High-frequency estimation of Itô semimartingale baseline for Hawkes processes

Yoann Potiron* Olivier Scaillet† Vladimir Volkov‡ and Seunghyeon Yu§

Abstract

We introduce Hawkes self-exciting processes with a baseline driven by an Itô semimartingale with possible jumps. Three measures for the intensity of these Hawkes processes, i.e., integrated intensity, integrated baseline, and integrated volatility of the baseline are studied. The statistics are based on empirical averages and preaveraging of local Poisson estimates constructed from high-frequency quote times. We also propose feasible statistics induced by central limit theory with in-fill asymptotic theory and develop a test for the absence of a Hawkes component and a test for constant baseline. An empirical application on high-frequency data of the E-mini S&P500 futures contracts finds rejection of both the null hypothesis of no Hawkes excitation and that of constant baseline contributing to the formal identification of patterns in high-frequency trading activity.

Keywords: Hawkes tests; in-fill asymptotics; high-frequency data; Itô semimartingale; self-exciting process; time-varying baseline

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

*Faculty of Business and Commerce, Keio University. E-mail: potiron@fbc.keio.ac.jp

†(CONTACT) Swiss Finance Institute, University of Geneva. E-mail: olivier.scaillet@unige.ch

‡School of Business and Economics, University of Tasmania and HSE University. E-mail: vladimir.volkov@utas.edu.au

§Kellogg School of Management, Northwestern University. E-mail: seunghyeon.yu@kellogg.northwestern.edu

1 Introduction

Quote intensity is a central measure in financial economics. Quotes are driven by an underlying stochastic process, which can be interpreted as the flow of news in the mixture of distribution hypothesis (see Clark (1973), Epps and Epps (1976) and Tauchen and Pitts (1983)). This implies that the quote intensity is random and persistent whenever the unobservable arrival of news is random and persistent. Some market microstructure theories suggest that optimal execution strategies are based on time-varying quote intensity (see Kyle (1985), Admati and Pfleiderer (1988) and Almgren and Chriss (2001)). Moreover, the increased volatility observed during distressed market conditions often coincides with abnormal increases in quote intensity.

To investigate high-frequency data, Engle and Russell (1998) and Engle (2000) introduce time series models, namely autoregressive conditional duration (ACD) models, based on quote times. More generally, point processes are widely used in econometrics to characterize quote times. These point processes in which the time between two quotes, i.e., the duration, is random can be seen as a natural extension of standard time series, in which the time between two quotes is fixed. The main stylized fact in this strand of literature is the presence of duration clustering over time. This motivates the so-called Hawkes self-exciting processes (see Hawkes (1971b) and Hawkes (1971a)). We define N_t as the cumulative number of quotes from the starting time 0 to the final time t and λ_t is the intensity. Then, a standard definition of a Hawkes self-exciting process is given by

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

Here, the real number $\mu > 0$ is the Poisson baseline. Also, the nonnegative function ϕ defined on $\mathbb{R}^+$ is the 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.

Targeting the presence of duration clustering over time, Hawkes processes are widely used in financial econometrics (see Yu (2004), Bowsher (2007), Embrechts et al. (2011), Chen and Hall (2013), Aït-Sahalia et al. (2014), Clinet and Potiron (2018b) and Corradi et al. (2020)). There are also applications in finance (see Large (2007), Aït-Sahalia et al. (2015) and Fulop et al. (2015)) and quantitative finance (see Chavez-Demoulin et al. (2005), Bacry et al. (2013), Jaisson and Rosenbaum (2015) and Morariu-Patrichi and Pakkanen (2022)). More recently, Cavaliere et al. (2023) develop a bootstrap approach; Clements et al. (2023) study nonparametric estimation; Christensen and Kolokolov (2024) propose an unbounded intensity model for more general point processes; Potiron and Volkov (2025) consider estimation of latency.[1]

Empirical evidence suggests that the baseline is time-varying during intraday trading activity. Chen and Hall (2013) show in their empirical study (Section 5.2, pp. 7–10) that goodness-of-fit results are in favor of their time-varying baseline model over a set 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 (2018b). Moreover, the empirical findings in the two aforementioned papers suggest that there may be a remaining stochastic effect. Finally, Rambaldi et al. (2015) and Rambaldi et al. (2018) document that there are frequent intensity bursts in the baseline. In the former paper, the authors focus on a short period around the macroeconomic news announcement and are able to capture the increase in trading activity after the news, both when the news has a significant effect on volatility and when this effect is negligible. In the latter paper, the authors develop a general method to detect intensity bursts and demonstrate empirically that only a relatively small proportion of these bursts can be related to news. The intensity remains locally bounded in this case. Finally, Christensen and Kolokolov (2024) propose a framework for intensity burst detection in which the intensity is potentially unbounded over short time intervals.

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

$$\mu_t = \mu_0 + \int_0^t b_s \, ds + \int_0^t \sigma_s \, dW_s + (\delta \mathbf{1}_{\{|\delta| \leq 1\}}) \star (\underline{\mu} - \underline{\nu})_t + (\delta \mathbf{1}_{\{|\delta| > 1\}}) \star \underline{\mu}_t. \tag{1.2}$$

[1] See also the working papers Erdemlioglu et al. (2025b), Erdemlioglu et al. (2025a), Potiron et al. (2025b), Potiron et al. (2025a). Finally, Potiron (2025b) and Potiron (2025c) consider parametric estimation of Hawkes processes, while Potiron (2025a) studies nonparametric estimation in theoretical econometrics.

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 locally bounded bursts. The reason why we rely on an Itô semimartingale rather than a more general process is due to the presence of a Hawkes component in the intensity (1.1). Namely, estimation of the integrated baseline is based on estimation of the integrated volatility of the baseline process. Then, the Itô semimartingale baseline (1.2) is the most general model where estimation of integrated volatility is tractable. It is a restricted semimartingale, and the semimartingale is known to be the most general class (see Jacod and Shiryaev (2013)). Potiron (2025a) consider a more general baseline, but for which estimation of the integrated baseline, our object of interest, is not feasible.

We introduce three measures for the intensity of Hawkes processes. We start from the integrated intensity measure

$$\Lambda_T = \int_0^T \lambda_t \, dt. \tag{1.3}$$

This measure is natural since the intensity of a quote plays an inverse role to the volatility of an asset price. It is well known that the volatility increase is associated with the increase in the number of quotes. Originally, this measure was introduced in risk management, where the integrated intensity can be interpreted as the arrival rate in a queuing system. As far as the authors know, it has not been studied in econometrics. Although the integrated intensity seems to be the most relevant quantity for applications, it cannot be estimated in the presence of a Hawkes component in the intensity due to the presence of duration clustering.

This motivates introduction of a second measure, i.e., the integrated baseline

$$B_T = \int_0^T \mu_t \, dt. \tag{1.4}$$

Compared to the integrated intensity measure (1.3), the Hawkes component of the intensity is removed. Thus, the integrated baseline measure is smoother, and more importantly, we can estimate it even in the presence of a Hawkes component in the intensity. This measure was introduced in Clinet and Potiron (2018b). See also Erdemlioglu et al. (2025b) and Potiron (2025a).

Since the estimation of the integrated baseline is based on the estimation of the integrated volatility of the baseline process, we also introduce another new measure, i.e., the integrated volatility of the baseline

$$\text{IV} = \int_0^T \sigma_t^2 \, dt. \tag{1.5}$$

Just like the volatility of an asset price, the volatility of the baseline can be seen as a measure of risk related to the quotes. By selecting this quantity, we also provide a semiparametric measure of the integrated variance.

Our limit theory builds on in-fill asymptotics when the horizon time T is fixed and the number of observations on $[0, T]$ increases as $n \to \infty$. This approach is motivated by financial applications based on high-frequency data (see Aït-Sahalia and Jacod (2014)). Boosting the intensity, i.e., the use of in-fill asymptotics, boils down to assuming that the intensity is directly proportional to the inverse of the time interval. In other words, the probability of having a quote is proportional to the inverse of the time interval. This seems natural when we consider a self-exciting process with a time-varying intensity applied to quote events. Financial markets reveal submitting many quotes in short time intervals due to a high baseline amplified by the self-excitation and phases with a lower baseline with fewer quotes in short time intervals. These phases create randomness, which explains why we look at a random baseline driven by an Itô semimartingale and not at a purely deterministic time-varying baseline.

Some work in statistics and econometrics accommodates in-fill asymptotics in Hawkes processes. In-fill asymptotic results in Chen and Hall (2013) are based on random observation times of order n. A single boosting of the baseline, i.e., $\lambda_t = \alpha_n \mu_t + \int_0^t \phi(t-s) dN_s$, is considered, where $\alpha_n \to \infty$ is a scaling sequence. Clinet and Potiron (2018b) introduce a joint boosting of the baseline and the kernel, i.e., $\lambda(t) = n\mu_t + \int_0^t n a_s \exp(-n b_s (t-s)) dN_s$. Kwan et al. (2023) revisit Chen and Hall (2013) with the same asymptotic setup as in Clinet and Potiron (2018b). Christensen and Kolokolov (2024), Potiron and Volkov (2025), Erdemlioglu et al. (2025b), Erdemlioglu et al. (2025a), Potiron et al. (2025b) and Potiron et al. (2025a) also use the in-fill asymptotic approach.

The most relevant paper in the literature on Hawkes processes with time-varying baseline is Clinet and Potiron (2018b), who consider a general stochastic baseline and time-varying kernel parameters in the exponential kernel case. Clinet and Potiron (2018b) propose estimation of the integrated baseline based on local parametric estimation. Erdemlioglu et al. (2025a) extend the framework to the case of a generalized gamma kernel. Chen and Hall (2013) and Kwan et al. (2023) allow for a nonrandom parametric baseline, relying on parametric estimation from the starting time to the horizon time. Cai et al. (2024) (Section 4.4) propose a test for constant baseline with high dimensional nonlinear Hawkes processes. However, the conditions on the parametric kernel are stronger since they use least squares estimation. Moreover, they also consider a nonrandom baseline, under specific conditions. Finally, Fang et al. (2024) consider group network Hawkes processes with a time-varying baseline.

Our main result (see Theorem 1) is a joint CLT of suitably rescaled empirical average and preaveraging of local Poisson estimates. This contribution is different from the aforementioned papers on Hawkes processes with a time-varying baseline. The key ingredient in deriving our CLT is the use of the martingale representation of the intensity. This is based on the convolution of the resolvent kernel and the martingale. This extends Theorem 2 in Bacry et al. (2013), which considers an invariant baseline and asymptotics when the final time $T \to \infty$, to the time-varying baseline and in-fill asymptotics. In particular, we study the estimation of the integrated baseline, which was not considered in Bacry et al. (2013). Thus, we obtain convergence of a three-dimensional process rather than a simpler convergence of a univariate process. Finally, the functional form of Theorem 1 makes the obtained result very general. In particular, the tightness condition for the triangular array of martingale increments is fully addressed in the proofs.

Furthermore, we provide five main corollaries: (i) estimation of the integrated intensity, (ii) estimation of the integrated baseline and (iii) estimation of the integrated volatility of the baseline. Then, we consider: (iv) hypothesis testing for the absence of a Hawkes component and (v) hypothesis testing for constant baseline. We obtain novel results (see Corollary 1) for (i) since the problem has not been studied in econometrics. Our result (ii) (see Corollary 2) complements Theorem 5.4 in Clinet and Potiron (2018b). Our result (iii) (see Corollary 3) extends Stoltenberg et al. (2022), who consider the estimation of covariation between the volatility of a price process and its intensity in the absence of Hawkes processes. Our strategy for (iv) (see Corollary 4) differs from Dachian and Kutoyants (2006), who consider parametric and nonparametric composite alternatives with asymptotics when the final time $T \to \infty$. Finally, the result (v) (see Corollary 5) complements Theorem 6 and Proposition 1 in Cai et al. (2024).

In our empirical study (see Section 6) on quote data of the E-mini S&P 500 futures contracts, we show the presence of a Hawkes component capturing clusters of trading and the baseline being time-varying. Furthermore, we find the intraday pattern to be U-shaped. We emphasize that the proposed tests are powerful even in cases of other intraday trading patterns. We demonstrate, based on sub-sample analysis, that both J-shaped and reverse J-shaped trading patterns can be confirmed using our tests. This highlights the potential for empirical applications of our methodology to other data sets.

A simulation study (see Section 9) verifies our CLT results, showing that the integrated intensity, baseline, and volatility have feasible and unfeasible moments converging to their theoretical counterparts in large samples. Hypothesis testing results show adequate size and power for the two proposed tests for the absence of a Hawkes component and constancy of baseline at different levels. We also confirm that the number of observations showing stable results in simulations is compatible with our empirical study.

The remainder of this paper is organized as follows. We provide the setup in Section 2, and we introduce the estimation strategy in Section 3. Our main CLT result is reported in Section 4. We investigate estimation problems (i)-(ii)-(iii) and testing problems (iv)-(v) in Section 5. In Section 6, an empirical application on high-frequency quote data of the E-mini S&P 500 futures contracts is presented. Section 7 concludes. All the proofs are gathered in the supplementary materials, Section 8. We also provide in Section 9 a finite sample analysis, which corroborates the asymptotic theory and belongs to the supplementary material.

2 Setup

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

We start with an introduction to the point process setup. We define N_t as a simple point process on the time interval $[0,T]$, i.e., a family $\{N(C)\}_{C\in\mathcal{B}([0,T])}$ of random variables with values in $\mathbb{N}$. Here, $\mathcal{B}([0,T])$ is the Borel σ-algebra on the compact space $[0,T]$, $N(C)=\sum_{i\in\mathbb{N}}\mathbf{1}_C(T_i)$, and $\{T_i\}_{i\in\mathbb{N}}$ is a sequence of $\mathbb{R}^+$-valued random event times such that, a.s. $T_0=0<T_1<\ldots<T_{N_T}<T<T_{N_T+1}$. The point process N_t counts the cumulative number of quotes on the time interval $[0,t]$. Ee denote by $\mathcal{B}=(\Omega,\mathcal{F},\{\mathcal{F}_t\}_{t\in[0,T]},\mathbb{P})$ a filtered probability space which satisfies the usual conditions. For any time $t\in[0,T]$, we denote the filtration generated by some stochastic process X as $\mathcal{F}_t^X=\sigma\{X_s:0\leq s\leq t\}$. We assume that, for any time $t\in[0,T]$, 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., $\lambda_t=\lim_{u\to 0}\mathbb{E}[\frac{N_{t+u}-N_t}{u}\mid\mathcal{F}_t]$ a.s.. For background on point processes, the reader can consult Daley and Vere-Jones (2003, 2008) and Jacod and Shiryaev (2013).

In this paper, we consider Hawkes processes with an Itô semimartingale baseline. Namely, we are concerned with 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, the nonnegative function ϕ defined on $\mathbb{R}^+$ is the exciting kernel. Also, we have that μ_t is the $\widetilde{\mathcal{F}}_t$ Itô semimartingale baseline process with $\widetilde{\mathcal{F}}_t\subset\mathcal{F}_t$. Since the baseline μ_t is 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

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

Here, we have that the baseline starting variable μ_0 is $\mathcal{F}_0$ measurable. Also, the drift b_t is an $\mathbb{R}$-valued predictable process and the volatility σ_t is an $\mathbb{R}^+$-valued predictable process on $(\Omega, \mathcal{F}, \{\widetilde{\mathcal{F}}_t\}_{t\in[0,T]}, \mathbb{P})$ such that both integrals defined in Equation (2.2) are well-defined for any time $t \in [0,T]$. Moreover, δ is an $\mathbb{R}$-valued predictable function on $\Omega \times \mathbb{R}_+ \times E$ such that both integrals in Equation (2.2) are well-defined for any time $t \in [0,T]$. 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).

As mentioned earlier, the reason why we rely on an Itô semimartingale is due to the presence of a Hawkes component in the intensity (1.1). Namely, estimation of the integrated baseline is based on estimation of the integrated volatility of the baseline process. Then, the Itô semimartingale baseline (1.2) is the most general model where estimation of integrated volatility is tractable. It is a slightly restricted semimartingale, and the semimartingale is known to be the most general class with stochastic integration (see Jacod and Shiryaev (2013)).

Although the introduced baseline process is general, the baseline has to stay positive for its own existence. However, we can ensure positiveness by choosing a baseline equal to the exponential of an Itô semimartingale or any other positive function. It results from the function of an Itô semimartingale being an Itô semimartingale by the Itô's lemma for a regular enough function. Nonetheless, we will introduce weaker conditions that cover a larger span of processes. Namely, we will have some restrictions for the existence of the baseline and the CLT. In particular, the baseline can still exhibit positive and negative jumps, as long as the baseline stays positive. In practice, this mainly inhibits the size of negative jumps. Moreover, we find that the baseline mean is much bigger than its volatility in the empirical study. Thus, the positive baseline is not very restrictive for its continuous component, and we can safely use it in our numerical study.

3 Estimation

In this section, we introduce the in-fill asymptotics. We also introduce empirical average and preaveraging of local Poisson estimates constructed from high-frequency quote times.

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 Equation (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 (2018b), Kwan et al. (2023) and Potiron and Volkov (2025), also based on joint boosting, by not imposing an exponential or a mixture of generalized gamma kernels. Christensen and Kolokolov (2024) consider boosting the baseline to detect intensity bursts without a Hawkes component. They are different from Chen and Hall (2013) in-fill asymptotics, which do not include boosting of the kernel. Boosting the intensity, i.e., the use of in-fill asymptotics, boils down to assuming that the intensity is directly proportional to the inverse of the time interval.

First, we propose an estimator for the empirical average constructed from high-frequency quote times for any time $t \in [0, T]$ as

$$\widehat{\text{Mean}}_t = N_t. \tag{3.2}$$

This natural estimator for the empirical average is novel in the literature of Hawkes processes in econometrics. The estimator is different from the local parametric estimator used in Clinet and Potiron (2018b), which was designed for exponential kernels and is biased when the kernel is not exponential.

Since the intensity explodes asymptotically, the estimator also diverges to infinity. More specifically, we define the diverging mean target for any time $t \in [0, T]$ as

$$\text{Mean}_t = n \frac{1}{1 - \|\phi\|_1} \int_0^t \mu_s \, ds. \tag{3.3}$$

Here, $\|f\|_1 = \int_0^\infty f(t)dt$ denotes the L^1 norm of any function f. To deduce the integrated baseline from the mean target, we first have to retrieve $\|\phi\|_1$, which is unknown to the econometrician. Thus, we will rely on a more complicated estimation procedure in what follows.

More specifically, we introduce $M = \lfloor T/\Delta_n \rfloor$ intervals with equal length Δ_n such that $\bigcup_{i=1}^M [(i-1)\Delta_n, i\Delta_n) \subset [0, T)$ for a finite horizon $T > 0$. Here, $\lfloor \cdot \rfloor$ denotes the floor function. For any index $i = 1, \dots, M$, we define an estimator for local Poisson estimates on the ith interval $[(i-1)\Delta_n, i\Delta_n)$ as

$$\widehat{\lambda}_i = \frac{1}{\Delta_n} (N_{i\Delta_n -} - N_{(i-1)\Delta_n}). \tag{3.4}$$

Then, we consider preaveraging of local Poisson estimates constructed from high-frequency quote times for any time $t \in [0, T]$ as

$$\widehat{\text{Var}}_{t,1} = \sum_{i=2}^{\lfloor t/\Delta_n \rfloor} \big(\Delta_i \widehat{\lambda}\big)^2 \mathbf{1}_{\{|\Delta_i \widehat{\lambda}| \leq \alpha \Delta_n^{-\overline{\omega}}\}}, \tag{3.5}$$

$$\widehat{\text{Var}}_{t,2} = \sum_{i=2}^{\lfloor t/(2\Delta_n) \rfloor} \left(\frac{\Delta_{2i-2}\widehat{\lambda} + \Delta_{2i-1}\widehat{\lambda}}{2} \right)^2 \mathbf{1}_{\{|(\Delta_{2i-1}\widehat{\lambda} + \Delta_{2i}\widehat{\lambda})/2| \leq \alpha \Delta_n^{-\overline{\omega}}\}}. \tag{3.6}$$

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. The two variance estimators with different scales are requested for the estimation of the integrated baseline. The preaveraging method is usually used to mitigate the effect of the market microstructure noise for the estimation of the integrated volatility. In our case, we rely on preaveraging to deal with the self-excitation of the Hawkes process. As far as the authors know, the preaveraging method has not been used for estimation with Hawkes processes. In addition, we consider a truncation in our variance estimators since they would be contaminated by the jumps otherwise.

