An unbounded intensity model for point processes*

Kim Christensen† Aleksey Kolokolov‡

June, 2022

Please do not circulate without permission.

Abstract

In this paper, we develop a model for point processes in which the arrival rate is allowed to explode without inducing an explosion in the point process itself (an intensity burst). In contrast to the standard Poisson postulates, where the probability of more than one event over a short time interval is negligible, the unbounded intensity causes an extreme concentration of points around a particular time instance (an intensity burst time). We derive a novel nonparametric inference for detecting intensity bursts. The asymptotic theory relies on a so-called heavy traffic condition from queueing theory. With Monte Carlo evidence, we show that our testing procedure exhibits size control under the null hypothesis, whereas it has high rejection rates under the alternative. We interpret the model in the context of a financial market and implement it on high-frequency data from leading futures contracts, where our test statistic captures abnormal surges in trading activity, e.g. the level of order submissions and trade executions, which are often observed during times of elevated stress. We detect a nontrivial amount of intensity bursts across asset classes and over time.

JEL Classification: C10; C80.

Keywords: Cox process; heavy traffic; high-frequency data; intensity burst; point process; trading activity.

---

*Christensen received funding from the Independent Research Fund Denmark (DFF 1028–00030B). This work was also supported by CREATES. We are indebted to participants at the "Volatility, Jumps, and Bursts" workshop at Lancaster University, the 2022 annual SoFiE conference at University of Cambridge, and the 33rd annual $(EC)^2$ at ESSEC Business School for helpful remarks.

†Aarhus University, Department of Economics and Business Economics, Denmark. kim@econ.au.dk.

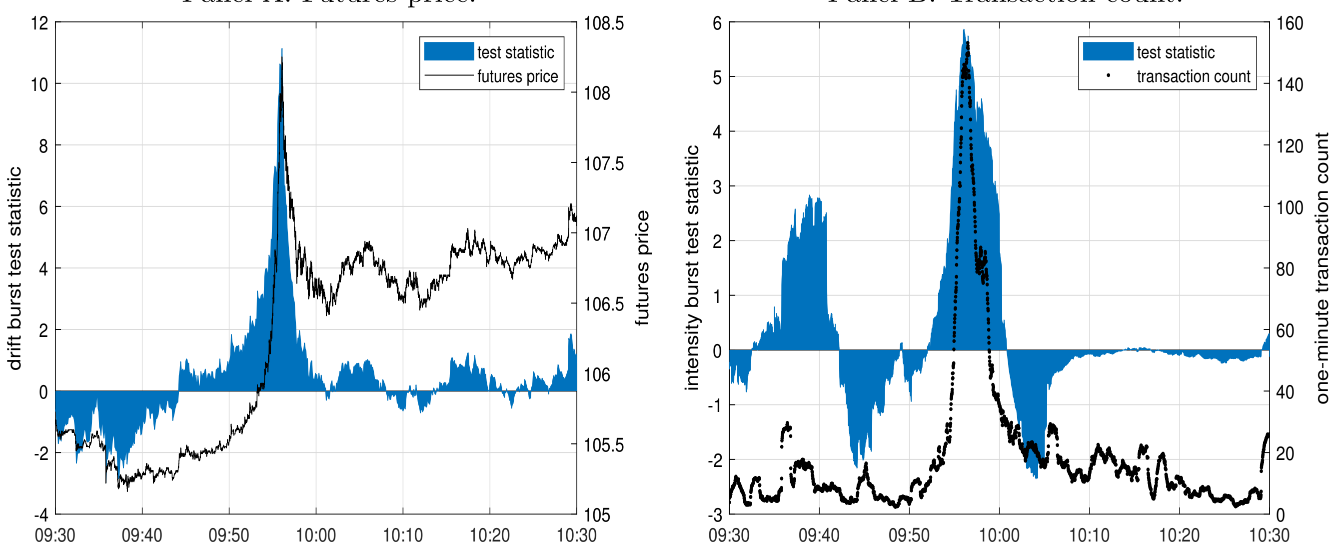
‡Alliance Manchester Business School, University of Manchester, United Kingdom. aleksey.kolokolov@manchester.ac.uk.

