrightarrow{\mathbb{P}} 1$ and $\widehat{\mathrm{AVar}}(\widehat{\|\phi\|_1} - \widehat{\|\phi\|_1^H} = O_{\mathbb{P}}(1)$. It implies that the test statistic S' explodes. $\square$