Then, we define the diverging target values for any time $t \in [0,T]$ as

$$\mathrm{Var}_{t,1} = n^2 \frac{1}{(1-\|\phi\|_1)^2} \int_0^t \left(\frac{2}{3}\sigma_s^2 + \frac{1}{1-\|\phi\|_1} \frac{2}{c}\mu_s \right) ds, \tag{3.7}$$

$$\mathrm{Var}_{t,2} = n^2 \frac{1}{(1-\|\phi\|_1)^2} \int_0^t \left(\frac{2}{3}\sigma_s^2 + \frac{1}{1-\|\phi\|_1} \frac{1}{2c}\mu_s \right) ds. \tag{3.8}$$

To deduce the existence and the form of the target values, we assume that there exists a positive real number $c > 0$ such that $n\Delta_n^2 \xrightarrow{\mathbb{P}} c$. In addition, the order of observation number n is unknown in practice. Thus, the length of intervals Δ_n cannot be chosen directly. Instead, we can estimate it as

$$\widehat{\Delta}_n = \frac{\sqrt{cT}}{\sqrt{N_T}}. \tag{3.9}$$

We use $c = 0.5$, which performs best in our simulation study (see Section 9). Unfortunately, we insist on this estimator of the length of intervals being inconsistent. The reason is that $n^{-1}N_T \xrightarrow{\mathbb{P}} \frac{1}{1-\|\phi\|_1}\int_0^T \mu_s \, ds$, and the limit on the right side of the equation is unknown in practice. However, we can show that $n\widehat{\Delta}_n^2 \xrightarrow{\mathbb{P}} cT(1-\|\phi\|_1)(\int_0^T \mu_s \, ds)^{-1}$. Thus, we choose to derive the theory of this paper on the length of intervals Δ_n.

Our target in this paper is the vector $(\mathrm{Mean_t}, \mathrm{Var}_{t,1}, \mathrm{Var}_{t,2})$ for any time $t \in [0,T]$. We define its non-diverging error, which is also standardized by its convergence rate, as

$$X_t = \begin{pmatrix} \Delta_n^{-1} n^{-1} \big(\widehat{\mathrm{Mean}}_t - \mathrm{Mean}_t\big) \\ \Delta_n^{-\frac{1}{2}} n^{-2} \big(\widehat{\mathrm{Var}}_{t,1} - \mathrm{Var}_{t,1}\big) \\ \Delta_n^{-\frac{1}{2}} n^{-2} \big(\widehat{\mathrm{Var}}_{t,2} - \mathrm{Var}_{t,2}\big) \end{pmatrix}.$$

For any time $t \in [0,T]$, we define the rescaled drift $\breve{\vartheta}_t$ as $\breve{\vartheta}_t = \frac{\mu_t}{c^2(1-\|\phi\|_1)^3}$, the rescaled volatility $\breve{\sigma}_t$ as $\breve{\sigma}_t = \frac{\sigma_t}{1-\|\phi\|_1}$, and the asymptotic variance process w_t as

$$w_t w_t^T = \begin{pmatrix} \frac{\mu_t}{(1-\|\phi\|_1)^3} & 0 & 0 \\ 0 & \breve{\sigma}_t^4 + 4\breve{\sigma}_t^2\breve{\vartheta}_t + 12\breve{\vartheta}_t^2 & \frac{29}{24}\breve{\sigma}_t^4 + \frac{3}{2}\breve{\sigma}_t^2\breve{\vartheta}_t + \frac{3}{2}\breve{\vartheta}_t^2 \\ 0 & \frac{29}{24}\breve{\sigma}_t^4 + \frac{3}{2}\breve{\sigma}_t^2\breve{\vartheta}_t + \frac{3}{2}\breve{\vartheta}_t^2 & 2\breve{\sigma}_t^4 + 2\breve{\sigma}_t^2\breve{\vartheta}_t + \frac{3}{2}\breve{\vartheta}_t^2 \end{pmatrix}. \tag{3.10}$$

We also define the non-diverging asymptotic variance as $\Sigma_t = \int_0^t w_s w_s^T ds$. In addition, we define the diverging asymptotic variance for any time $t \in [0,T]$ as

$$\Sigma_{t,d} = \begin{pmatrix} n\Sigma_{t,11} & 0 & 0 \\ 0 & n^2\Sigma_{t,22} & n^2\Sigma_{t,23} \\ 0 & n^2\Sigma_{t,23} & n^2\Sigma_{t,33} \end{pmatrix}.$$

We define its estimator for any time $t \in [0, T]$ as

$$\widehat{\Sigma}_{t,d} = \begin{pmatrix} \widehat{\Sigma}_{t,d,11} & 0 & 0 \\ 0 & \widehat{\Sigma}_{t,d,22} & \widehat{\Sigma}_{t,d,23} \\ 0 & \widehat{\Sigma}_{t,d,23} & \widehat{\Sigma}_{t,d,33} \end{pmatrix}. \tag{3.11}$$

Here, the components of the matrix are defined as $\widehat{\Sigma}_{t,d,11} = \widehat{\text{Mean}}_t$, $\widehat{\Sigma}_{t,d,22} = \frac{3}{4}\widehat{\kappa}_{t,4,1} - 3\widehat{\eta}_t\widehat{\kappa}_{t,3,1} + 9\widehat{\eta}_t^2\widehat{\kappa}_{t,2,1}$, $\widehat{\Sigma}_{t,d,23} = \frac{29}{32}\widehat{\kappa}_{t,4,1} - \frac{69}{8}\widehat{\eta}_t\widehat{\kappa}_{t,3,1} + \frac{63}{8}\widehat{\eta}_t^2\widehat{\kappa}_{t,2,1}$ and $\widehat{\Sigma}_{t,d,33} = \frac{3}{2}\widehat{\kappa}_{t,4,2} - 6\widehat{\eta}_t\widehat{\kappa}_{t,3,2} + 18\widehat{\eta}_t^2\widehat{\kappa}_{t,2,2}$. We also define $\widehat{\kappa}_{t,2,1} = \Delta_n^{-3}\sum_{i=2}^{\lfloor t/\Delta_n \rfloor} \widehat{\lambda}_i^2 \mathbf{1}_{\{|\Delta_i\widehat{\lambda}| \leq n\varpi_i\}}$ and $\widehat{\kappa}_{t,2,2} = \Delta_n^{-3}\sum_{i=2}^{\lfloor t/(2\Delta_n) \rfloor} (\frac{\widehat{\lambda}_{i-1}+\widehat{\lambda}_i}{2})^2 \mathbf{1}_{\{|(\Delta_{2i-1}\widehat{\lambda}+\Delta_{2i}\widehat{\lambda})/2| \leq n\varpi_i\}}$. Additionally, we define $\widehat{\kappa}_{t,3,1} = \Delta_n^{-2}\sum_{i=2}^{\lfloor t/\Delta_n \rfloor} \widehat{\lambda}_i(\Delta_i\widehat{\lambda})^2 \mathbf{1}_{\{|\Delta_i\widehat{\lambda}| \leq n\varpi_i\}}$, and

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

Let $\widehat{\kappa}_{4,1} = \Delta_n^{-1}\sum_{i=1}^{\lfloor T/\Delta_n \rfloor}(\Delta_i\widehat{\lambda})^4 \mathbf{1}_{\{|\Delta_i\widehat{\lambda}| \leq n\varpi_i\}}$, $\widehat{\kappa}_{4,2} = \Delta_n^{-1}\sum_{i=1}^{\lfloor T/(2\Delta_n) \rfloor}(\frac{\Delta_{2i-2}\widehat{\lambda}+\Delta_{2i-1}\widehat{\lambda}}{2})^4 \mathbf{1}_{\{|(\Delta_{2i-1}\widehat{\lambda}+\Delta_{2i}\widehat{\lambda})/2| \leq n\varpi_i\}}$, while $\widehat{\eta} = \frac{2}{3}\frac{\Delta_n^2(\widehat{\text{Var}}_1 - \widehat{\text{Var}}_2)}{\widehat{\text{Mean}}}$. Finally, we define the estimator of the non-diverging asymptotic variance as

$$\widehat{\Sigma}_t = \begin{pmatrix} \frac{\Delta_n^2}{c}\widehat{\Sigma}_{t,d,11} & 0 & 0 \\ 0 & \frac{\Delta_n^4}{c^2}\widehat{\Sigma}_{t,d,22} & \frac{\Delta_n^4}{c^2}\widehat{\Sigma}_{t,d,23} \\ 0 & \frac{\Delta_n^4}{c^2}\widehat{\Sigma}_{t,d,23} & \frac{\Delta_n^4}{c^2}\widehat{\Sigma}_{t,d,33} \end{pmatrix}. \tag{3.12}$$

4 Theory

In this section, our main result characterizes feasible statistics induced by central limit theory for empirical average and preaveraging of local Poisson estimates.

As the Itô semimartingale is not bounded below, it can become nonpositive. This is incompatible with the constraint of a positive baseline for getting a well-defined Hawkes process. Thus, we restrict the class of Itô semimartingales considered. As already mentioned, we can ensure positiveness by choosing a baseline equal to the exponential of an Itô semimartingale or any other positive function. Thus, we introduce a set of conditions required for the existence of Hawkes processes with a time-varying baseline driven by an Itô semimartingale.

Condition 1. (a) The baseline is a.s. positive on $[0, T]$ a.e., i.e., $\mathbb{P}(\mu_t > 0 \forall t \in [0, T] \text{ a.e.}) = 1$.

(b) The baseline is integrable on the time interval $[0, T]$ a.s., i.e., $\mathbb{P}(\int_0^T \mu_s\, ds < \infty) = 1$.

(c) For any time $t \in [0,T]$, 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., $N_t = \int_{[0,t]\times\mathbb{R}} \mathbf{1}_{[0,\lambda_s]}(x)\underline{N}(ds \times dx)$.

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

Condition 1 (a) implies that the baseline of the point process is positive. Thus, the point process is well-defined. Moreover, Condition 1 (a) is weaker than the condition from Clinet and Potiron (2018b), who require that the baseline belongs to a compact space. Condition 1 (b) is already a condition in the simpler case of heterogeneous Poisson processes without a self-exciting kernel (see Daley and Vere-Jones (2003)), and is also required to establish existence in Clinet and Potiron (2018b) (see Assumption E (ii), p. 3476). Condition 1 (c) is a technical condition which corresponds to Poisson imbedding (see Brémaud and Massoulié (1996), Section 3, pp. 1571-1572). In particular, it assumes independence between the two filtrations $\widetilde{\mathcal{F}}_t$ and $\mathcal{F}_t^{\underline{N}}$. It is already required in Clinet and Potiron (2018b) (see the last sentence before Theorem 5.1, p. 3476). The stochastic process N_t is constructed as the point process counting the points of the Poisson process $\underline{N}$ below the curve $t \to \lambda_t$. Finally, Condition 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)).

We now provide a new existence result to the literature. It is obtained by extending the machinery of Poisson imbedding for classical Hawkes processes (see Brémaud and Massoulié (1996)) to the Itô semimartingale baseline case. It complements Theorem 5.1 (p. 3476) in Clinet and Potiron (2018b), in which the kernel is exponential and Theorem 1 (p. 99) in Cai et al. (2024) in which the form of the kernel is also restricted. See also Proposition 1 in Potiron (2025a) and Proposition 1 in Erdemlioglu et al. (2025b)

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

We define an alternative drift as $b_t^\delta = b_t - \int_E \delta(t,z)\mathbf{1}_{\{|\delta(t,z)|\leq 1\}}F_t(dz)$ for any time $t \in [0,T]$. This alternative drift depends on the compensated measure of the jump process. Moreover, we define $V_a^b(f)$ as the total variation of the function f from the starting time a to the final time b. We introduce a set of conditions required to derive the CLT for the empirical average and preaveraging of local Poisson estimates.

Condition 2. (a) The kernel satisfies the short-range condition, i.e., $\int_0^\infty t\phi(t)dt < \infty$.

(b) For any $k \in \mathbb{N}$ with $k \geq 2$, the L^1 norm of ϕ^k is finite, i.e., $\|\phi^k\|_1 < \infty$.

(c) There exists a positive real number $c > 0$ such that $n\Delta_n^2 \xrightarrow{\mathbb{P}} c$.

(d) There exists a real number $\beta \in [0, 1)$ such that $\sup_{0\leq t\leq T} \int \min(|x|^\beta, 1)F_t(dx)$ is a.s. finite.

(e) The truncation level satisfies $\varpi \in (0, 1/(8 - 4\beta))$.

(f) For any positive integer $k \in \mathbb{N}_*$ and any time $t \in [0, T]$, we have $\mathbb{E}\big[|b_t|^k\big] < \infty$ and $\mathbb{E}[\sigma_t^k] < \infty$.

(g) We have that $\mathbb{E}\big[\exp\big(\frac{1}{2}\int_0^T \frac{(b_s^\delta)^2}{\sigma_s^2}ds\big)\big] < \infty$.

(h) The volatility process is a semimartingale, i.e., $\sigma_t^2 = A_t + M_t^{(\sigma)}$ in which A_t is an $\mathcal{F}_t$ 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 positive real number $\gamma > 0$ and any positive integer $k \in \mathbb{N}_*$.

Condition 2 (a) imposes restrictions on the kernel shape $\phi(t)$. Namely, the kernel has to decrease fast enough to 0. This corresponds exactly to Assumption (A2) in Bacry et al. (2013) (p. 2480). This is required to obtain the martingale form of the intensity process. Condition 2 (a) is also used in Jaisson and Rosenbaum (2015). Condition 2 (b) also imposes restrictions on the kernel shape $\phi(t)$. Namely, the L^1 norm of any power of the kernel has to be finite. Condition 2 (c) is used to deduce the existence and the form of the variance target values Var_1 and Var_2. This condition is commonly used in the proofs based on local estimation. This arises in the estimation of conditional means and variances, proving the convergence of martingale brackets and controlling fourth moments in the Lindeberg condition.

Condition 2 (d) and Condition 2 (e) are due to the presence of jumps in the Itô semimartingale baseline. Condition 2 (d) is slightly more restrictive than Assumption 2 in Jing et al. (2014). Condition 2 (e) corresponds exactly to the condition used in Theorem 3 from Jing et al. (2014). The exponent β in Condition 2 (d) is an upper bound on the generalized Blumenthal-Getoor index, an adaptation from Lévy processes to the semimartingale setting (see Lemma 3.2.1 in Jacod and Protter (2012)). The Blumenthal-Getoor index of a Lévy process is defined as $\beta = \inf \left\{ r \geq 0 \ : \ \int_{|x|<1} |x|^r \nu(\mathrm{d}x) < \infty \right\}$, where ν is the Lévy measure. We can interpret the index β as the smallest power at which the distribution of the small jumps has a finite βth moment. Thus, it is related to the jump activity of the baseline process, with lower values of the index β being more binding. The index $\beta = 1$ separates jump processes with sample paths of finite and infinite variation, i.e., it concerns their absolute summability, while $r = 0$ implies that the jumps are finitely active on finite time intervals a.s.. Moreover, the index β plays a crucial role in determining rate conditions for the tuning parameters in our estimation procedure and in establishing the rate of convergence of our estimator. In particular, we do not allow for highly irregular baselines such as Levy-driven components with infinite activity.

Condition 2 (f) requires the boundedness for any moment of the drift process and the volatility process. This is used in the proof of the CLT. Condition 2(g) is required to apply the Girsanov theorem in our proofs. More specifically, this corresponds to Novikov condition (see Novikov (1972)). We apply the Girsanov theorem after removing the jumps from the baseline. The use of the Girsanov theorem in high-frequency econometrics to remove the drift from the Itô semimartingale is common (see Section 2.2 in Mykland and Zhang (2009), Potiron and Mykland (2017), Clinet and Potiron (2018a) and Clinet and Potiron (2019)). In these papers, the authors restrict to a continuous Itô semimartingale. They further assume that the volatility process is itself a continuous Itô semimartingale bounded away from 0 uniformly over time a.s. and that the drift process is locally bounded. In particular, the conditions on the volatility process and the drift process to apply the Girsanov theorem are not very sharp. On the contrary, we propose sharper conditions on the volatility process and the drift process. In addition, we consider the condition on the alternative drift b_s^δ due to the presence of jumps in the Itô semimartingale. Finally, Condition 2(h) restricts to a semimartingale volatility process, and this is used in the proof of the CLT.

In Theorem 1, we provide our main result, which is a joint CLT of suitably the rescaled empirical average and preaveraging of local Poisson estimates. This contribution is orthogonal to the aforementioned papers on Hawkes processes with time-varying baselines. The key ingredient in deriving our CLT is the use of the martingale representation of the intensity. This is based on the convolution of the resolvent kernel and the martingale. This extends Theorem 2 in Bacry et al. (2013), which considers an invariant baseline and asymptotics when the final time $T \to \infty$, to the time-varying baseline and in-fill asymptotics. In particular, we investigate the estimation of the integrated baseline, which was not studied in Bacry et al. (2013). Thus, we obtain convergence of a three-dimensional process rather than the more simple convergence of a unidimensional process.

We denote the collection of cadlag functions starting from the space $[0,T]$ to the space $\mathbb{R}^k$ by $\mathbb{D}([0,T],\mathbb{R}^k)$ for any positive integer k. Moreover, we denote $\overset{\mathcal{D}-s}{\longrightarrow}$ as the $\mathcal{F}_t$ stable weak convergence for the Skorokhod topology on the Skorokhod space $\mathbb{D}([0,T],\mathbb{R}^k)$. Finally, we define the convergence in probability uniformly for any time $t \in [0,T]$ as *u.c.p.*, i.e., $X_n \overset{u.c.p.}{\to} X$ if $\sup_{t\in[0,T]} |X_{t,n} - X_t| \overset{\mathbb{P}}{\to} 0$ as $n \to \infty$.

Theorem 1. *We assume that Condition 1 and Condition 2 hold. Then, there is a canonical* 3*-dimensional standard Brownian extension of* $\mathcal{B}$ *with the canonical standard Brownian motion* $\widetilde{W}_t$ *such that as* $n \to \infty$ *we have jointly for any time* $t \in [0,T]$ *that*

$$X_t \overset{\mathcal{D}-s}{\longrightarrow} \int_0^t w_s d\widetilde{W}_s. \tag{4.1}$$

We also have the uniform consistency of the estimator of non-diverging asymptotic variance as $n \to \infty$ *jointly for any time* $t \in [0,T]$*, i.e.,*

$$\widehat{\Sigma}_t \overset{u.c.p.}{\to} \Sigma_t. \tag{4.2}$$

We have the normalized CLT with feasible variance as $n \to \infty$ *jointly for any time* $t \in [0,T]$*, i.e.,*

$$\widehat{\Sigma}^{-1/2} \begin{pmatrix} \frac{\Delta_n}{c}\left(\widehat{\text{Mean}} - \text{Mean}\right)_t \\ \frac{\Delta_n^{7/2}}{c^2}\left(\widehat{\text{Var}_1} - \text{Var}_1\right)_t \\ \frac{\Delta_n^{7/2}}{c^2}\left(\widehat{\text{Var}_2} - \text{Var}_2\right)_t \end{pmatrix} \overset{\mathcal{D}-s}{\longrightarrow} \widetilde{W}_t. \tag{4.3}$$

The convergence rate is Δ_n^{-1} for estimation of the mean, while $\Delta_n^{-\frac{1}{2}}$ for estimation of the variances. By Condition 2(c), these convergence rates are equivalent to $\sqrt{n}$ and $n^{1/4}$. The convergence is slower for estimation of variances since variance estimation is based on mean estimation. Finally, the functional convergence in Theorem 1 implies as $n \to \infty$ and for any time $t \in [0, T]$ that

$$\widehat{\Sigma}^{-1/2} \begin{pmatrix} \frac{\Delta_n}{c}\left(\widehat{\text{Mean}} - \text{Mean}\right) \\ \frac{\Delta_n^{7/2}}{c^2}\left(\widehat{\text{Var}}_1 - \text{Var}_1\right) \\ \frac{\Delta_n^{7/2}}{c^2}\left(\widehat{\text{Var}}_2 - \text{Var}_2\right) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_3(0, 1). \tag{4.4}$$

Here, we define $\mathcal{N}_k(0, 1)$ as a standard normal vector of dimension k.