1 Introduction

The volatility in financial markets observed during distressed conditions, such as flash crashes, drift bursts, or gradual jumps, typically coincide with abnormal increases in trading activity. For example, in Panel A of Figure 1 we plot a short-lived rally in the delivery price of the most active Crude Oil futures contract (settlement in May 2012) on March 23, 2012 from 9:30am to 10:30am Chicago time. The price of oil accelerated to about 108\$/barrel after a report showed that Iranian oil exports had tumbled about 300,000 barrels per day during the month. Christensen, Oomen, and Renò (2022) propose a drift burst model to explain such rapid and sustained price movements. In their model, the spot drift of a log-price process is allowed to explode, but the log-price itself remains finite. They also develop an econometric approach for detecting such events from noisy high-frequency data. As illustrated in the figure, this event is detected from the transaction price series using their drift burst test statistic, which is highly significant.

Figure 1: Example of an intensity burst in Crude Oil.

Panel A: Futures price. Panel B: Transaction count.



Note. This figure shows the Crude Oil futures contract (May 2012 expiration) over a joint drift and intensity burst on March 23, 2012. In the left panel, we plot the futures price from 9:30am to 10:30am Chicago time along with the drift burst test statistic proposed in Christensen, Oomen, and Renò (2022). In the right panel, we plot a nonparametric estimator of the time-varying trading intensity (measured by the average number of transactions per second) along with the intensity burst test statistic developed in this paper.

In this paper, we extend the concept of bursts to the intensity of a point process. We propose an intensity burst model, which describes an explosion in the intensity of the process. We interpret our model in the context of a financial market, where the flow of market events (e.g., trade arrivals) is modeled by a doubly stochastic Poisson process (i.e. a Poisson process with stochastic intensity of the Cox form) with two independent components: a base intensity describing the "normal" times and an exploding intensity describing "distress". In the normal times, the events are allowed to be inhomogeneous and cluster or exhibit serial correlation. However, since the base intensity of the process is locally bounded, it implies that none of the point clusters observed in a fixed time interval (interpreted as, e.g., a trading day) is substantially larger than the others. The intensity burst component is, in contrast, unbounded in a vicinity of a stopping time producing a pronounced cluster of events in a neighbourhood of that instant.

A realization of a point process observed over a trading day is a configuration of ordered points in a fixed time interval. We propose a nonparametric estimation for intensity burst detection within such data. We study both pointwise identification, that is testing for the presence of an intensity burst in the vicinity of a single candidate time instance, and a uniform identification, that is testing for the presence of at least one intensity burst in a fixed time interval. The inference is done with the help of a novel heavy traffic condition (e.g. Kingman, 1961), where independent copies of an underlying Cox process are stacked.

Our framework allows to independently screen financial high-frequency data for large and unexpected increases in trading activity that may or may not be associated with unusual price volatility. In contrast to Christensen, Oomen, and Renò (2022), our approach is based directly on the arrival times of trades, and it does not require a "mark" in the form of an associated transaction price (or other relevant information). In Panel B of Figure 1, we report a nonparametric estimator of the time-varying trade intensity (measured by the average number of transaction count per one minute) during the above hour. It shows that trading activity increases sharply during the oil price appreciation and then reverts back to its original level soon after. The figure also reports real-time estimates of the intensity burst test statistis developed in this paper. As evident, our new econometric approach also identifies this event as a significant intensity burst.

Our paper is related to a recent strand of papers that study clustering of events for point processes. In the parametric setting, Engle and Russell (1998) propose an autoregressive conditional duration (ACD) model for irregularly spaced transaction data that accommodates persistence in the interarrival times. Clinet and Potiron (2018) propose an inference procedure for the doubly stochastic self-exciting point processes, whereas Potiron and Mykland (2020) study local parametric estimation of high-frequency data. Rambaldi, Pennesi, and Lillo (2015) and Rambaldi, Filimonov, and Lillo (2018) propose a parametric self-exciting process, where an "intensity burst" is associated with an underlying Hawkes process that is perturbed by an exogenous (potentially random) number of points at a (potentially random) time point. While their model is also targeted for capturing sharp accelerations in trading activity, the intensity stays bounded. In contrast, our approach is nonparametric, and we allow the intensity to explode under the burst alternative. Moreover, in our setting the number of burst points can be endogenous and depend on underlying state variables (such as liquidity and volatility), which we describe by an adapted stochastic process in the burst intensity.