5 Applications

In this section, we first investigate estimation problems (i), (ii) and (iii), and then testing problems (iv) and (v).

5.1 Estimation of the integrated intensity

We start with the estimation of the diverging integrated intensity for any time $t \in [0, T]$:

$$\Lambda_t = \int_0^t \lambda_s \, ds. \tag{5.1}$$

This measure is natural since the intensity of a quote plays an inverse role to the volatility of an asset price. It is well known that when volatility increases, the number of quotes increases, and vice versa. Originally, this measure was introduced in management science where the integrated intensity can be interpreted as the arrival rate in a queuing system. As far as the authors know, it has not been studied in econometrics. Although the integrated intensity seems to be the most relevant quantity for applications, it cannot be estimated in the presence of a Hawkes component in the intensity. This is explained by the presence of duration clustering in time. Thus, we consider the case where the point process N_t is not a Hawkes process, i.e., $\phi(t) = 0$.

We define the asymptotic variance of the non-diverging mean estimator for any time $t \in [0, T]$:

$$AVar(\widehat{\text{Mean}_t}) = n^{-1} \int_0^t \lambda_s \, ds. \tag{5.2}$$

Moreover, we define the asymptotic variance estimator of $AVar(\widehat{\text{Mean}_t})$ for any time $t \in [0, T]$ as

$$\widehat{AVar}(\widehat{\text{Mean}_t}) = \frac{\Delta_n^2 \widehat{\text{Mean}_t}}{c}. \tag{5.3}$$

The following corollary gives the CLT for the mean estimator. The convergence rate is Δ_n^{-1}, which by Condition 2 (c) is asymptotically equivalent to $\sqrt{n}$. This result is novel to the literature.

Corollary 1. *We assume that Condition 1 and Condition 2 hold. We also assume that $\phi(t) = 0$ for any time $t \in \mathbb{R}^+$. Then, we have as $n \to \infty$ jointly for any time $t \in [0, T]$ that*

$$\frac{\Delta_n^{-1} n^{-1} (\widehat{\text{Mean}_t} - \Lambda_t)}{\sqrt{AVar(\widehat{\text{Mean}_t})}} \xrightarrow{\mathcal{D}-s} \widetilde{W}_t^{(1)}. \tag{5.4}$$

We have the normalized CLT with feasible variance as $n \to \infty$ jointly for any time $t \in [0, T]$, i.e.,

$$\frac{\Delta_n (\widehat{\text{Mean}_t} - \Lambda_t)}{c\sqrt{\widehat{AVar}(\widehat{\text{Mean}_t})}} \xrightarrow{\mathcal{D}-s} \widetilde{W}_t^{(1)}. \tag{5.5}$$

5.2 Estimation of the integrated baseline

We continue with the estimation of the diverging integrated baseline for any time $t \in [0, T]$ as

$$B_t = n \int_0^t \mu_s \, ds. \tag{5.6}$$

This is the most important application from Theorem 1. Compared to the integrated intensity measure (1.3), the Hawkes component of the intensity is removed. Thus, the integrated baseline measure is smoother, and more importantly, we can estimate it even in the presence of a Hawkes component in the intensity. This measure was introduced in Clinet and Potiron (2018b) in the context of parametric Hawkes processes with time-varying parameters and an exponential kernel. See also Erdemlioglu et al. (2025b), who consider an extension with gamma kernels.

To deduce the integrated baseline from the mean target, we first have to retrieve $\|\phi\|_1$ which is unknown to the econometrician. Thus, we first estimate the L^1 norm of the kernel for any time $t \in [0,T]$ as

$$\widehat{\|\phi\|}_{t,1} = 1 - \sqrt{\frac{3\widehat{\text{Mean}}_t}{2\Delta_n^2\big(\widehat{\text{Var}}_1 - \widehat{\text{Var}}_2\big)_t}}. \tag{5.7}$$

In Equation (5.7), we use the estimator from Hardiman and Bouchaud (2014) and we replace their variance by $\frac{2}{3}\big(\widehat{\text{Var}}_1 - \widehat{\text{Var}}_2\big)$ since we have a time-varying baseline in this paper. The study of this estimator with its applications in finance is beyond the scope of this paper and available in Potiron et al. (2025a). Then, we can estimate the diverging integrated baseline for any time $t \in [0,T]$ as

$$\widehat{B}_t = (1 - \widehat{\|\phi\|}_{t,1})\widehat{\text{Mean}}_t. \tag{5.8}$$

In addition, we define the asymptotic variance of $\widehat{\|\phi\|}_{t,1}$ for any time $t \in [0,T]$ as

$$AVar(\widehat{\|\phi\|}_{t,1}) = \nabla f_{t,1}(x_t)^T \Sigma_t \nabla f_{t,1}(x_t). \tag{5.9}$$

Here, we have $f_{t,1}(x_t) = 1 - \sqrt{\frac{3n\text{Mean}_t}{2c(x_{t,2}-x_{t,3})}}$ and $x_t = [0, \text{Var}_{t,1}, \text{Var}_{t,2}]^T$. Moreover, we define the estimator of the asymptotic variance for any time $t \in [0,T]$ as

$$\widehat{AVar}(\widehat{\|\phi\|}_{t,1}) = \nabla f_{t,2}(\widehat{x}_t)^T \widehat{\Sigma}_t \nabla f_{t,2}(\widehat{x}_t). \tag{5.10}$$

Here, we have $f_{t,2}(x) = 1 - \sqrt{\frac{3\widehat{\text{Mean}}_t}{2\Delta_n^2(x_{t,2}-x_{t,3})}}$ and $\widehat{x}_t = [0, \widehat{\text{Var}}_{t,1}, \widehat{\text{Var}}_{t,2}]^T$.

In the following corollary, we give the CLT for the integrated baseline. This complements Theorem 5.4 in Clinet and Potiron (2018b). The convergence rate is $\Delta_n^{-1/2}$, which by Condition 2(c) is asymptotically equivalent to $n^{1/4}$.

Corollary 2. *We assume that Condition 1 and Condition 2 hold. Then, there is a canonical* 1*-dimensional standard Brownian extension of $\mathcal{B}$ with the canonical standard Brownian motion $\widetilde{W}_{B,t}$ satisfying as $n \to \infty$ jointly for any time $t \in [0,T]$ that*

$$\frac{\Delta_n^{-\frac{1}{2}} n^{-1}(\widehat{B}_t - B_t)}{\sqrt{AVar(\widehat{\|\phi\|}_{t,1}) n^{-2}\text{Mean}_t^2}} \xrightarrow{\mathcal{D}-s} \widetilde{W}_{B,t}. \tag{5.11}$$

We have the normalized CLT with feasible variance as $n \to \infty$ jointly for any time $t \in [0, T]$, i.e.,

$$\frac{\Delta_n^{-\frac{1}{2}}(\widehat{B}_t - B_t)}{\sqrt{\widehat{AVar}(\widehat{\|\phi\|}_{t,1})\widehat{\text{Mean}}_t^2}} \xrightarrow{\mathcal{D}-s} \widetilde{W}_{B,t}. \tag{5.12}$$

5.3 Integrated volatility of the baseline

Since the estimation of the integrated baseline is based on the estimation of the integrated volatility of the baseline process, we also introduce another new measure, i.e., the diverging integrated volatility of the baseline defined for any time $t \in [0, T]$ as

$$\text{IV}_t = n^2 \int_0^t \sigma_s^2 \, ds. \tag{5.13}$$

Just like the volatility of an asset price, we can see the volatility of the baseline as a measure of risk related to the quotes. To the best of our knowledge, this is a novel measure in econometrics. The complementing setup is Stoltenberg et al. (2022), who consider estimation of covariances between the volatility of a price process and its intensity in the absence of Hawkes processes.

We estimate the diverging integrated volatility of the baseline for any time $t \in [0, T]$ as

$$\widehat{\text{IV}}_t = (1 - \widehat{\|\phi\|}_{t,1})^2 \big(2\widehat{\text{Var}}_{t,2} - \frac{1}{2}\widehat{\text{Var}}_{t,1}\big). \tag{5.14}$$

Here, the use of two variance estimators with different scales is necessary. We define the asymptotic variance of the integrated volatility estimator for any time $t \in [0, T]$ as

$$AVar(\widehat{\text{IV}}_t) = \nabla g_{t,1}(x_t)^T \Sigma_t \nabla g_{t,1}(x_t). \tag{5.15}$$

Here, we have $g_{t,1}(x) = \big(\frac{3n\text{Mean}_t}{2c(x_2 - x_3)_t}\big)\big(2x_{t,3} - \frac{1}{2}x_{t,2}\big)$. Moreover, we define the asymptotic variance estimator for any time $t \in [0, T]$ as

$$\widehat{AVar}(\widehat{\text{IV}}_t) = c^2 \nabla g_{t,2}(\widehat{x}_t)^T \widehat{\Sigma}_t \nabla g_2(\widehat{x}_t). \tag{5.16}$$

Here, we have $g_{t,2}(x_t) = \big(\frac{3\widehat{\text{Mean}}_t}{2\Delta_n^2(x_2 - x_3)_t}\big)\big(2x_{t,3} - \frac{1}{2}x_{t,2}\big)$.

The following corollary gives the CLT of the integrated volatility of the baseline. The convergence rate is $\Delta_n^{-1/2}$, which by Condition 2(c) is asymptotically equivalent to $n^{1/4}$.

Corollary 3. *We assume that Condition 1 and Condition 2 hold. Then, there is a canonical 1-dimensional standard Brownian extension of $\mathcal{B}$ with the canonical standard Brownian motion $\widetilde{W}_{IV,t}$ satisfying as $n \to \infty$ jointly for any time $t \in [0,T]$ that*

$$\frac{\Delta_n^{-\frac{1}{2}} n^{-2} \big(\widehat{\mathrm{IV}}_t - \mathrm{IV}_t\big)}{\sqrt{AVar(\widehat{\mathrm{IV}}_t)}} \xrightarrow{\mathcal{D}-s} \widetilde{W}_{IV,t}. \tag{5.17}$$

We have the normalized CLT with feasible variance as $n \to \infty$ jointly for any time $t \in [0,T]$, i.e.,

$$\frac{\Delta_n^{7/2} \big(\widehat{\mathrm{IV}}_t - \mathrm{IV}_t\big)}{c^2 \sqrt{\widehat{AVar}(\widehat{\mathrm{IV}}_t)}} \xrightarrow{\mathcal{D}-s} \widetilde{W}_{IV,t}. \tag{5.18}$$

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. A related paper provides 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 nonparametric composite alternatives under asymptotics when the final time $T \to \infty$. Their framework is simpler since they consider a stationary Poisson process with a known intensity under the null hypothesis.

We define the null hypothesis and the alternative hypothesis as

$$H_0 : \{\text{absence of a Hawkes component, i.e., } \|\phi\| = 0\},$$

$$H_1 : \{\text{presence of a Hawkes component, i.e., } \|\phi\| > 0\}.$$

We introduce the test statistic jointly for any time $t \in [0,T]$ as

$$S_t = \frac{\Delta_n^{-1} \widehat{\|\phi\|_{t,1}^2}}{\widehat{AVar}(\widehat{\|\phi\|_{t,1}})}). \tag{5.20}$$

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 4. *We assume that Condition 1 and Condition 2 hold. Then, the test statistic S_t converges in distribution for any time $t \in [0,T]$ 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. More specifically, we have for any $0 < \alpha < 1$ as $n \to \infty$ that*

$$\mathbb{P}(S_t > q(1-\alpha) \mid H_0) \to \alpha, \tag{5.21}$$

$$\mathbb{P}(S_t > q(1-\alpha) \mid H_1) \to 1. \tag{5.22}$$

5.5 Test for constant baseline

Finally, we introduce a test for constant baseline. Cai et al. (2024) (Section 4.4) propose a test for constant baseline with high dimensional nonlinear Hawkes processes. However, the conditions on the parametric kernel are stronger since they use least squares estimation. Moreover, they also consider a nonrandom baseline, with specific conditions. We consider a test based on the Hausman principle (see 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 (see Aït-Sahalia and Xiu (2019), Clinet and Potiron (2019) and Clinet and Potiron (2021)).

We define the null hypothesis and the alternative hypothesis for any time $t \in [0, T]$ as

$$H'_{t,0} : \{\text{The baseline } \mu_s \text{ is constant on } [0,t]\},$$

$$H'_{t,1} : \{\text{The baseline } \mu_s \text{ is not constant on } [0,t]\}.$$

We first define a variance estimator in case of a constant baseline for any time $t \in [0, T]$ as

$$\widehat{\text{Var}}_t = \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \left(\widehat{\lambda}_i - \frac{\widehat{\text{Mean}}_t}{t} \right)^2. \tag{5.25}$$

Following Hardiman and Bouchaud (2014), we then estimate the L^1 norm of the kernel in case of a constant baseline for any time $t \in [0, T]$ as

$$\widehat{\|\phi\|_{t,1}^H} = 1 - \sqrt{\frac{\widehat{\text{Mean}}_t}{\Delta_n^2 \widehat{\text{Var}}_t}}. \tag{5.26}$$

We also define the matrix $\widehat{\Sigma}'_t$ for any time $t \in [0,T]$ as

$$\widehat{\Sigma}'_t = \begin{pmatrix} \frac{\Delta_n^2}{c}\widehat{\text{Mean}}_t & 0 & 0 & 0 \\ 0 & \frac{\Delta_n^4}{c^2}\widehat{\text{Var}}_{t,1} & \frac{2\Delta_n^4}{c^2}\widehat{\text{Var}}_{t,1} & \frac{\Delta_n^4}{2c^2}\widehat{\text{Var}}_{t,1} \\ 0 & \frac{2\Delta_n^4}{c^2}\widehat{\text{Var}}_{t,1} & \frac{6\Delta_n^4}{c^2}\widehat{\text{Var}}_{t,1} & \frac{3\Delta_n^4}{4c^2}\widehat{\text{Var}}_{t,1} \\ 0 & \frac{\Delta_n^4}{2c^2}\widehat{\text{Var}}_{t,1} & \frac{3\Delta_n^4}{4c^2}\widehat{\text{Var}}_{t,1} & \frac{3\Delta_n^4}{4c^2}\widehat{\text{Var}}_{t,1} \end{pmatrix}. \tag{5.27}$$

Finally, we define the asymptotic variance of $\widehat{\|\phi\|_{t,1}} - \widehat{\|\phi\|^H_{t,1}}$ for any time $t \in [0,T]$ as

$$\widehat{AVar}(\widehat{\|\phi\|_{t,1}} - \widehat{\|\phi\|^H_{t,1}}) = \big(\nabla f_t(\widehat{x}_{t,1}) - \nabla g_t(\widehat{x}_{t,2})\big)^T \widehat{\Sigma}'_t\big(\nabla f_t(\widehat{x}_{t,1}) - \nabla g_t(\widehat{x}_{t,2})\big).$$

Here, we have that $f_t(x) = 1 - \sqrt{\frac{3}{2}\frac{\widehat{\text{Mean}}_t}{\Delta_n^2(x_3 - x_4)}}$, $\widehat{x}_{t,1} = \big(0,\ \ 0,\ \ \widehat{\text{Var}}_{t,1},\ \ \widehat{\text{Var}}_{t,2}\big)^T$, $g_t(x) = 1 - \sqrt{\frac{\widehat{\text{Mean}}_t}{\Delta_n^2 x_2}}$ and $\widehat{x}_{t,2} = \big(0,\ \ \widehat{\text{Var}},\ \ 0,\ \ 0\big)^T$. We introduce the test statistic for any time $t \in [0,T]$ as

$$S'_t = \frac{\Delta_n^{-1}(\widehat{\|\phi\|_{t,1}} - \widehat{\|\phi\|^H_{t,1}})^2}{\widehat{AVar}(\widehat{\|\phi\|_{t,1}} - \widehat{\|\phi\|^H_{t,1}})}. \tag{5.28}$$

We first give the following CLT which extends Theorem 1 under the null hypothesis $H'_{t,0}$.

Proposition 2. *We assume that Condition 1 and Condition 2 hold. Under the null hypothesis $H'_{t,0}$ for any time $t \in [0,T]$, we have as $n \to \infty$ that*

$$(\widehat{\Sigma}'_t)^{-\frac{1}{2}} \begin{pmatrix} \Delta_n^{-1}\big(\widehat{\text{Mean}} - \text{Mean}\big)_t \\ \Delta_n^{-\frac{1}{2}}\big(\widehat{\text{Var}} - \text{Var}\big)_t \\ \Delta_n^{-\frac{1}{2}}\big(\widehat{\text{Var}}_1 - \text{Var}_1\big)_t \\ \Delta_n^{-\frac{1}{2}}\big(\widehat{\text{Var}}_2 - \text{Var}_2\big)_t \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_4\left(0, 1\right). \tag{5.29}$$

The following corollary gives the limit theory of the test for constant baseline. This complements Theorem 6 and Proposition 1 in Cai et al. (2024).

Corollary 5. *We assume that Condition 1 and Condition 2 hold. Then, the test statistic S'_t for any time $t \in [0,T]$ converges in distribution to a chi-squared random variable with one degree of freedom under the null hypothesis $H'_{t,0}$ and is consistent under the alternative hypothesis $H'_{t,1}$, i.e., for any $0 < \alpha < 1$ as $n \to \infty$, we have*

$$\mathbb{P}(S'_t > q(1-\alpha) \mid H'_{t,0}) \to \alpha, \tag{5.30}$$

$$\mathbb{P}(S'_t > q(1-\alpha) \mid H'_{t,1}) \to 1. \tag{5.31}$$

6 Empirical application

6.1 Testing

Our empirical application focuses on the S&P 500 E-mini futures, which 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 Tickdata-market database. The data set covers the period from January 2020 to December 2021. All quotes are considered during normal trading hours.

In Figure 1, 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 Equation (9.1) in Section 9. The most pronounced bursts occur at the beginning of the trading session and just before closing.

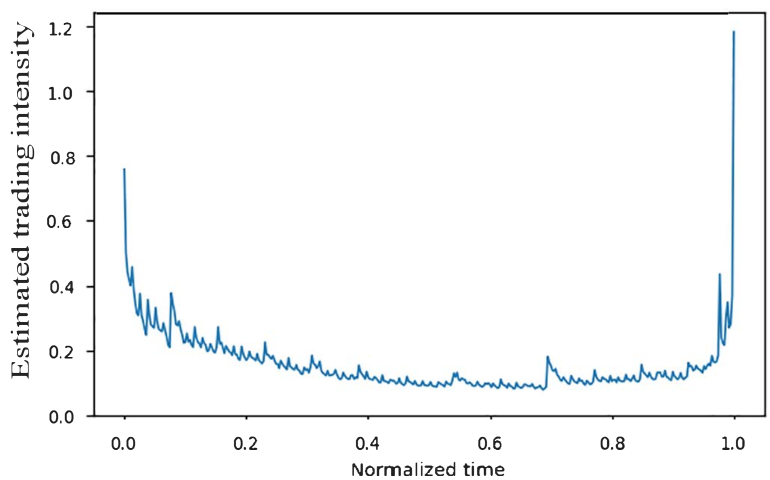

Figure 1: Estimated intensity of the S&P500 E-mini futures quotes. The number of the mid-quote price changes in millions per minute is shown.

Now, we turn to testing the hypotheses formulated in Sections 5.4 and 5.5, namely the absence of a Hawkes component and constant baseline. For each day in the sample, we perform the tests following Corollaries 4 and 5. Figure 2 shows corresponding test statistics revealing rejection of the null hypothesis in both cases. Namely, we confirm the presence of a Hawkes component (blue line) and the time-varying baseline (orange line) for all days at the 5% level.

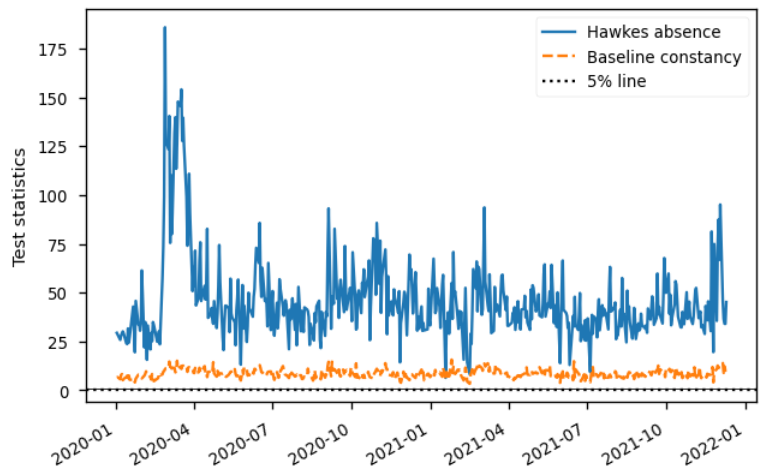

Figure 2: Test statistics for the null hypothesis in the two tests from Sections 5.4 and 5.5 with the 5% critical value. Test statistics are computed for each day in the sample.

To verify that a multiple statistical inference problem does not distort our test results, 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. We have conducted another robustness check of our test results following Bajgrowicz et al. (2016), and the results are in agreement with the Bonferroni corrected tests.

6.2 Sub-sample analysis and robustness

The empirical findings favor Hawkes processes with a time-varying baseline. Now, we check how different types of stock market trading activities affect the result of the time-varying baseline and the presence of a Hawkes component in the data. Indeed, different trading patterns can create various trading strategies and provide different implications for asset pricing. Although the reported U-shaped behavior is identified for our data set, other stock exchanges create different patterns, namely J-shaped, associated with high trading activity earlier in the day and lower activity later in the day (see Huber

(2007)), or reverse J-shaped. In our case the J-shaped pattern can be associated with the trading between 9.30 and 12.45 and the reverse J-shaped pattern starts with trading at 12.45 and finishes at 16.00. We split each day into these two sub-samples and repeat the two tests discussed earlier.

Figure 3 reports the results of the two tests for the morning trading hours. At the 5% significance level, we confirm that the J-shaped pattern is not rejected for all days. In this case, better-informed traders trade in the opening, while worse-informed delay and trade later in the day. The identified trading pattern can be associated with very short-lived private information.

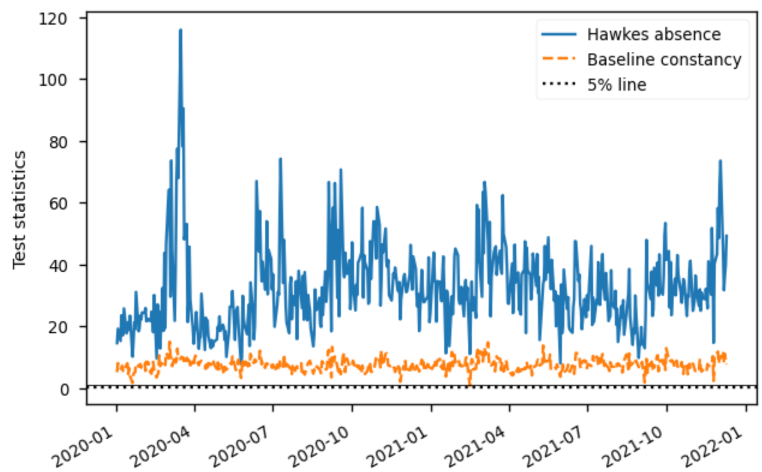

Figure 3: Test statistics for the null hypothesis in the two tests from Sections 5.4 and 5.5 with the 5% critical value. Morning trading hours, 9.30 to 12.45.

Figure 4 reports the results of the two tests for the afternoon hours. Again, we confirm the reverse J-pattern for all days. This finding provides a partial explanation of the significant hour-of-the-day effect on the CME. These intraday 'seasonal' patterns emerge as consequences of the interacting strategic decisions of informed and liquidity traders. Similar patterns can be assessed in bid/ask spreads, volumes, and volatility over the trading day, but this will require incorporating these characteristics (commonly called marks in the point processes literature) into the model. This development is beyond the scope of this paper and left for future research.

To summarize, the empirical findings favor Hawkes processes with time-varying baselines.

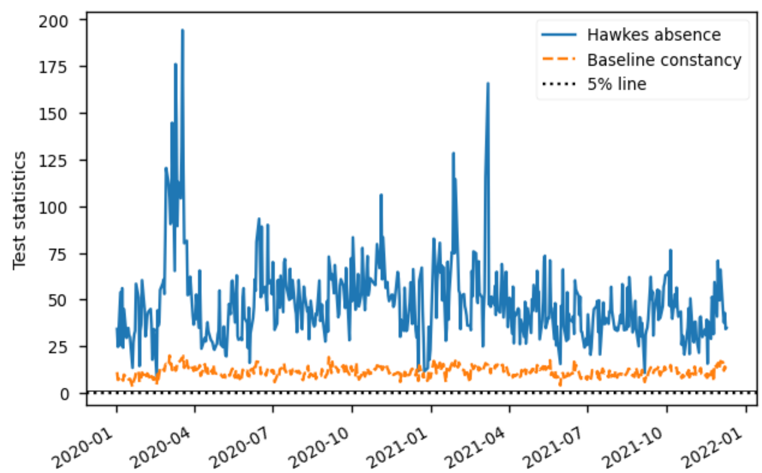

Figure 4: Test statistics for the null hypothesis in the two tests from Sections 5.4 and 5.5 with the 5% critical value. Afternoon trading hours, 12.45 to 16.00.

7 Conclusion

In this paper, we have considered Hawkes processes with Itô semimartingale baseline. This baseline can accommodate time variation, stochasticity, and locally bounded bursts. We have derived CLT for empirical average and preaveraging 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 constant baseline. The empirical application shows that the absence of a Hawkes component and constant baseline is always rejected, highlighting the presence of U-shaped and J-shaped patterns of trading.

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. The third author acknowledges that the publication was prepared within the framework of the Academic Fund Program at HSE University (grant N 2500037). **Data availability statement**: The data that support the findings of this study are available in the Tickdatamarket database. Restrictions apply to the availability of these data, which were used under license for this study.

References

Admati, A. R. and Pfleiderer, P. (1988). A theory of intraday patterns: Volume and price variability. *The Review of Financial Studies*, 1(1):3–40.

Aït-Sahalia, Y., Cacho-Diaz, J., and Laeven, R. J. (2015). Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, 117(3):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 Xiu, D. (2019). A Hausman test for the presence of market microstructure noise in high frequency data. *Journal of Econometrics*, 211(1):176–205.

Almgren, R. and Chriss, N. (2001). Optimal execution of portfolio transactions. *Journal of Risk*, 3:5–40.

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(1):7937–7941.

Bacry, E., Delattre, S., Hoffmann, M., and Muzy, J.-F. (2013). Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes and their Applications*, 123(7):2475–2499.

Bajgrowicz, P., Scaillet, O., and Treccani, A. (2016). Jumps in high-frequency data: Spurious detections, dynamics, and news. *Management Science*, 62(8):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*, pages 33–68. Springer.

Bowsher, C. G. (2007). Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141(2):876–912.

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

Cai, B., Zhang, J., and Guan, Y. (2024). Latent network structure learning from high-dimensional multivariate point processes. *Journal of the American Statistical Association*, 119(545):95–108.

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(2):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(4):1006–1024.

Christensen, K. and Kolokolov, A. (2024). An unbounded intensity model for point processes. *Journal of Econometrics*, 244(1):105840.

Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, pages 135–155.

Clements, A., Hurn, A., Lindsay, K., and Volkov, V. (2023). Estimating a non-parametric memory kernel for mutually-exciting point processes. *Journal of Financial Econometrics*, 21(5):1759–1790.

Clinet, S. and Potiron, Y. (2018a). Efficient asymptotic variance reduction when estimating volatility in high frequency data. *Journal of Econometrics*, 206(1):103–142.

Clinet, S. and Potiron, Y. (2018b). Statistical inference for the doubly stochastic self-exciting process. *Bernoulli*, 24(4B):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(1):289–337.

Clinet, S. and Potiron, Y. (2021). Disentangling sources of high frequency market microstructure noise. *Journal of Business & Economic Statistics*, 39(1):18–39.

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(2):391–408.

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

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

Engle, R. F. (2000). The econometrics of ultra-high-frequency data. *Econometrica*, 68(1):1–22.

Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, pages 1127–1162.

Epps, T. W. and Epps, M. L. (1976). The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distributions hypothesis. *Econometrica*, pages 305–321.

Erdemlioglu, D., Potiron, Y., Xu, T., and Volkov, V. (2025a). Estimation of latency for Hawkes processes with a polynomial periodic kernel. *Working paper available at https://www.fbc.keio.ac.jp/∼potiron/Erdemlioglu2025workingpaperestimationlatency.pdf*.

Erdemlioglu, D., Potiron, Y., Xu, T., and Volkov, V. (2025b). Estimation of time-dependent latency with locally stationary Hawkes processes. *Working paper available at https://www.fbc.keio.ac.jp/∼potiron/Erdemlioglu2025workingpaper.pdf*.

Fang, G., Xu, G., Xu, H., Zhu, X., and Guan, Y. (2024). Group network Hawkes process. *Journal of the American Statistical Association*, 119(547):2328–2344.

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

Filimonov, V. and Sornette, D. (2012). Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review E*, 85(5):056108.

Fulop, A., Li, J., and Yu, J. (2015). Self-exciting jumps, learning, and asset pricing implications. *The Review of Financial Studies*, 28(3):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(3):285–301.

Hardiman, S. J. and Bouchaud, J.-P. (2014). Branching-ratio approximation for the self-exciting Hawkes process. *Physical Review E*, 90(6):062807.

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

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

Hawkes, A. G. (1971b). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90.

Hawkes, A. G. and Oakes, D. (1974). A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3):493–503.

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

Huber, J. (2007). 'j'-shaped returns to timing advantage in access to information-experimental evidence and a tentative explanation. *Journal of Economic Dynamics and Control*, 31:2536–2572.

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

Jacod, J. and Protter, P. E. (2012). *Discretization of processes*. Springer Berlin, Heidelberg.

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

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

Jing, B.-Y., Liu, Z., and Kong, X.-B. (2014). On the estimation of integrated volatility with jumps and microstructure noise. *Journal of Business & Economic Statistics*, 32(3):457–467.

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.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, pages 1315–1335.

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

Morariu-Patrichi, M. and Pakkanen, M. S. (2022). State-dependent Hawkes processes and their application to limit order book modelling. *Quantitative Finance*, 22(3):563–583.

Mykland, P. A. and Zhang, L. (2009). Inference for continuous semimartingales observed at high frequency. *Econometrica*, 77(5):1403–1445.

Novikov, A. A. (1972). On an identity for stochastic integrals. *Teoriya Veroyatnostei i ee Primeneniya*, 17(4):761–765.

Potiron, Y. (2025a). Nonparametric inference for Hawkes processes with a stochastic time-dependent baseline. *Working paper available at https://www.fbc.keio.ac.jp/~ potiron/Potiron2025nonparametricworkingpaper.pdf*.

Potiron, Y. (2025b). Parametric inference for Hawkes processes with a general kernel. *Revise and Resubmit for the Annals of the Institute of Statistical Mathematics.*

Potiron, Y. (2025c). Parametric inference for nonlinear Hawkes processes. *Working paper available at https://www.fbc.keio.ac.jp/~ potiron/Potiron2025nonlinearworkingpaper.pdf*.

Potiron, Y. and Mykland, P. A. (2017). Estimation of integrated quadratic covariation with endogenous sampling times. *Journal of Econometrics*, 197(1):20–41.

Potiron, Y., Scaillet, O., Volkov, V., and Yao, W. (2025a). Estimation of branching ratio matrix for mutually exciting Hawkes processes with Itô semimartingale baseline. *Working paper available at https://www.fbc.keio.ac.jp/~ potiron/Potiron2025branchingworkingpaper.pdf*.

Potiron, Y., Scaillet, O., Volkov, V., and Yu, S. (2025b). Estimation of branching ratio for Hawkes processes with Itô semimartingale baseline. *Working paper available at https://www.fbc.keio.ac.jp/~ potiron/Potiron2025estimationworkingpaper.pdf*.

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(3):032318.

Rambaldi, M., Pennesi, P., and Lillo, F. (2015). Modeling foreign exchange market activity around macroeconomic news: Hawkes-process approach. *Physical Review E*, 91(1):012819.

Stoltenberg, E. A., Mykland, P. A., and Zhang, L. (2022). A CLT for second difference estimators with an application to volatility and intensity. *Annals of Statistics*, 50(4):2072–2095.

Tauchen, G. E. and Pitts, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica*, pages 485–505.

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(2):546–588.

Van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3. Cambridge university press.

Yu, J. (2004). Empirical characteristic function estimation and its applications. *Econometric Reviews*, 23(2):93–123.

SUPPLEMENTARY MATERIAL: This supplementary material (see Section 8) provides the proofs for the theoretical results, namely Proposition 1, Theorem 1, Proposition 2, and the five corollaries of Section 5 in "High-frequency estimation of Ito semimartingale baseline for Hawkes processes" by Yoann Potiron, Olivier Scaillet, Vladimir Volkov, and Seunghyeon Yu. In Section 9, we carry out a finite sample analysis, which corroborates the asymptotic theory.

8 Proofs

This section of the supplementary material gives detailed proofs for the theoretical results, namely Proposition 1, Theorem 1, Proposition 2, and the five corollaries of Section 5. They rely on theory developed in Bacry et al. (2013), Barndorff-Nielsen et al. (2006), Brémaud and Massoulié (1996), Clinet and Potiron (2018b) , Erdemlioglu et al. (2025b), Girsanov (1960), Jacod and Shiryaev (2013), Jacod and Protter (2011), Novikov (1972), Potiron (2025b), and Todorov and Tauchen (2011).

To start, we introduce some notations we will use for convenience in the proofs. We have that f, g and h are temporary functions, which may vary in the proofs. Moreover, C denotes a generic constant that does not depend on n and may differ.

We begin with the proof of the existence of a Hawkes process 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 (2018b) in which the kernel is exponential to the general kernel case. It also complements the proof of Theorem 1 in Cai et al. (2024) in which the form of the kernel is also restricted. See also Erdemlioglu et al. (2025b), who consider a generalized gamma kernel and Potiron (2025a) who studies a general baseline process.

Proof of Proposition 1. The strategy of the proof consists of defining a suitable sequence of 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 the point process N_t admits λ_t as an $\mathcal{F}_t$-intensity in the sense of Equation (2.1).

We first define for any time $t \in [0,T]$ the starting intensity $\lambda^0(t) = \mu_t$ and N_t^0, namely the point process counting the points of the Poisson process $\underline{N}$ below the curve $t \to \lambda_t^0$ as $N_t^0 = \int_{[0,t]\times\mathbb{R}} \mathbf{1}_{[0,\lambda_s^0]}(x)\underline{N}(ds \times dx)$. We then define recursively the sequence of $(N_t^k, \lambda_t^k)_{k\geq 1}$ as

$$\begin{aligned} \lambda_t^{k+1} &= \mu_t + \int_0^{t-} \phi(t-s)dN_s^k \\ N_t^{k+1} &= \int_{[0,t]\times\mathbb{R}} \mathbf{1}_{[0,\lambda_s^{k+1}]}(x)\underline{N}(ds \times dx). \end{aligned} \tag{8.1}$$

First, we have that the intensity λ_t^k is a.s. positive as an application of Condition 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 the point process N_t^k is $\mathcal{F}_t$-adapted, the process λ_t^k is $\mathcal{F}_t$-predictable and an $\mathcal{F}_t$-intensity of the point process 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 time $t \in [0,T]$.

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, we have $\rho_t^{k+1} = \mathbb{E}\Big[\int_0^t \phi(t-s)(\lambda_s^k - \lambda_s^{k-1})ds\Big|\widetilde{\mathcal{F}}_T\Big] = \int_0^t \phi(t-s)\rho_s^k ds$, where the first equality is obtained by Lemma 10.1 (p. 2) in the supplement of Clinet and Potiron (2018b), when $\mathcal{G} = \widetilde{\mathcal{F}}_T$, along with Equation (8.1), and the second equality by Tonelli's theorem and the definition of ρ_t^k. If we define the sequence of random variables Φ_t^k as $\Phi_t^k = \int_0^t \rho_s^k ds$, we have by another application of Tonelli's theorem that

$$\Phi_t^{k+1} = \int_0^t \Big(\int_0^{t-s} \phi(u)du\Big)\rho_s^k ds. \tag{8.2}$$

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 Equation (8.2) implies that $\Phi_t^{k+1} \leq \|\phi\|_1 \Phi_t^k$ a.s..

Then, we can deduce that $F : \Phi_t^k \to \Phi_t^{k+1}$ is a.s. a contraction function since Condition 1 (d) states that $\|\phi\|_1 < 1$. 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

$$\mathbb{E}\Big[\int_0^t \lambda_s ds\Big|\widetilde{\mathcal{F}}_T\Big] \leq \int_0^t \mu_s ds + \|\phi\|_1 \mathbb{E}\Big[\int_0^t \lambda_s ds\Big|\widetilde{\mathcal{F}}_T\Big]. \tag{8.3}$$

By rearranging the terms in Equation (8.3), we get that

$$\mathbb{E}\Big[\int_0^t \lambda_s ds \Big| \widetilde{\mathcal{F}}_T\Big] \leq (1-\|\phi\|_1)^{-1}\int_0^t \mu_s ds. \tag{8.4}$$

Given Condition 1 (b), the expression on the left side of Equation (8.4) is finite a.s.. Given that its conditional expectation is finite, $\int_0^t \lambda_s ds$ is finite a.s.. Moreover, the intensity λ_t is $\mathcal{F}_t$-predictable as a limit of such processes. The point process N_t counts the points of the Poisson process $\underline{N}$ under the curve $t \mapsto \lambda_t$ by applying the monotone convergence theorem. Therefore, the point process N_t 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 the point process N_t is finite a.s.. It remains to show that the intensity λ_t is of the form (2.1). The monotonicity properties of the processes N_t^k and λ_t^k ensure that $\lambda_t^k \leq \mu_t + \int_0^t \phi(t-s)\, dN_s$ and $\lambda_t \geq \mu_t + \int_0^t \phi(t-s)\, dN_s^k$ for any nonnegative integer $k \in \mathbb{N}$ and any time $t \in [0,T]$. Finally, this gives Equation (2.1) by taking the limit $k \to +\infty$ in both inequalities. □

Before giving the first lemma in the proof of the main CLT, we introduce some definitions. We 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 two integrable functions f and g, we define the convolution of f and g as $f * g_t = \int_{-\infty}^\infty f(t-s)g(s)\, ds$. In particular, we define recursively $f^{*1} = f$ and f^{*k} is the convolution product of $f^{*(k-1)}$ with the function f for any integer $k \geq 2$. For an integrable function f and a stochastic process X, we define the convolution of f and X as $f * dX_t = \int_{-\infty}^\infty f(t-s)\, dX_s$. In addition, we denote by $\psi \colon \mathbb{R}_+ \to \mathbb{R}_+$ the resolvent kernel of ϕ which satisfies $\psi(t) = \phi(t) + \phi * \psi_t$ for any time $t \in \mathbb{R}_+$. Similarly, we denote by ψ^n the resolvent kernel of ϕ^n, namely $\psi^n(t) = \phi^n(t) + \phi^n * \psi_t^n$ for any time $t \in \mathbb{R}_+$. Moreover, 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 $\Delta_i\Psi^n(-t) = \int_{(i-1)\Delta_n}^{i\Delta_n} \psi^n(s-t)\, ds$. Finally, we denote the 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 any nonnegative integer $n \in \mathbb{N}$, any time $t \in [0,T]$ and some positive real number $C \in \mathbb{R}_+$ which does not depend on n and t.