The heavy traffic condition is a natural precursor for the infill theory employed in nonparametric estimation of volatility, see, e.g., Jacod and Protter (2012). In that setting, the number of observations of an arbitrage-free price process over a, fixed or shrinking, time interval goes to infinity with the mesh going to zero. Indeed, a realization of our point process falls in the class of stochastic sampling schemes adopted by Hayashi, Jacod, and Yoshida (2011). A common problem with ultra high-frequency data is market microstructure noise, which tends to distort such estimates. In fact, a branch of that literature advocates to estimate volatility via duration-based measures (e.g., Andersen, Dobrev, and Schaumburg, 2008; Hong, Nolte, Taylor, and Zhao, 2021). Apart from developing a statistical test for an intensity burst, we contribute to this field by extending the observed asymptotic variance of Mykland and Zhang (2017) to local estimation (see also Christensen, Podolskij, Thamrongrat, and Veliyev, 2017).

The roadmap of the paper is as follows. Section 2 introduces the unbounded intensity model for point processes and describes the theoretical foundation of an intensity burst. Section 3 develops an identification strategy to conduct inference both for a pointwise test statistic and a maximal test statistic. We also propose a estimator of the asymptotic variance. Section 4 includes an extensive simulation study that demonstrates the finite sample properties of our test. An empirical application is conducted in Section 5, while we conclude the paper in Section 6. The proofs are deferred to the Appendix, where we also present some supplemental empirical results.

2 Intensity burst of a point process

We suppose a random scattering of ordered points is observed on the interval $[0, T]$. It is assumed to be the realization of an underlying point process $N = (N_t)_{t \geq 0}$, which is defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ and adapted to the filtration. In this paper, we assume N to be a doubly stochastic Poisson—or Cox—process with associated random intensity—or rate—process $\lambda = (\lambda_t)_{t \geq 0}$, where λ is an adapted and strictly positive real-valued stochastic process. That is, conditional on λ, N is an inhomogeneous Poisson process with rate function λ, i.e. the conditional characteristic function of the increment $N_{t+\Delta} - N_t$ is given by

$$\varphi_{N_{t+\Delta} - N_t}(u) = \mathbb{E}\Big(e^{iu(N_{t+\Delta} - N_t)} \mid \mathcal{F}_t^{\lambda}\Big) = \exp\left((e^{iu} - 1) \int_t^{t+\Delta} \lambda_s \mathrm{d}s\right), \tag{1}$$

where $\mathcal{F}_t^{\lambda} = \sigma(\{\lambda_s; s \leq t\})$.

λ_t can be thought of as the expected number of points arriving over the next short time interval $[t, t + \Delta]$, based on historical information about the rate process, since

$$\lambda_t = \lim_{\Delta \to 0} \frac{\mathbb{E}\big[N_{t+\Delta} - N_t \mid \mathcal{F}_t^{\lambda}\big]}{\Delta}. \tag{2}$$

Hence, when λ_t is locally bounded, the instantaneous expectation of the number of points is finite. In order to define intensity bursts, we therefore suppose that λ_t is locally unbounded in the vicinity of a particular time instant, such that the expectation of the number of points in the vicinity of that time point is locally infinite.

Assumption 1. *λ_t can be decomposed as follows:*

$$\lambda_t = \mu_t + \beta_t, \tag{3}$$

where $\mu = (\mu_t)_{t\geq 0}$ is a locally bounded and strictly positive real-valued stochastic process, and

$$\beta_t = \frac{\sigma_t}{|\tau_{\text{ib}} - t|^{\alpha}}, \tag{4}$$

for a stopping time τ_{ib}, where $\sigma = (\sigma_t)_{t\geq 0}$ is a non-negative locally bounded stochastic process and $0 < \alpha < 1$ is a constant.

In the context of an financial market, N may describe the number of order submissions or trade executions in a security over the course of a trading session. With this interpretation, Assumption 1 extends the notion of a drift burst in the asset log-price from Christensen, Oomen, and Renò (2022) to the intensity generating the log-price. μ is the arrival rate of trades during "normal" market conditions, whereas β represents the arrival rate of trades during "distressed" market conditions. The stopping time τ_{ib} is called the intensity burst time.

If $\sigma_t = 0$, almost surely for all t, the explosive term β_t is absent and N_t again evolves as a Cox process. If $\sigma_t > 0$, almost surely in a neighbourhood of τ_{ib},

$$\int_{\tau_{\text{ib}}-\Delta}^{\tau_{\text{ib}}+\Delta} \mu_s \text{d}s = O_p(\Delta) \qquad \text{and} \qquad \int_{\tau_{\text{ib}}-\Delta}^{\tau_{\text{ib}}+\Delta} \beta_s \text{d}s = O_p\big(\Delta^{1-\alpha}\big), \tag{5}$$

as $\Delta \to 0$.