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

Lemma 1. *Under Condition 1 and Condition 2 (a), we have for any time $t \in \mathbb{R}_+$ as $n \to \infty$ that*

$$\psi(t) \geq 0, \tag{8.5}$$

$$\psi^n(t) = n\psi(nt), \tag{8.6}$$

$$\Psi^n(t) = \widehat{\psi}(0) + \underline{O}\left(1 \wedge \frac{1}{nt}\right), \tag{8.7}$$

$$\Delta_i \Psi^n(-t) = \underline{O}\left(1 \wedge \frac{1}{n((i-1)\Delta_n - t)}\right). \tag{8.8}$$

Proof of Lemma 1. Since $\|\phi\|_1 < 1$ by Condition 1 (d), the function $\theta\colon f \mapsto (\mu + \phi * f)$ is a contraction function. Thus, we can apply Banach fixed-point theorem to get a fixed-point $\psi = f_\infty$ with recursion $f_k = \theta(f_{k-1})$. Then, we obtain recursively that

$$\begin{aligned} f_k = \theta(f_{k-1}) &= \mu + \phi * f_{k-1} = \mu + \phi * (\theta(f_{k-2})) = \mu + \phi * (\mu + \phi * f_{k-2}) \\ &= \mu + \sum_{l=1}^{k-2} \phi^{*l} * \mu + \phi^{*(k-1)} * f_1. \end{aligned}$$

For the initial value, we can set $f_1 = 0$. Then, $f_k(t)$ is nonnegative for any integer $k > 1$ and any time $t \in \mathbb{R}_+$ since the kernel ϕ is nonnegative. Thus, we have shown Equation (8.5). By the scaling property of the Laplace transform, we have that $\widehat{\phi^n}(s) = \widehat{\phi}(s/n)$, and hence Equation (8.6) holds.

To show Equation (8.7), we have that

$$\begin{aligned} \widehat{\psi}(0) - \widehat{\psi}\left(\frac{1}{nt}\right) &= \int_0^\infty (1 - e^{-\frac{s}{nt}})\psi(s)\,ds \geq \int_{nt}^\infty (1 - e^{-\frac{s}{nt}})\psi(s)\,ds \\ &\geq (1 - e^{-1}) \int_{nt}^\infty \psi(s)\,ds. \end{aligned} \tag{8.9}$$

This is because the two functions ψ and ϕ are nonnegative. From Condition 2 (a), we obtain that $\widehat{\phi}'(0) = \int_0^\infty t\phi(t)dt < \infty$. Then, we can apply Taylor's theorem for $\widehat{\psi}(s) = \widehat{\phi}(s)/(1 - \widehat{\phi}(s))$ with the remainder $\widehat{\psi}(s) = \widehat{\psi}(0) + \widehat{\psi}'(0)s + h(s)s$, where $\lim_{s\to 0} h(s) = 0$. As the function $\widehat{\psi}$ is decreasing, i.e., $\widehat{\psi}(s) \geq \widehat{\psi}(t)$ for any time $0 \leq s < t$, and $\widehat{\psi}(0) < \infty$, we can deduce that

$$0 \leq \widehat{\psi}(0) - \widehat{\psi}\left(\frac{1}{nt}\right) = -\left(\widehat{\psi}'(0) + h\left(\frac{1}{nt}\right)\right)\frac{1}{nt}$$

$$\leq \left(|\widehat{\psi}'(0)| + \left|h\left(\frac{1}{nt}\right)\right|\right)\frac{1}{nt}\mathbf{1}_{\{nt\geq 1\}} + \widehat{\psi}(0)\mathbf{1}_{\{nt<1\}}.$$

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

$$\left|\widehat{\psi}(0) - \widehat{\psi}\left(\frac{1}{nt}\right)\right| \leq C\left(1 \wedge \frac{1}{nt}\right). \tag{8.10}$$

It implies for sufficiently big n that

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

Here, we have used successively Equation (8.6), the change of variable $ns \to s$, Expressions (8.9) and (8.10). Thus, we have proven Equation (8.7). With the same arguments, we can also show that Equation (8.8) holds. □

We define the point process N_t compensated by its intensity λ_t as $M_t = N_t - \int_0^t \lambda_u du$ for any time $t \in [0, T]$. By definition of a compensator, we have that the process M_t is an $\mathcal{F}_t$-martingale a.s.. Then, we denote the sum of the baseline and the convolution of the resolvent kernel and baseline by $\nu_t^n = \mu_t + \psi^n * \mu_t$ for any time $t \in [0, T]$. Moreover, we define the limit of ν_t^n as $\nu_t = (1+\widehat{\psi}(0))\mu_t$. Finally, we introduce $\underline{O}_{L^k}$, which is the strict big O in L^k-norm $\underline{O}_{L^k}$ defined through $f(n,i) = \underline{O}_{L^k}\big(g(n,i)\big) \iff \mathbb{E}\big[\big(f(n,i)\big)^k\big]^{\frac{1}{k}} = \underline{O}\big(g(n,i)\big)$.

The following lemma exhibits a martingale representation of the intensity λ_t. It is based on the convolution of the resolvent kernel and the martingale. It extends Lemma 3 in Bacry et al. (2013) and Proposition 2.1 (p. 606) in Jaisson and Rosenbaum (2015), which considers an invariant baseline and asymptotics when the final time $T \to \infty$, to the time-varying baseline and in-fill asymptotics case.

Lemma 2. *Under Condition 1 and Condition 2 (a), we have for any time $t \in [0, T]$ that the intensity λ_t has the martingale representation*

$$\lambda_t = n\nu_t^n + \psi^n * dM_t. \tag{8.11}$$

Moreover, we have

$$\nu_t^n - \nu_t = \underline{O}_{L^k}(1/\sqrt{n}). \tag{8.12}$$

Proof of Lemma 2. By Lemma 3 in Bacry et al. (2013), 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 consider the process λ_t, which satisfies for any time $t \in [0, T]$ that

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

Applying Lemma 3 in Bacry et al. (2013) with the function $h(t)$ defined as $h(t) = n\mu_t + \phi^n * dM_t$, we have for any time $t \in [0, T]$ that

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

Thus, we can obtain Equation (8.11).

We now show Equation (8.12). First, we get

$$\psi^n * \mu_t = \int_{-\infty}^t \psi^n(t-s)\mu_s\, ds = \int_0^\infty \psi^n(s)\mu_{t-s}\, ds = \int_0^\infty \psi(s)\mu_{t-\frac{s}{n}}\, ds.$$

Here, we use Equation (8.6) from Lemma 1. Then, we can deduce that

$$\nu_t^n - \nu_t = \psi^n * \mu_t - \widehat{\psi}(0)\mu_t = \int_0^\infty \psi(s)\mu_{t-\frac{s}{n}} ds - \widehat{\psi}(0)\mu_t = \int_0^\infty \psi(s)\big(\mu_{t-\frac{s}{n}} - \mu_t\big) ds. \tag{8.13}$$

By the use of Burkholder-Davis-Gundy inequality (see Expression (2.1.32) in Jacod and Protter (2012) (p. 39)) along with the baseline μ_t being an Itô semimartingale, we get Equation (8.12). $\square$

The following lemma extends Lemma 10.3 from Clinet and Potiron (2018b) (pp. 4-6 in the supplement) in which the kernel is exponential and Lemma 6 in Erdemlioglu et al. (2025b) in which the kernel is generalized gamma.

Lemma 3. *We assume that Condition 1 and Condition 2 (b) hold. Then, there exists C such that $\sup_{t\in[0,T]} \mathbb{E}\lambda_t^k \leq Cn^k$ for any nonnegative integer $k \in \mathbb{N}$.*

Proof of Lemma 3. Since the baseline is an Itô semimartingale, we can deduce that $\sup_{t\in[0,T]} \mathbb{E}\mu_t^k \leq Cn^k$. Moreover, we can extend the arguments from the proof of Lemma 6 in Erdemlioglu et al. (2025b) along with Condition 1 (d) and Condition 2 (b) to obtain $\sup_{t\in[0,T]} \mathbb{E}\big[\big(\int_0^t \phi(t-s)\,dN_s\big)^k\big] \leq Cn^k$. Finally, we can deduce the lemma since the baseline μ_t and the Hawkes component $\int_0^t \phi(t-s)\,dN_s$ are independent by Condition 1 (c). □

For any interval number $i = 1, \dots, M$, we denote by X_i the process X evaluated at the end of the ith interval, i.e., $X_i = X_{i\Delta_n}$. We also define $\bar{X}_i$ as the average of X_t on the ith interval, i.e., $\bar{X}_i = \Delta_n^{-1} \int_{(i-1)\Delta_n}^{i\Delta_n} X_t\,dt$. Moreover, we define the estimator of rescaled integrated intensity on the ith interval as $\widehat{\bar{\nu}}_i^n = \frac{\widehat{\lambda}_i}{n}$. In addition, we define the rescaled increment of the martingale on the ith interval as $\varepsilon_i = \frac{1}{n\Delta_n} \int_{(i-1)\Delta_n}^{i\Delta_n} dM_t$. Furthermore, we denote the rescaled increment related to the Hawkes component on the ith interval by

$$\epsilon_i = \frac{1}{n\Delta_n}\Big\{ \int_0^{(i-1)\Delta_n} \Delta_i\Psi^n(-t)\,dM_t + \int_{(i-1)\Delta_n}^{i\Delta_n} \Psi^n(i\Delta_n - t)\,dM_t\Big\}.$$

Finally, we define the sum of ε_i and ϵ_i as $u_i = \varepsilon_i + \varepsilon_i$.

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 component ϵ_i.

Lemma 4. *Under Condition 1 and Condition 2 (a), we have for any interval number $i = 1, \dots, M$ the decomposition*

$$\widehat{\bar{\nu}}_i^n = \bar{\nu}_i^n + u_i. \tag{8.15}$$

Proof of Lemma 4. It is obtained using Lemma 2 and Fubini's theorem. □

We denote the L^p-norm of X by $\|X\|_p$. We also 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 $X_n = O_{L^p}(\alpha_n) \iff \|X_n\|_p = O(\alpha_n)$. Moreover, we use $\mathbb{E}_{i-1}$, Var_{i-1} and Cov_{i-1} in place of $\mathbb{E}[\cdot|\mathcal{F}_{(i-1)\Delta_n}]$, $\mathrm{Var}[\cdot|\mathcal{F}_{(i-1)\Delta_n}]$ and $\mathrm{Cov}[\cdot|\mathcal{F}_{(i-1)\Delta_n}]$ for any interval number $i = 1, \dots, M$. Finally, we introduce $\vartheta_t^n = (1 + \widehat{\psi}(0))^2 \nu_t^n$ for any $t \in [0, T]$.

The following lemma provides moments of u_i.

Lemma 5. *We assume that Condition 1, Condition 2 (a) and (b) hold. Then, we have for any nonnegative integer $k \in \mathbb{N}$ as $n \to \infty$ that*

$$\mathbb{E}[|u_i|^k] \leq \frac{C}{(n\Delta_n)^{k/2}}, \tag{8.16}$$

$$\mathbb{E}_{i-1}[u_i \mid \mathcal{F}^\mu] = \underline{O}_{L^k}\left(\frac{\log n}{n\Delta_n}\right), \tag{8.17}$$

$$\mathbb{E}_{i-1}[u_i u_j \mid \mathcal{F}^\mu] = \underline{O}_{L^k}\left(\frac{\log n}{(n\Delta_n)^2}\right) \quad \text{for any } i < j, \tag{8.18}$$

$$\mathbb{E}_{i-1}[u_i^2 \mid \mathcal{F}^\mu] = \frac{1}{n\Delta_n}\bar{\vartheta}_i^n + \underline{O}_{L^k}\left(\frac{\log n}{(n\Delta_n)^2}\right), \tag{8.19}$$

$$\mathbb{E}_{i-1}[u_i^3 \mid \mathcal{F}^\mu] = \frac{(1+3\widehat{\psi}(0))(1+\widehat{\psi}(0))}{(n\Delta_n)^2}\bar{\vartheta}_i^n + \underline{O}_{L^k}\left(\frac{(\log n)^2}{(n\Delta_n)^3}\right), \tag{8.20}$$

$$\mathbb{E}_{i-1}[u_i^4 \mid \mathcal{F}^\mu] = \frac{3}{(n\Delta_n)^2}(\bar{\vartheta}_i^n)^2 + \underline{O}_{L^k}\left(\frac{1}{(n\Delta_n)^3} + \frac{(\log n)^3}{(n\Delta_n)^4}\right). \tag{8.21}$$

Proof of Lemma 5. Without loss of generality and for convenience of notation, we assume that the baseline μ_t and thus ν_t are nonrandom throughout this proof. We first calculate the moments of ϵ_i. We have that $\mathbb{E}_{i-1}[\epsilon_i] = \underline{O}_{L^k}\left(\frac{\log n}{n\Delta_n}\right)$ 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 Lemmas 3 and 4 and Itô isometry for point processes. We thus obtain that

$$\mathbb{E}\Big[\big(\mathbb{E}_{i-1}[\epsilon_i]\big)^2\Big] \leq \frac{Cn}{(n\Delta_n)^2}\left(\int_0^{(i-1)\Delta_n - n^{-1}} \frac{1}{n((i-1)\Delta_n - t)} dt + \int_{(i-1)\Delta_n - n^{-1}}^{(i-1)\Delta_n} dt\right) \leq \frac{Cn}{(n\Delta_n)^2}\left(\frac{\log n}{n} + \frac{1}{n}\right).$$

For $k > 2$, by Lemma 2.1.5 in Jacod and Protter (2011), Lemma 3, and Hölder's inequality, we have that $(n\Delta_n)^k \mathbb{E}\Big[\big(\mathbb{E}_{i-1}[\epsilon_i]\big)^k\Big] \leq \underline{O}\big((\log n)^k\big)$. To calculate the moments of ε_i, we can use the same arguments and Lemma 2. Thus, we can deduce Expression (8.16). With the same arguments, we can get Equations (8.17), (8.18), (8.19), (8.20), and (8.21). □

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

Lemma 6. *We have $\bar{\nu}_i - \nu_{i-1} = \int_0^{\Delta_n} r_t\, d\nu_{i-1+t}, \nu_i - \bar{\nu}_i = \int_0^{\Delta_n} l_t\, d\nu_{i-1+t}$ and $\Delta_i \bar{\nu} = \int_0^{2\Delta_n} lr_t\, d\nu_{i-2+t}$ for any interval number $i = 1, \ldots, M$.*

Proof of Lemma 6. We have

$$\bar{\nu}_i - \nu_{i-1} = \frac{1}{\Delta_n}\bigg(\int_{(i-1)\Delta_n}^{i\Delta_n}\bigg(\nu_{i-1} + \int_{(i-1)\Delta_n}^{t} d\nu_s\bigg)dt - \nu_{i-1}\Delta_n\bigg) = \frac{1}{\Delta_n}\bigg(\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{(i-1)\Delta_n}^{t} d\nu_s\, dt\bigg)$$
$$= \frac{1}{\Delta_n}\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{s}^{i\Delta_n} dt\, d\nu_s = \frac{1}{\Delta_n}\int_{(i-1)\Delta_n}^{i\Delta_n}(i\Delta_n - s)\, d\nu_s = \int_0^{\Delta_n} r_s\, d\nu_{i-1+s}.$$

Then, symmetry yields $\nu_i - \bar{\nu}_i = \int_0^{\Delta_n} l_t\, d\nu_{i-1+t}$. We also have that

$$\bar{\nu}_i - \bar{\nu}_{i-1} = \bar{\nu}_i - \nu_{i-1} + \nu_{i-1} - \bar{\nu}_{i-1} = \int_{\Delta_n}^{2\Delta_n} r_t\, d\nu_{i-2+t} + \int_0^{\Delta_n} l_t\, d\nu_{i-2+t} = \int_0^{2\Delta_n} lr_t\, d\nu_{i-2+t}.$$

Finally, we can support the third assertion with similar arguments. □

Without loss of generality and with a slight abuse of notation, we rewrite ν_t itself as an Itô semimartingale with Grigelionis representation of the form (2.2). This is possible since ν_t is the product of a deterministic constant and the baseline μ_t by definition. Condition 2 (d) implies that the jumps are absolutely summable, i.e., $\sum_{s\leq t}|\Delta\nu_s| < \infty$, a.s. for any time $t \in [0,T]$. Moreover, it implies that we can express the baseline ν_t as

$$\nu_t = \int_0^t b_s^\delta\, ds + \int_0^t \sigma_s\, dW_s + \sum_{s\leq t}\Delta\nu_s. \tag{8.22}$$

Thus, we can define the continuous part of the baseline ν_t as $\nu_t^{(c)} = \int_0^t b_s^\delta\, ds + \int_0^t \sigma_s\, dW_s$ and the discontinuous part of the baseline ν_t as $\nu_t^{(d)} = \sum_{s\leq t}\Delta\nu_s$. For any interval number $i = 1,\ldots,M$, we also define $\widehat{\bar{\nu}}_i^{(c)}$ as $\widehat{\bar{\nu}}_i^{(c)} = \bar{\nu}_i^{(c)} + u_i$ and $\widehat{\bar{\nu}}_i$ as $\widehat{\bar{\nu}}_i = \widehat{\bar{\nu}}_i^{(c)} + \bar{\nu}_i^{(d)}$. Moreover, we denote by ΔX_i the difference for the process X between the start and the end of the ith interval, i.e., $\Delta_i X = X_i - X_{i-1}$. Finally, we define $X_{t,n} = \underline{o}_{\mathbb{P}}(1)$ if $X_{t,n} \overset{u.c.p.}{\to} 0$ as $n \to \infty$.

The next lemma shows that we can remove the discontinuous part of the baseline ν_t in the remainder of the proofs. This is based on the proof of Theorem 3 in Jing et al. (2014) which study estimation of integrated volatility in the presence of market microstructure noise.

Lemma 7. *We assume that Condition 1, Condition 2 (a), (c), (d), (e) and (f) hold. Then, we have as $n \to \infty$ uniformly for any time $t \in [0,T]$ that*

$$\Delta_n^{-\frac{1}{2}}\sum_{i=1}^{\lfloor t/\Delta_n\rfloor}\big(\Delta_i\widehat{\bar{\nu}}^n\big)^2\mathbf{1}_{\{|\Delta_i\widehat{\bar{\nu}}^n|\leq\alpha\Delta_n^{\varpi}\}} = \Delta_n^{-\frac{1}{2}}\sum_{i=1}^{\lfloor t/\Delta_n\rfloor}\big(\Delta_i\widehat{\bar{\nu}}^{(c)}\big)^2 + \underline{o}_{\mathbb{P}}(1). \tag{8.23}$$

Proof of Lemma 7. We first show as $n \to \infty$ that

$$\Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \left(\Delta_i \widehat{\bar{\nu}}\right)^2 \mathbf{1}_{\{|\Delta_i \widehat{\bar{\nu}}| \le \alpha \Delta_n^{\varpi}\}} = \Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \left(\Delta_i \widehat{\bar{\nu}}^{(c)}\right)^2 + \underline{o}_{\mathbb{P}}(1). \tag{8.24}$$

From the definition of $\widehat{\bar{\nu}}$, we can deduce that

$$4(\Delta_i \widehat{\bar{\nu}})^2 = (\Delta_i \widehat{\bar{\nu}}^{(c)})^2 + 2(\Delta_i \widehat{\bar{\nu}}^{(c)})(\Delta_i \bar{\nu}^{(d)}) + (\Delta_i \bar{\nu}^{(d)})^2.$$

Thus, we get

$$\begin{aligned}\Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} (\Delta_i \widehat{\bar{\nu}})^2 \mathbf{1}_{\{|\Delta_i \widehat{\bar{\nu}}| \le \alpha \Delta_n^{\varpi}\}} &= \Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} (\Delta_i \widehat{\bar{\nu}}^{(c)})^2 \mathbf{1}_{\{|\Delta_i \widehat{\bar{\nu}}| \le \alpha \Delta_n^{\varpi}\}} \\ &\quad + \Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} 2(\Delta_i \widehat{\bar{\nu}}^{(c)})(\Delta_i \bar{\nu}^{(d)}) \mathbf{1}_{\{|\Delta_i \widehat{\bar{\nu}}| \le \alpha \Delta_n^{\varpi}\}} \\ &\quad + \Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} (\Delta_i \bar{\nu}^{(d)})^2 \mathbf{1}_{\{|\Delta_i \widehat{\bar{\nu}}| \le \alpha \Delta_n^{\varpi}\}}.\end{aligned}$$

We introduce $(\mathrm{I})_t = \Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} (\Delta_i \widehat{\bar{\nu}}^{(c)})^2 \mathbf{1}_{\{|\Delta_i \widehat{\bar{\nu}}| \le \alpha \Delta_n^{\varpi}\}}$. In what follows, we first show that $(\mathrm{I})_t = \Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \left(\Delta_i \widehat{\bar{\nu}}^{(c)}\right)^2 + \underline{o}_{\mathbb{P}}(1)$ as $n \to \infty$. The domination $\mathbf{1}_{\{|x| > a\}} \le 2^k |x|^k / a^k$ along with Lemma 5 and Lemma 6 for any positive integer $k > 0$ gives as $n \to \infty$ that

$$\begin{aligned}\left| (\mathrm{I})_t - \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \left(\Delta_i \widehat{\bar{\nu}}^{(c)}\right)^2 \right| &= \underline{O}_{L^1}\left(\Delta_n^{-\frac{3}{2}} \frac{\Delta_n^{(k+2)/2} + (n\Delta_n)^{-(k+2)/2}}{\varpi^k} \right. \\ &\quad \left. + \underline{O}_{\mathbb{P}}\left(\frac{\Delta_n^{-\frac{1}{2} - \frac{1}{p}}}{\varpi} \left(\Delta_n + (n\Delta_n)^{-1}\right) \right) \right).\end{aligned} \tag{8.25}$$

By Condition 2 (e) and choosing sufficiently large integers k and p, we get $(\mathrm{I})_t = \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \left(\Delta_i \widehat{\bar{\nu}}^{(c)}\right)^2 + \underline{o}_{\mathbb{P}}(1)$.

We now introduce $(\mathrm{II})_t = \Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} 2(\Delta_i \widehat{\bar{\nu}}^{(c)})(\Delta_i \bar{\nu}^{(d)}) \mathbf{1}_{\{|\Delta_i \widehat{\bar{\nu}}| \le \alpha \Delta_n^{\varpi}\}}$. First, we have by an extension of the proof of Theorem 3 in Jing et al. (2014), Hölder's inequality, and Condition 2 (d) as $n \to \infty$ that

$$|(\mathrm{II})_t| = \underline{O}_{\mathbb{P}}\left(\frac{\Delta_n^{-\frac{1}{2} - \frac{1}{p}}}{\varpi^k} \left(\Delta_n^{(k+1)/2} + (n\Delta_n)^{-(k+1)/2}\right) \right) + \underline{O}_{\mathbb{P}}\left(\Delta_n^{-\frac{1}{2}} \varpi^{2-\beta}\right).$$

Thus, we can deduce that $(\mathrm{II}) = \underline{o}_{\mathbb{P}}(1)$ as $n \to \infty$ by Condition 2 (e) and with a sufficiently large integer k. Finally, we can show that $\Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} (\Delta_i \bar{\nu}^{(d)})^2 \mathbf{1}_{\{|\Delta_i \widehat{\bar{\nu}}| \le \alpha \Delta_n^{\varpi}\}} = \underline{o}_{\mathbb{P}}(1)$ as $n \to \infty$ with the same arguments as for the proof of $(\mathrm{II})_t = \underline{o}_{\mathbb{P}}(1)$. Thus, we have shown that Equation (8.24) holds.

We show now as $n \to \infty$ that

$$\Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \left(\Delta_i \widehat{\bar{\nu}}^n\right)^2 \mathbf{1}_{\{|\Delta_i \widehat{\bar{\nu}}^n| \leq \alpha \Delta_n^{\varpi}\}} = \Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \left(\Delta_i \widehat{\bar{\nu}}\right)^2 \mathbf{1}_{\{|\Delta_i \widehat{\bar{\nu}}| \leq \alpha \Delta_n^{\varpi}\}} + o_{\mathbb{P}}(1). \tag{8.26}$$

We first prove that $\left(\mathbb{E}\left|\Delta_i(\bar{\nu}^{n(c)} - \bar{\nu}^{(c)})\right|^k\right)^{\frac{1}{k}} \leq Cn^{-\frac{5}{8}}$ for all nonnegative integer $k > 0$ by Equation (8.13), Burkholder-Davis-Gundy inequality along with Condition 2 (c) and (f). By similar arguments, we show that $\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \left|\Delta_i(\bar{\nu}^{n(d)} - \bar{\nu}^{(d)})\right| = O_{\mathbb{P}}\left(n^{-\frac{3}{8}}\right)$. We define $(\mathrm{IV})_t$ as

$$(\mathrm{IV})_t = \Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \left(\Delta_i \widehat{\bar{\nu}}^n\right)^2 \Big(\mathbf{1}_{\{|\Delta_i \widehat{\bar{\nu}}^n| \leq \alpha \Delta_n^{\varpi}\}} - \mathbf{1}_{\{|\Delta_i \widehat{\bar{\nu}}| \leq \alpha \Delta_n^{\varpi}\}}\Big).$$

To prove Equation (8.26), it is then sufficient to show that $(\mathrm{IV})_t = o_{\mathbb{P}}(1)$ as $n \to \infty$. By some algebraic manipulation, we first get that

$$\mathbf{1}_{\{|\Delta_i \widehat{\bar{\nu}}^n| \leq \alpha \Delta_n^{\varpi}\}} - \mathbf{1}_{\{|\Delta_i \widehat{\bar{\nu}}| \leq \alpha \Delta_n^{\varpi}\}} = \mathbf{1}_{\{|\Delta_i \widehat{\bar{\nu}}| > \alpha \Delta_n^{\varpi}, |\Delta_i \widehat{\bar{\nu}}^n| \leq \alpha \Delta_n^{\varpi}\}} - \mathbf{1}_{\{|\Delta_i \widehat{\bar{\nu}}| \leq \alpha \Delta_n^{\varpi}, |\Delta_i \widehat{\bar{\nu}}^n| > \alpha \Delta_n^{\varpi}\}}.$$

We also note that $\left\{|\Delta_i \widehat{\bar{\nu}}| > \alpha \Delta_n^{\varpi}, |\Delta_i \widehat{\bar{\nu}}^n| \leq \alpha \Delta_n^{\varpi}\right\} \subset \left\{|\Delta_i(\widehat{\bar{\nu}}^n - \widehat{\bar{\nu}})| > \alpha \Delta_n^{\varpi}, |\Delta_i \widehat{\bar{\nu}}^n| \leq \alpha \Delta_n^{\varpi}\right\} \cup \left\{|\Delta_i(\widehat{\bar{\nu}}^n - \widehat{\bar{\nu}})| \leq \alpha \Delta_n^{\varpi}, |\Delta_i \widehat{\bar{\nu}}| \in (\alpha \Delta_n^{\varpi}, 2\alpha \Delta_n^{\varpi}]\right\}$. Moreover, we have that $\left\{|\Delta_i \widehat{\bar{\nu}}| \leq \alpha \Delta_n^{\varpi}, |\Delta_i \widehat{\bar{\nu}}^n| > \alpha \Delta_n^{\varpi}\right\} \subset \left\{|\Delta_i(\widehat{\bar{\nu}}^n - \widehat{\bar{\nu}})| > \alpha \Delta_n^{\varpi}, |\Delta_i \widehat{\bar{\nu}}| \leq \alpha \Delta_n^{\varpi}\right\} \cup \left\{|\Delta_i(\widehat{\bar{\nu}}^n - \widehat{\bar{\nu}})| \leq \alpha \Delta_n^{\varpi}, |\Delta_i \widehat{\bar{\nu}}^n| \in (\alpha \Delta_n^{\varpi}, 2\alpha \Delta_n^{\varpi}]\right\}$. Then, we obtain as $n \to \infty$ that

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

We can also show as $n \to \infty$ that $\Delta_n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \left(\Delta_i \widehat{\bar{\nu}}^n\right)^2 \mathbf{1}_{\{|\Delta_i \widehat{\bar{\nu}}| \leq \alpha \Delta_n^{\varpi}, |\Delta_i \widehat{\bar{\nu}}^n| > \alpha \Delta_n^{\varpi}\}} = o_{\mathbb{P}}(1)$, so we can deduce that $(\mathrm{IV})_t \overset{u.c.p.}{\to} 0$ as $n \to \infty$. □

The next lemma shows that we can remove the drift from the baseline ν_t if we assume Novikov Condition 2 (g) (see Novikov (1972)), which is required to apply the Girsanov theorem (see Girsanov (1960)). We consider an equivalent probability measure $\mathbb{P}^*$ under which the baseline 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 8. *Under Condition 2 (g) and if we assume that the statement of Theorem 1 holds under the equivalent probability measure $\mathbb{P}^*$, the same statement holds under the probability measure $\mathbb{P}$.*

Proof of Lemma 8. We define $\overline{M}_t$ as $\overline{M}_t = \exp\big(\int_0^t \frac{b_s^\delta}{\sigma_s} dW_s - \frac{1}{2}\int_0^t \frac{(b_s^\delta)^2}{\sigma_s^2} ds\big)$ for any time $0 \leq t \leq T$, which by Condition 2(g) satisfies the Novikov condition and thus is a positive martingale. By the Girsanov theorem, we can thus consider an equivalent probability measure $\mathbb{P}^*$. 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 $W_t^* = W_t + \int_0^t \frac{b_s^\delta}{\sigma_s} dW_s$ is a standard Wiener process under the equivalent probability measure $\mathbb{P}^*$. To show that the statement of Theorem 1 holds under the probability measure $\mathbb{P}$, it is sufficient to prove for any eveet $E \in \mathcal{F}_T$, any measurable function h and jointly for any time $t \in [0, T]$ as $n \to \infty$ that

$$\mathbb{E}_{\mathbb{P}}\Big[h(X_t)\mathbf{1}_E\Big] \to \mathbb{E}_{\mathbb{P}}\Big[h\Big(\int_0^t w_s d\widetilde{W}_s\Big)\mathbf{1}_E\Big]. \tag{8.27}$$

By a change of probability in the expectation, we obtain

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

Since $\overline{M}_T^{-1} \in \mathcal{F}_T$ and the statement of Theorem 1 holds under the equivalent probability measure $\mathbb{P}^*$, we can deduce as $n \to \infty$ that

$$\mathbb{E}_{\mathbb{P}^*}\Big[h(X_t)\mathbf{1}_E\overline{M}_T^{-1}\Big] \to \mathbb{E}_{\mathbb{P}^*}\Big[h\Big(\int_0^t w_s d\widetilde{W}_s\Big)\mathbf{1}_E\overline{M}_T^{-1}\Big].$$

Finally, we obtain $\mathbb{E}_{\mathbb{P}^*}\Big[h\Big(\int_0^t w_s d\widetilde{W}_s\Big)\mathbf{1}_E\overline{M}_T^{-1}\Big] = \mathbb{E}_{\mathbb{P}}\Big[h\Big(\int_0^t w_s d\widetilde{W}_s\Big)\mathbf{1}_E\Big]$, by another change of probability in the expectation. Thus, we have shown Equation (8.27). □

By Lemmas 7 and 8, we can assume that the baseline ν_t is continuous with no drift in what follows. We write the $\mathcal{F}_t$-martingale X_t for any time $0 \leq t \leq T$ as

$$X_t = \begin{pmatrix} \Delta_n^{-1} n^{-1}\big(\widehat{\text{Mean}} - \text{Mean}\big)_t \\ \Delta_n^{-\frac{1}{2}} n^{-2}\big(\widehat{\text{Var}}_1 - \text{Var}_1\big)_t \\ \Delta_n^{-\frac{1}{2}} n^{-2}\big(\widehat{\text{Var}}_2 - \text{Var}_2\big)_t \end{pmatrix} = \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \big(\xi_i - \mathbb{E}_{i-1}[\xi_i]\big). \tag{8.28}$$

Here, ξ_i is a vector of dimension 3 defined by

$$\xi_i = \begin{pmatrix} \Delta_n^{-1}\big(\widehat{\nu}_i \Delta_n - \int_{(i-1)\Delta_n}^{i\Delta_n} \nu_t \, dt\big) \\ \Delta_n^{-\frac{1}{2}}\Big((\Delta_i\widehat{\nu})^2 - \mathbb{E}_{i-1}\big[(\Delta_i\widehat{\nu})^2\big] + \mathbb{E}_i\big[(\Delta_{i+1}\widehat{\nu})^2\big] - \int_{(i-1)\Delta_n}^{i\Delta_n} \big(\frac{2}{3}\sigma_t^2 + 2\breve{\vartheta}_t\big) dt\Big) \\ \cdots \end{pmatrix}. \tag{8.29}$$

Moreover, $\mathbf{F}_t$ is the discretized filtration defined as $\mathbf{F}_t = \mathcal{F}_{\Delta_n \lfloor t/\Delta_n \rfloor}$ for any time $t \in [0,T]$. In the vector ξ_i, we do not explicit the third component, which is similar to the second component. The proof of Theorem 1 will be based on an application of Theorem IX.7.28 (pp. 590-591) in Jacod and Shiryaev (2013).

We first show that Condition (7.27) holds with $B_t = 0$ in the following proposition. This proves that the sum of the biases converges to 0 in probability and uniformly over time.

Proposition 3. *We assume that Condition 1 and Condition 2 hold. Then, we have for any $j = 1, 2, 3$ as $n \to \infty$ that*

$$\sup_{0 \le t \le T} \left| \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}_{i-1}[\xi_{i,j}] \right| \xrightarrow{\mathbb{P}} 0. \tag{8.30}$$

Proof of Proposition 3. By Equation (8.15) from Lemma 4, we get

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

Then, we deduce Equation (8.30) in the case $j = 1$ from the martingale definition. By Lemma 6, we obtain as $n \to \infty$ that

$$\begin{aligned} \mathbb{E}_{i-1}[\xi_{i,2}] = {} & \Delta_n^{-\frac{1}{2}} \int_{i-1}^{i+1} \Big(lr_t^2 - \frac{1}{3}\Big) \mathbb{E}_{i-1}[\sigma_t^2]\, dt + \Delta_n^{-\frac{1}{2}} \mathbb{E}_{i-1}\left[u_{i+1}^2 - \frac{\bar{\vartheta}_{i+1}}{n\Delta_n} + u_i^2 - \frac{\bar{\vartheta}_i}{n\Delta_n}\right] \\ & + O_{L^2}\left(\Delta_n^{\frac{1}{2}} \frac{\log n}{n\Delta_n}\right). \end{aligned} \tag{8.31}$$

Since $\int_{i-1}^{i+1} \big(lr_t^2 - \frac{1}{3}\big)\sigma_{i-1}^2\, dt = 0$ and by Condition 2 (h), we obtain as $n \to \infty$ that

$$\sup_{0 \le t \le T} \left| \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \Delta_n^{-\frac{1}{2}} \int_{i-1}^{i+1} \Big(lr_t^2 - \frac{1}{3}\Big) \mathbb{E}_{i-1}[\sigma_t^2]\, dt \right| = o_{\mathbb{P}}(1).$$

For the second term in the right side of Equation (8.31), we first get as $n \to \infty$ that $\big(\frac{\log n}{n^2}\big)^{\frac{2}{7}} = o_{\mathbb{P}}(\Delta_n)$ which holds by Condition 2 (c). Then, we obtain by Lemma 5 as $n \to \infty$ that

$$\sup_{0 \le t \le T} \left| \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \Delta_n^{-\frac{1}{2}} \mathbb{E}_{i-1}\left[u_{i+1}^2 - \frac{\bar{\vartheta}_{i+1}}{n\Delta_n} + u_i^2 - \frac{\bar{\vartheta}_i}{n\Delta_n}\right] \right| = o_{\mathbb{P}}(1).$$

Thus, we can deduce as $n \to \infty$ 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. This proves that the sum of the covariances converges to the asymptotic covariance in probability.

Proposition 4. *We assume that Condition 1 and Condition 2 hold. Then, we have for any time $0 \leq t \leq T$ as $n \to \infty$ that*

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathrm{Cov}_{i-1}[\xi_i] \xrightarrow{\mathbb{P}} \int_0^t w_u w_u^T \, du. \tag{8.32}$$

Proof of Proposition 4. Since $\xi_{i,1} = u_i$ and $\mathbb{E}_{i-1}[u_i] = \underline{O}_{L^2}\big(\frac{\log n}{n\Delta_n}\big)$ as $n \to \infty$, we have by Condition 2 (c) as $n \to \infty$ that

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \big(\mathbb{E}_{i-1}[\xi_{i,1}]\big)^2 = \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \underline{O}_{L^1}\Big(\frac{(\log n)^2}{(n\Delta_n)^2}\Big) = o_{L^1}(1).$$

Thus, Riemann integrability and Condition 2 (c) yield as $n \to \infty$ that

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathrm{Var}_{i-1}[\xi_{i,1}] \xrightarrow{\mathbb{P}} \int_0^t \frac{\mu_u}{(1 - \|\phi\|_1)^3} du.$$

From Equation (8.31), we obtain as $n \to \infty$ that

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

By Equation (8.21) from Lemma 5, Condition 2 (c), (f) and (h), we obtain as $n \to \infty$ that

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

The proof of $\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathrm{Var}_{i-1}[\xi_{i,3}]$ follows with the same arguments if we replace Δ_n by $2\Delta_n$. Since the baseline ν_t is a martingale, this yields as $n \to \infty$ that

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

and $\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathrm{Cov}_{i-1}[\xi_{i,1}, \xi_{i,3}] = o_{\mathbb{P}}(1)$. Finally, we obtain with the same arguments that

$$\sum_{i=1}^{\lfloor t/(2\Delta_n) \rfloor} \mathrm{Cov}_{i-1}[\xi_{i,2}, \xi_{i,3}] \xrightarrow{\mathbb{P}} \frac{1}{2} \int_0^t \frac{29}{24}\breve{\sigma}_u^4 + \frac{3}{2}\breve{\sigma}_u^2 \breve{\vartheta}_u + \frac{3}{2}\breve{\vartheta}_u^2 du.$$

□

We show now that Condition (7.30) holds in the proposition that follows. This proves the Lindeberg condition for the CLT.

Proposition 5. *We assume that Condition 1 and Condition 2 hold. Then, we have for any time $0 \leq t \leq T$ as $n \to \infty$ that*

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}_{i-1}[\|\xi_i\|^2 \mathbf{1}_{\{\|\xi_i\|>\varepsilon\}}] \xrightarrow{\mathbb{P}} 0. \tag{8.33}$$

Proof of Proposition 5. By Hölder's inequality, we have

$$\mathbb{E}\Big|\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}_{i-1}[\|\xi_i\|^2 \mathbf{1}_{\{\|\xi_i\|>\varepsilon\}}\Big| \leq C\sqrt{\frac{\mathbb{E}\|\xi_i\|^4}{\Delta_n^2}}\sqrt{\mathbb{P}\Big\{\frac{\|\xi_i\|}{\Delta_n^{1/2}} > \frac{\varepsilon}{\Delta_n^{1/2}}\Big\}}.$$

Thus, it is sufficient to show as $n \to \infty$ that $\mathbb{E}\|\xi_i\|^4 = O(\Delta_n^2)$, whose proof follows from similar arguments to the ones used in the proof of Proposition 4. □

We show that Condition (7.31) holds in the following proposition. This is based on the proof of Proposition 4.1 (pp. 15-16) in Barndorff-Nielsen et al. (2006) and the proof of Equation (6.10) in Todorov and Tauchen (2011).

Proposition 6. *We assume that Condition 1 and Condition 2 hold. Then, we have for any time $0 \leq t \leq T$ and for any bounded $\mathcal{F}_t$-martingale M' of dimension 3 as $n \to \infty$ that*

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}_{i-1}[\xi_i^T \Delta_i M'] \xrightarrow{\mathbb{P}} 0. \tag{8.34}$$

Proof of Proposition 6. When $M_t' = W_t$, 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 $\mathbb{E}\big(\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}_{i-1}[\xi_{i,2}\Delta_i W]\big)^2 = o(1)$ as $n \to \infty$ from the proof of Proposition 4. We consider now the case when M_t' is a continuous martingale orthogonal to W_t. Since $\mathbb{E}|\sigma_t - \sigma_{i-1}|^k \leq C\Delta_n^{k\gamma}$ for any nonnegative integer $k > 0$, we can approximate locally and replace σ_t^2 in ξ_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 ξ_i', and its conditional expectation as $\xi_t^{(M)} = \mathbb{E}_t[\xi_i]$ for any time $0 \leq t \leq T$. By Theorem III.4.34 (p. 189) in Jacod and Shiryaev (2013), we can express $\xi_t^{(M)}$ into a stochastic integration of W_t and M_t. In particular, the orthogonality of M'_t implies that $(d\xi_t^{(M)})^T(dM'_t) = 0$ for any time $0 \leq t \leq T$. Thus, we can deduce that

$$\mathbb{E}_{i-1}[\xi_i'^T \Delta_i M'_t] = \mathbb{E}_{i-1}\left[\int_{(i-1)\Delta_n}^{i\Delta_n} (d\xi_t^{(M)})^T(dM'_t)\right] = 0.$$

If ξ_i is C-tight and M'_t is a discontinuous martingale orthogonal to $M_t = N_t - \int_0^t \lambda_s \, ds$, then $\sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \xi_i$ and $\sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \Delta_i M'$ are jointly tight. This is obtained by the same arguments as in the proof of Equation (6.10) in Todorov and Tauchen (2011) and Corollary VI.3.33 (p. 353) in Jacod and Shiryaev (2013). Moreover, the left-hand side of Equation (8.34) converges as $n \to \infty$ to the predictable quadratic variation of the limit of $\sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \xi_i$ and $\sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \Delta_i M'$. Then, this limit is zero due to the orthogonality of continuous and discontinuous martingales. Finally, the C-tightness of ξ_i is implied by $\sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \mathbb{E}_{i-1}|\xi_{i,j}|^k \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$ for some nonnegative integer $k > 2$ and this can be derived from the 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 $M'_t = M_t \Delta_n$. □

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

Proposition 7. *We assume that Condition 1 and Condition 2 hold. Then, we have as $n \to \infty$ that*

$$\frac{1}{\Delta_n} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} (\Delta_i \widehat{\nu})^4 \overset{u.c.p.}{\to} \int_0^t \left\{ \frac{4}{3}\sigma_s^4 + 8\sigma_s^2 \vartheta_s + 12\vartheta_s^2 \right\} ds, \tag{8.35}$$

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \widehat{\nu}_i (\Delta_i \widehat{\nu})^2 \overset{u.c.p.}{\to} \int_0^t \left\{ \frac{2}{3}\sigma_s^2 \nu_s + 2\nu_s \vartheta_s \right\} ds, \tag{8.36}$$

$$\Delta_n \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \widehat{\nu}_i^2 \overset{u.c.p.}{\to} \int_0^t \nu_s^2 \, ds. \tag{8.37}$$

Proof of Proposition 7. First, we have that Lemma 7 also holds for powers greater than 2. Thus, we can assume, without loss of generality, that the baseline ν_t is continuous. For Expression (8.35), we can show as $n \to \infty$ that

$$\Delta_n^{-1} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} (\Delta_i \widehat{\nu})^4 = \Delta_n^{-1} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \left\{ (\Delta_i \bar{\nu})^4 + 6(\Delta_i \bar{\nu})^2 (\Delta_i u)^2 + (\Delta_i u)^4 \right\} + o_{\mathbb{P}}(1)$$

$$
\begin{aligned}
&= \Delta_n^{-1} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \Big\{ 6 \int_{i-2}^{i} (lr_s^{(\nu)})^2 \big(dlr_s^{(\nu)}\big)^2 + 6\Big(\int_{i-2}^{i} lr_s^2 \sigma_s^2 \, ds \Big) \Big(\mathbb{E}_{i-1}[u_i^2] + \mathbb{E}_{i-2}[u_{i-1}^2] \Big) \\
&\quad + \Big(\mathbb{E}_{i-1}[u_i^4] + \mathbb{E}_{i-2}[u_{i-1}^4] + 6\mathbb{E}_{i-1}[u_i^2]\mathbb{E}_{i-2}[u_{i-1}^2] \Big) \Big\} + o_{\mathbb{P}}(1) \\
&= \Delta_n^{-1} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \Big\{ 3 \Big(\int_{i-2}^{i} lr_s^2 \sigma_s^2 \, ds \Big)^2 + 6 \Big(\sigma_{i-2}^2 \int_{i-2}^{i} lr_s^2 ds \Big) \Big(\frac{\vartheta_{i-2}}{cn\Delta_n} + \frac{\vartheta_{i-2}}{cn\Delta_n} \Big) \\
&\quad + \Big(3 \Big(\frac{\vartheta_{i-2}}{cn\Delta_n} \Big)^2 + 3 \Big(\frac{\vartheta_{i-2}}{cn\Delta_n} \Big)^2 + 6 \Big(\frac{\vartheta_{i-2}}{cn\Delta_n} \Big)^2 \Big) \Big\} + o_{\mathbb{P}}(1) \\
&\overset{u.c.p.}{\to} \int_0^t \Big\{ \frac{4}{3} \sigma_s^4 + 8\sigma_s^2 \vartheta_s + 12 \vartheta_s^2 \Big\} dt.
\end{aligned}
$$

Here, we use Lemma 6 in the second equality and Condition 2 (c) in the convergence.

For Expression (8.36), we have as $n \to \infty$ that

$$
\begin{aligned}
\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \widehat{\nu}_i (\Delta_i \widehat{\nu})^2 &= \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \bar{\nu}_i \Big((\Delta_i \bar{\nu})^2 + (\Delta_i u)^2 \Big) + o_{\mathbb{P}}(1) \\
&= \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \bar{\nu}_i \Big(\int_{i-2}^{i} lr_s^2 \sigma_s^2 \, ds + \mathbb{E}_{i-1} u_i^2 + \mathbb{E}_{i-2} u_{i-1}^2 \Big) + o_{\mathbb{P}}(1) \\
&= \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \nu_{i-2} \Big(\sigma_{i-2}^2 \int_{i-2}^{i} lr_s^2 \, ds + \frac{\vartheta_{i-2}}{cn\Delta_n} + \frac{\vartheta_{i-2}}{cn\Delta_n} \Big) + o_{\mathbb{P}}(1) \\
&\overset{\mathbb{P}}{\to} \int_0^t \Big\{ \frac{2}{3} \sigma_s^2 \nu_s + 2\nu_s \vartheta_s \Big\} dt.
\end{aligned}
$$

Here, we use Lemma 6 in the second equality and Condition 2 (c) in the convergence.

Finally, we get for Expression (8.37) as $n \to \infty$ that

$$
\Delta_n \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \widehat{\nu}_i^2 = \Delta_n \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \bar{\nu}_i^2 + o_{\mathbb{P}}(1) \overset{u.c.p.}{\to} \int_0^t \nu_s^2 \, ds.
$$

□

In what follows, we deliver the proof of Theorem 1, which is based on an application of Theorem IX.7.28 (pp. 590-591) in Jacod and Shiryaev (2013).

Proof of Theorem 1. We first prove Expression (4.1). This is based on an application of Theorem IX.7.28 (pp. 590-591) in Jacod and Shiryaev (2013). We now verify that all the conditions, namely the five Conditions (7.27) to (7.31) 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 ξ_i is componentwise square-integrable, because $\widehat{\nu}_i$, ν_t, and σ_t have bounded fourth moments by Lemma 3. In addition, we have that Condition (7.27) holds by Proposition 3. Moreover, we can deduce that Condition (7.28) is satisfied with the use of Proposition 4. We also get that Condition (7.30) holds by Proposition 5. Finally, we obtain that Condition (7.31) is satisfied by applying Proposition 6.

We now give the proof of Expression (4.2). First, we have by Expression (4.1) and Condition 2 (c) as $n \to \infty$ that

$$\widehat{\Sigma}_{t,11} \overset{u.c.p.}{\to} \int_0^t \frac{\mu_s}{(1-\|\phi\|_1)^3} ds.$$

For the component $\widehat{\Sigma}_{t,22}$, we obtain by Expression (8.35) from Proposition 7 as $n \to \infty$ that

$$\frac{3}{4\Delta_n} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} (\Delta_i \widehat{\nu})^4 \overset{u.c.p.}{\to} \int_0^t \left(\sigma_s^4 + 6\sigma_s^2 \breve{\vartheta}_s + 9\breve{\vartheta}_s^2\right) dt.$$

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 get by the use of Expression (8.36) from Proposition 7 as $n \to \infty$ that

$$\frac{3}{4\Delta_n} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} (\Delta_i \widehat{\nu})^4 - \frac{3}{1-\|\phi\|_1} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \widehat{\nu}_i (\Delta_i \widehat{\nu})^2 \overset{u.c.p.}{\to} \int_0^t \left(\sigma_s^4 + 4\sigma_s^2 \breve{\vartheta}_s + 3\breve{\vartheta}_s^2\right) ds.$$

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 can deduce by Expression (8.37) from Proposition 7 as $n \to \infty$ that

$$\frac{3}{4\Delta_n} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} (\Delta_i \widehat{\nu})^4 - \frac{3}{1-\|\phi\|_1} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \widehat{\nu}_i (\Delta_i \widehat{\nu})^2 + \frac{9}{(1-\|\phi\|_1)^2} \Delta_n \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \widehat{\nu}_i^2$$
$$\overset{u.c.p.}{\to} \int_0^t \left(\sigma_s^4 + 4\sigma_s^2 \breve{\vartheta}_s + 12\breve{\vartheta}_s^2\right) ds.$$

Since $\frac{2}{3} \frac{\widehat{\text{Var}}_{t,1} - \widehat{\text{Var}}_{t,2}}{\widehat{\text{Mean}}_t} \overset{u.c.p.}{\to} \frac{1}{(1-\|\phi\|_1)^2}$ as $n \to \infty$ and by Condition 2 (c), this yields as $n \to \infty$ that

$$\begin{aligned} \widehat{\Sigma}_{t,22} &= \frac{\Delta_n^4}{c^2} \left(\frac{3}{4} \widehat{\kappa}_{t,4,1} - 3\widehat{\eta}_t \widehat{\kappa}_{t,3,1} + 9\widehat{\eta}_t^2 \widehat{\kappa}_{t,2,1} \right) \\ &\overset{u.c.p.}{\to} \int_0^t \left(\sigma_s^4 + 4\sigma_s^2 \breve{\vartheta}_s + 12\breve{\vartheta}_s^2\right) ds. \end{aligned} \tag{8.38}$$

Moreover, we can show the other components with similar arguments. Finally, we can show Expression (4.4) by an extension of Slutsky's theorem to the case of functional convergence with Expressions (4.1) and (4.2). □

We now give the proofs from Section 5. We start with the proof of Corollary 1. This is obtained by a direct application of Theorem 1 on the first component of the vector X.

Proof of Corollary 1. We focus on the first component of the vector X and the case when the kernel is equal to $\phi(t) = 0$ for any time $t \in \mathbb{R}^+$. Then, a direct application of Theorem 1 on the first component of the vector X gives the corollary. □

We continue with the proof of Corollary 2. This is obtained by applying the delta method (see Theorem 20.8 (p. 297) in Van der Vaart (1998)) to Theorem 1.

Proof of Corollary 2. We focus on the function $f_{t,1}(x_t) = 1 - \sqrt{\frac{3n\text{Mean}_t}{2c(x_2 - x_3)_t}}$ where $x_t = [0, \text{Var}_{t,1}, \text{Var}_{t,2}]^T$. We also focus on the function $f_2(\widehat{x}_t) = 1 - \sqrt{\frac{3\widehat{\text{Mean}}_t}{2\Delta_n^2(\widehat{x}_2 - \widehat{x}_3)_t}}$ where $\widehat{x}_t = [0, \widehat{\text{Var}}_{t,1}, \widehat{\text{Var}}_{t,2}]^T$. Then, we can show that the conditions from Theorem 20.8 (p. 297) in Van der Vaart (1998) applied on the functions $f_{t,1}(x_t)$ and $f_{t,2}(\widehat{x}_t)$ hold. In particular, we can show the Hadamard differentiability. Finally, an application of Theorem 20.8 (p. 297) in Van der Vaart (1998) on the functions $f_{t,1}(x_t)$ and $f_{t,2}(\widehat{x}_t)$ to Theorem 1 yields the corollary. □

We now deliver the proof of Corollary 3. This is also obtained by applying the delta method to Theorem 1.

Proof of Corollary 3. We focus on the function $g_{t,1}(x_t) = \left(\frac{3n\text{Mean}_t}{2c(x_2 - x_3)_t}\right)\left(2x_{t,3} - \frac{1}{2}x_{t,2}\right)$. We also focus on the function $g_{t,2}(x_t) = \left(\frac{3\widehat{\text{Mean}}_t}{2\Delta_n^2(x_2 - x_3)_t}\right)\left(2x_{t,3} - \frac{1}{2}x_{t,2}\right)$. Then, we can show that the conditions from Theorem 20.8 (p. 297) in Van der Vaart (1998) applied on the functions $g_{t,1}(x_t)$ and $g_{t,2}(x_t)$ hold. In particular, we can show the Hadamard differentiability. Finally, an application of Theorem 20.8 (p. 297) in Van der Vaart (1998) on the functions $g_{t,1}(x_t)$ and $g_{t,2}(x_t)$ to Theorem 1 provides the corollary. □

In addition, we provide the proof of Corollary 4. The proof of the convergence under the null hypothesis H_0 is also obtained by applying the delta method to Theorem 1.

Proof of Corollary 4. We focus on the function $f_{t,1}(x_t) = 1 - \sqrt{\frac{3n\text{Mean}_t}{2c(x_2 - x_3)_t}}$. We also focus on the function $f_{t,2}(x_t) = 1 - \sqrt{\frac{3\widehat{\text{Mean}}_t}{2\Delta_n^2(x_2-x_3)_t}}$. Then, we can show that the conditions from Theorem 20.8 (p. 297) in Van der Vaart (1998) applied on the functions $f_{t,1}(x_t)$ and $f_{t,2}(x_t)$ hold. In particular, we can show the Hadamard differentiability. Finally, an application of Theorem 20.8 (p. 297) in Van der Vaart (1998) on the functions $f_{t,1}(x_t)$ and $f_{t,2}(x_t)$ to Theorem 1 yields as $n \to \infty$ that $S_t \xrightarrow{\mathcal{D}} \chi_1^2$ under the null hypothesis H_0. Under the alternative hypothesis H_1, we can show as $n \to \infty$ that $S_t \to \infty$ since $\widehat{AVar}(\widehat{\|\phi\|_{t,1}}) = O_{\mathbb{P}}(1)$. It comes from $\frac{\widehat{\text{Mean}}_t}{\Delta_n^2(\widehat{\text{Var}}_1 - \widehat{\text{Var}}_2)_t} = O_{\mathbb{P}}(1)$ and $\Delta_n^{-2}\widehat{\text{Mean}}_t = O_{\mathbb{P}}(1)$ as $n \to \infty$. ☐

Moreover, we give the proof of Proposition 2. The proof slightly extends the proof of Theorem 1.

Proof of Proposition 2. By Theorem 1, it is sufficient to consider only the components of $\widehat{\text{Var}}_t$. We have

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

For the second term, we can deduce as $n \to \infty$ 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 5 as $n \to \infty$ that $\mathbb{E}_{i-1}[u_i^2 - \Delta_n\breve{\vartheta}_t c^{-1}] = \underline{O}_{L^1}\left(\frac{\log n}{(n\Delta_n)^2}\right)$. Thus, we obtain as $n \to \infty$ that

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

We also have as $n \to \infty$ that

$$\begin{aligned}
\Delta_n^{-1}\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}_{i-1}\Big[(u_i^2 - \Delta_n\breve{\vartheta}_t c^{-1})^2\Big] &= \Delta_n^{-1}\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}_{i-1}\big[u_i^4 - 2\breve{\vartheta}_t c^{-1}\Delta_n u_i^2 + (\breve{\vartheta}_t c^{-1}\Delta_n)^2\big] \\
&= \Delta_n^{-1}\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} 2\left(\frac{\vartheta_t}{cn\Delta_n}\right)^2 + o_{\mathbb{P}}(1)
\end{aligned}$$

$$= 2\frac{\vartheta_t}{cn\Delta_n^2}T + o_{\mathbb{P}}(1).$$

Thus, we obtain $AVar(\widehat{\mathrm{Var}_t}) = 2\vartheta_t T$ by Condition 2 (c). For $ACov(\widehat{\mathrm{Var}_t}, \widehat{\mathrm{Var}_{t,1}})$, we first have that $(\Delta_i\widehat{\nu})^2 = (\Delta_i u)^2$. Then, we obtain as $n \to \infty$ that

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

Thus, we get as $n \to \infty$ that

$$\begin{aligned}
&\Delta_n^{-1}\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}_{i-1}\Big[(u_i^2 - \Delta_n\breve{\vartheta}_t c^{-1})2(u_i^2 - u_{i-1}u_i - \Delta_n\breve{\vartheta}_t c^{-1})\Big] \\
&= 2\Delta_n^{-1}\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}_{i-1}\left[u_i^4 - u_{i-1}u_i^3 - 2\frac{\vartheta_t}{cn\Delta_n}u_i^2 + u_{i-1}u_i\frac{\vartheta_t}{cn\Delta_n} + \left(\frac{\vartheta_t}{cn\Delta_n}\right)^2\right] \\
&= 4\Delta_n^{-1}\sum_{i=1}^{\lfloor t/\Delta_n \rfloor}\left(\frac{\vartheta_t}{cn\Delta_n}\right)^2 + o_{\mathbb{P}}(1) \\
&\xrightarrow{\mathbb{P}} 4\vartheta_t^2 T. \qquad (8.39)
\end{aligned}$$

Here, we use Condition 2 (c) in the convergence. We obtain $Acov(\widehat{\mathrm{Var}_t}, \widehat{\mathrm{Var}_{t,2}}) = \frac{1}{2}\left(\frac{\vartheta_t}{c}\right)^2 t$ 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 satisfied with the same arguments as in the proof of Theorem 1. Thus, we get Equation (5.29). □

Finally, we give the proof of Corollary 5. The proof of the convergence under the null hypothesis H_0 is obtained by applying the delta method to Proposition 2.

Proof of Corollary 5. We focus on the function $f_t(\widehat{x}_{t,1}) = 1 - \sqrt{\frac{3}{2}\frac{\widehat{\mathrm{Mean}_t}}{\Delta_n^2(\widehat{x}_{t,1,3} - \widehat{x}_{t,1,4})}}$ where

$$\widehat{x}_{t,1} = \big(0,\ 0,\ \widehat{\mathrm{Var}_{t,1}},\ \widehat{\mathrm{Var}_{t,2}}\big)^T.$$

We also focus on the function $g_t(\widehat{x}_{t,2}) = 1 - \sqrt{\frac{\widehat{\mathrm{Mean}_t}}{\Delta_n^2\widehat{x}_{t,2,2}}}$ where $\widehat{x}_{t,2} = \big(0,\ \widehat{\mathrm{Var}_t},\ 0,\ 0\big)^T$. Then, we can show that the conditions from Theorem 20.8 (p. 297) in Van der Vaart (1998) applied on the functions $f_t(\widehat{x}_{t,1})$ and $g_t(\widehat{x}_{t,2})$ hold. In particular, we can show the Hadamard differentiability. Finally, an application of Theorem 20.8 (p. 297) in Van der Vaart (1998) on the functions $f_t(\widehat{x}_{t,1})$ and $g_t(\widehat{x}_{t,2})$ to Equation (5.29) to Proposition 2 yields as $n \to \infty$ that $S_t' \xrightarrow{\mathcal{D}} \chi_1^2$ under the null hypothesis $H_{t,0}'$.

Under the alternative hypothesis $H'_{t,1}$, we get as $n \to \infty$ that

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

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

9 Simulation study

In this section, we conduct a simulation study to document how the estimators and tests behave in finite samples.

9.1 Simulation design

We consider the following simulation design to replicate features observed in financial markets. All the models we introduce satisfy the conditions of the theory discussed in the previous sections. 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 for checking the asymptotic approximation and from 10,000 to 1,000,000 for hypothesis testing. 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

$$\lambda_t \; = \; n(1 - \|\phi\|_1)\big(\mu_t^{\mathrm{C}} + \mu_t^{\mathrm{B}}\big) + \int_0^t n\phi(n(t-s)) dN_s. \tag{9.1}$$

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

$$d\mu_t^{\mathrm{C}} \;=\; 30(b_t - \mu_t^{\mathrm{C}})dt + 3\sqrt{\mu_t^{\mathrm{C}}}dW_t. \tag{9.2}$$

The nonrandom function b_t is a solution of the ordinary differential equation $dr_t = 30(b_t - r_t)dt$ with U-shape r_t defined as $r_0 = \mu_0^{\mathrm{C}}$ and

$$r_t = 20\bigg((t - 0.53)^4 + \frac{1}{24}\bigg). \tag{9.3}$$

The tune parameter 0.53 centers the U-shaped component in the middle of the trading day. The drift term in Equation (9.2) ensures mean reversion of the process μ_t^{C} to the function b_t. Moreover, the function b_t pushes the process μ_t^{C} to follow the U-shape nonrandom term r_t. In Equation (9.2), the diffusion term $\sqrt{\mu_t^{\mathrm{C}}}dW_t$ captures the random fluctuation. The Feller condition (Feller, 1951) is satisfied with $30 \times b_t \geq 3^2$ for any time $t \in [0, T]$, thus the process μ_t^{C} is positive.

In Equation (9.2), the process μ_t^{B} correspond to the intensity bursts (Rambaldi et al. (2018)). They are defined as a sudden occurrence of a large number of exogenous points for a short period of time, i.e., around one second. We assume that the bursts are locally bounded. The arrival time of bursts z_i is sampled from a homogeneous Poisson process with rate $2/T$. The size of the bursts Z_i are drawn from $\max(\mathcal{N}(200n, (50n)^2), 50n)$. Thus, we have that the bursts Z_i are positive. The intensity bursts have the form

$$\mu_t^{\mathrm{B}} = \sum_{z_i \leq t} Z_i \mathbf{1}_{\{(t - z_i) \in [0, 1/(3600 \times 6.5)]\}}. \tag{9.4}$$

The parameter values are taken from our empirical application and the results from Rambaldi et al. (2018) (p. 6), where the authors report an average number of bursts between 1.95–3.25 for a 6.5-hour period. With this choice, the intensity remains locally bounded and the conditions of the theory hold.

Kernels are specified as follows. An exponential kernel is 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 empirical application and in the results of Filimonov and Sornette (2012). Finally, we set the truncation parameters as $\alpha = \alpha_0\big(\sum_{i=2}^{\lfloor t/\Delta_n \rfloor} \big(\Delta_i\widehat{\lambda}\big)^2/T\big)^{1/2}$ with

$\alpha_0 = 3$ and $\varpi = 0.48$. The values $\alpha_0 = 3$ and $\varpi = 0.48$ are equal to the values used in Aït-Sahalia and Xiu (2019) and Clinet and Potiron (2019). The value of α is set to α_0 multiplied by the preaveraging of local Poisson estimates without truncation. The idea is similar to the idea used in Aït-Sahalia and Xiu (2019) and Clinet and Potiron (2019). The truncation parameter influences the variance estimation and test calibration. Thus, we have evaluated the sensitivity of test size and power to variations in the parameters ϖ, α, and α_0. However, we do not observe any meaningful change.

We consider the following model variants 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 U-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 U-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 U-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 U-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 U-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 U-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	U-shape	U-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 Equation (9.4), but in the case of the 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, results do not show any significant differences.

Figure 5 provides a comparison between simulated intensity from 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.

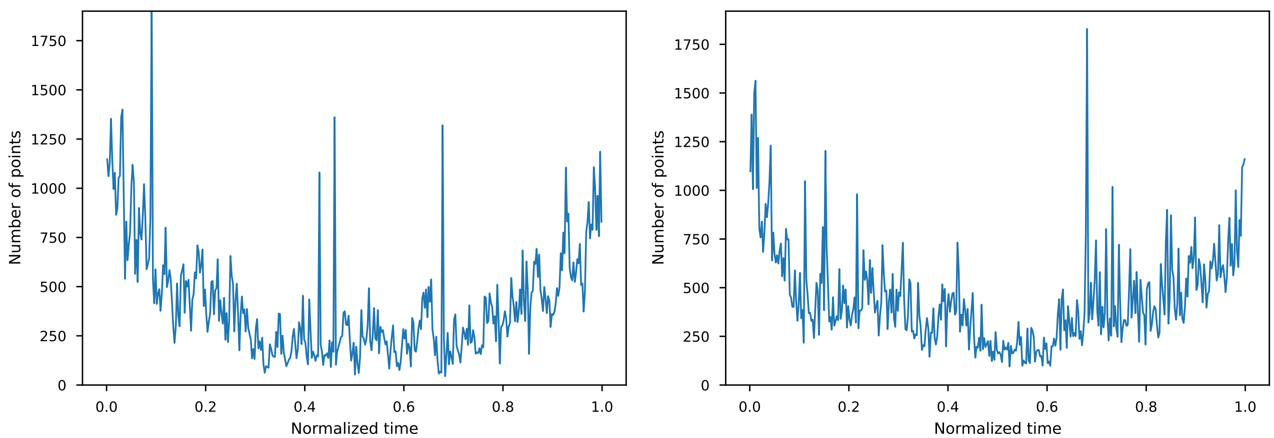

Figure 5: Comparison between simulated intensity with Model 9 (left panel) and intensity based on AAPL data on April 1st 2016 (right panel).

9.2 Asymptotic approximation

Table 2 and Figure 6 report the summary statistics and the histogram for the integrated intensity 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%. 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

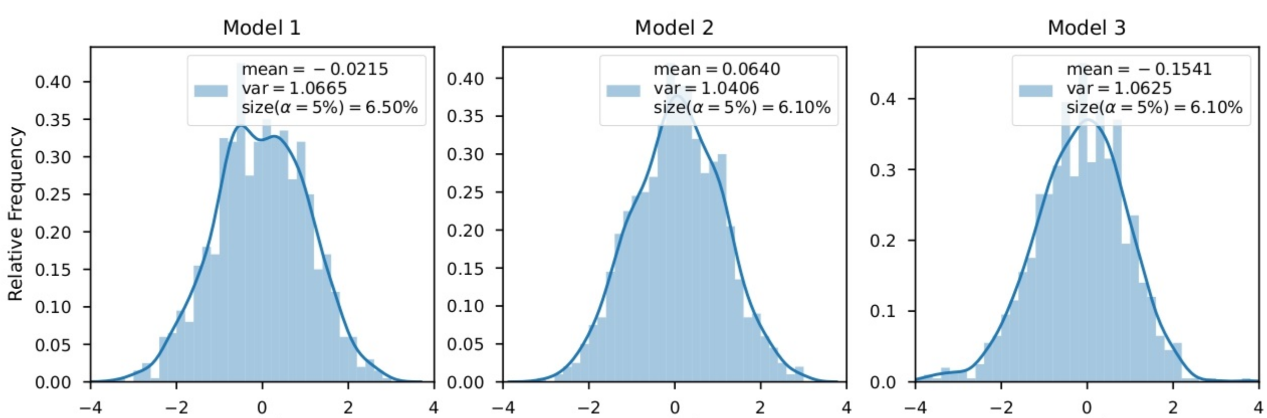

Figure 6: Histogram of the normalized CLT with feasible variance (5.5) for the integrated intensity using Models 1-3. The observation number n is 150,000, and the number of replications is 1,000.

Table 3 and Figure 7 report the summary statistics and the histogram for the integrated baseline using 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 3: Summary statistics for the integrated baseline using 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

Table 4 and Figure 8 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 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

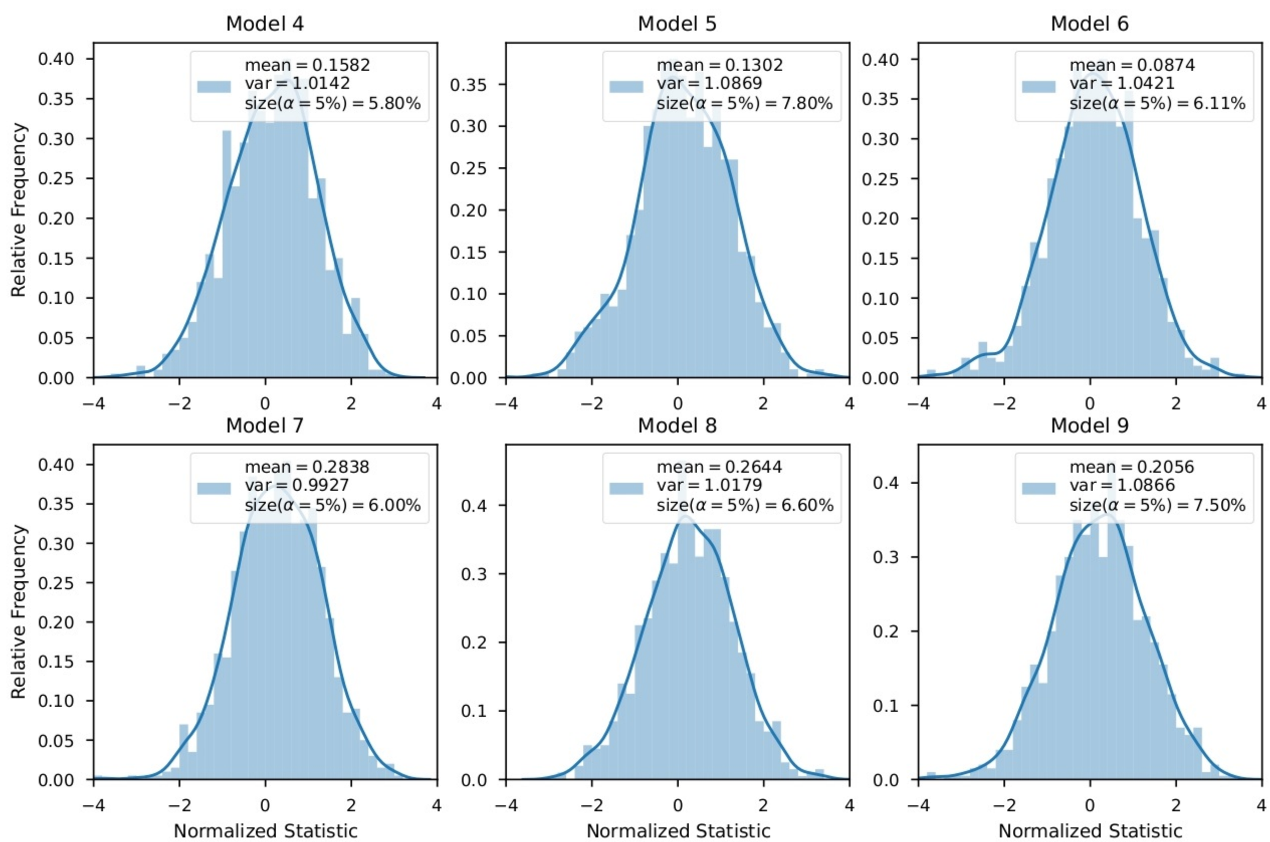

Figure 7: Histogram of the normalized CLT with feasible variance (5.12) for the integrated baseline using Models 4-9. The observation number n is 150,000, and the number of replications is 1,000.

statistics with feasible variance. Overall, the variance is reasonably close to unity.

Table 4: Summary statistics for the integrated volatility of the baseline using 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

9.3 Hypothesis testing

Table 5 reports the percentage of rejections at the 5% level of the null hypothesis for the two tests using Models 1-9. The order of the observation number n is 10,000, 50,000, 150,000, and 1,000,000. The size ranges from 4.2% to 8.7%, with an average of 5.5%. 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 constant baseline. Overall, the test for the absence of a Hawkes component is slightly oversized, while the size of the test for constant baseline is adequate. The power varies between 99.8% and 100% for testing the absence of a Hawkes component and between 62.9% and 100% for testing constant baseline, and thus is more

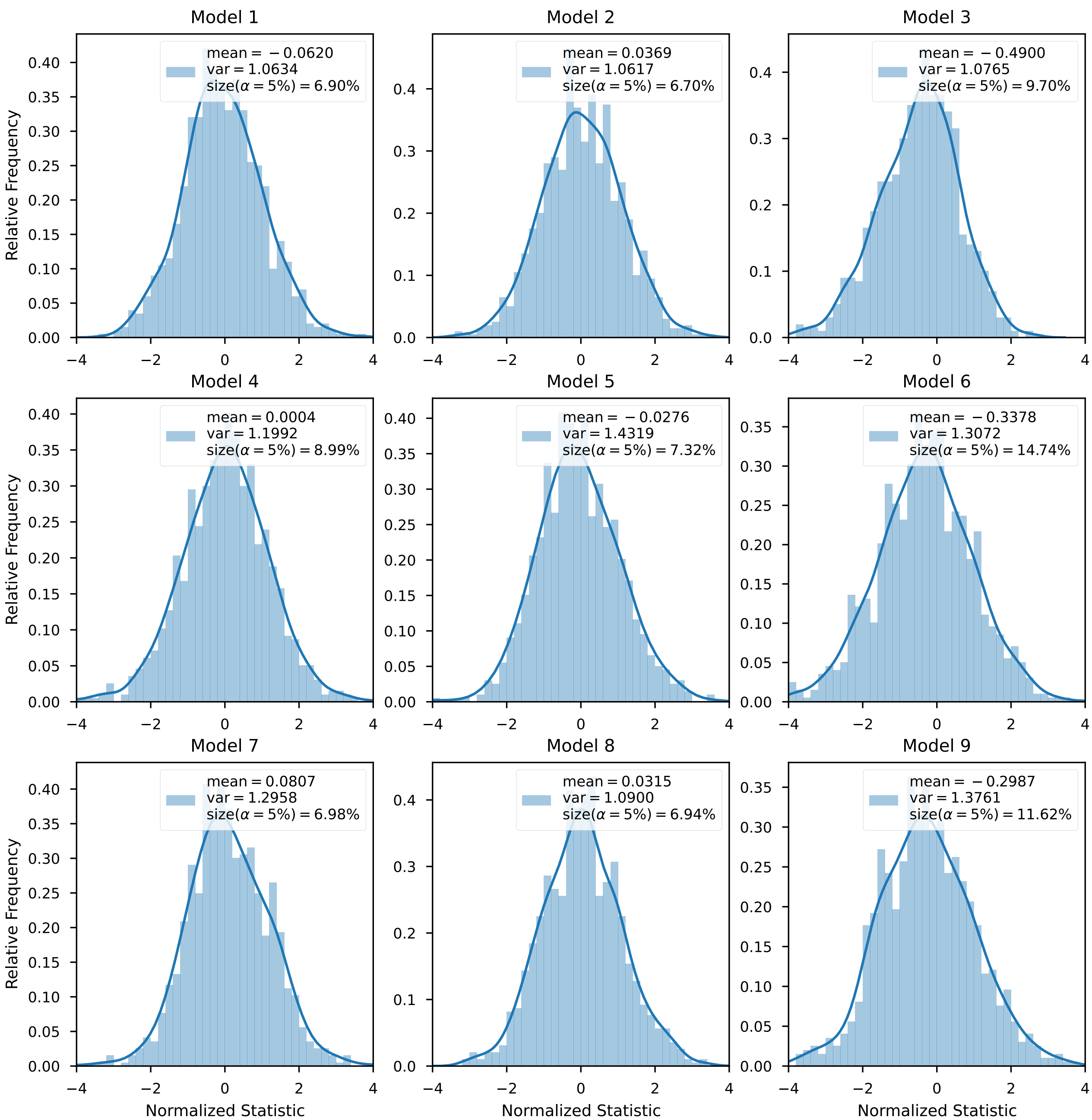

Figure 8: Histogram of the normalized CLT with feasible variance (5.18) for the integrated volatility of the baseline with Models 1-9. The order of the observation number n is 150,000, and the number of replications is 1,000.

adequate for the former test.

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 10,000, 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
10,000	5.7	5.1	8.7	99.9	99.9	99.9	99.8	99.9	99.9
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 constant baseline									
	Size			Power					
n	1	4	7	2	3	5	6	8	9
10,000	5.2	5.1	4.6	99.9	98.3	62.9	87.4	70.3	88.6
50,000	4.2	4.7	4.9	100	100	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

Table 6 reports the percentage of rejections at the 10% level of the null hypothesis for the two tests using Models 1-9. The order of observation number n is 10,000, 50,000, 150,000, and 1,000,000. The size ranges from 9.3% to 14.9%, with an average of 10.9%. It has an average of 12.4% for the test for the absence of a Hawkes component, and an average of 10.3% for testing constant baseline. Overall, the test for the absence of a Hawkes component is slightly oversized, while the size of the test for constant baseline is adequate. The power varies between 74.8% and 100% for testing constant baseline and between 99.8% and 100% for testing the absence of a Hawkes component, and thus is more adequate for the latter test.

Overall, the testing results confirm adequate size and power for both tests for the sample size $n = 50,000$ or above. This number of observations is available in our empirical study.

Table 6: Percentage of rejections at the 10% level of the null hypothesis for the two tests using Models 1-9. The order of n is 10,000, 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
10,000	11.0	10.3	14.9	99.9	99.9	99.8	99.8	99.9	99.9
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 constant baseline									
	Size			Power					
n	1	4	7	2	3	5	6	8	9
10,000	10.2	10.9	10.5	99.9	98.7	74.8	92.2	81.5	93.9
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