Consequently, the expected number of points explodes in a neighborhood of τ_{ib}. However, due to the integrability condition on β_t in Assumption 1, $\int_0^T \lambda_s \text{d}s < \infty$, so N itself is non-explosive and well-defined.

It is well-known that one cannot consistently estimate the intensity of a point process over a finite time interval, not even in the homogenous case. Hence, inference of point processes usually proceeds under long-span asymptotics ($T \to \infty$). However, as we are interested in identifying intensity bursts over small time intervals, we follow the convention in the high-frequency literature and assume T is fixed. Instead, we impose an alternative "heavy traffic" assumption, in which the intensity of the underlying point process diverges (e.g., Kingman, 1961). To formalize this idea, we introduce an auxiliary parameter n and consider a sequence of Cox processes $N^n = (N_t^n)_{t\geq 0}$, for $n = 1, 2, \dots$. For a fixed n, the rate function of N_t^n is equal to $n\lambda_t$, and the observed configuration of points is then a realization of N^n. The asymptotic theory is then developed by supposing that $n \to \infty$.

To guarantee the existence of the required sequence of Cox processes, N_t^n can conveniently be constructed as follows:

$$N_t^n = \sum_{i=1}^{n} N_{t,i}, \tag{6}$$

where $(N_{t,i})_{i=1}^n$ are independent copies of N_t.

The above assumption can be motivated by a financial market populated with n economic agents, each trading independently according to their optimal strategy, i.e. an adapted stochastic process represented by $(\lambda_t)_{t\geq 0}$. Our asymptotic theory should then be understood as letting the number of agents increase without bound ($n \to \infty$).

The heavy traffic condition is intimately connected with the literature on high-frequency estimation of volatility. Inference on realized variance typically proceeds under infill asymptotics (Andersen and Bollerslev, 1998; Barndorff-Nielsen and Shephard, 2002). In that setting, we suppose a log-price process is observed at $m = m(n)$ discrete time points $t_0, t_1, \dots, t_m$ that partition the time interval $[0, T]$, where $m \xrightarrow{p} \infty$ such that $\sup_i |t_i - t_{i-1}| \xrightarrow{p} 0$ as $m \to \infty$. Thus, the point process N^n together with the heavy traffic assumption is a standard model for generating stochastic sampling times satisfying the usual conditions solicited in financial econometrics. In fact, a realization of N^n falls within the class of stochastic sampling schemes studied by Hayashi, Jacod, and Yoshida (2011). Hence, heavy traffic is a natural precursor for analysis of volatility from financial high-frequency data.

The detection of intensity bursts amounts to the following hypothesis:

$$\mathscr{H}_0 : N_t^n(\omega) \in \Omega_0 \qquad \text{and} \qquad \mathscr{H}_1 : N_t^n(\omega) \in \Omega_1, \tag{7}$$

where Ω_0 and Ω_1 are complementary subsets of Ω:

$$\Omega_0 = \left\{ \omega \in \Omega : \int_0^T \beta_t^2 \mathrm{d}t = 0 \right\} \qquad \text{and} \qquad \Omega_1 = \left\{ \omega \in \Omega : \int_0^T \beta_t^2 \mathrm{d}t > 0 \right\}. \tag{8}$$

We propose a testing procedure to figure out which subset the realization of N_t^n belongs to, i.e. whether it has an intensity burst component ($N_t^n(\omega) \in \Omega_1$) or not ($N_t^n(\omega) \in \Omega_0$).

3 Identification

In this section, we now develop a nonparametric approach to detect intensity bursts. First, we propose a pointwise test statistic that conducts a statistical test for the presence of an intensity burst at a single candidate time instance. Second, we refine this to a uniform test statistic that conducts a test for the presence of at least one intensity burst in $[0, T]$.

3.1 Pointwise test

The identification of intensity bursts is based on a backward-looking estimator of the local intensity, to allow for online detection, which is defined as:

$$\widehat{\lambda}_t = \frac{N^n(t - \delta_n, t)}{n\delta_n}, \tag{9}$$

where $N^n(a, b) = N_b^n - N_a^n$ is the number of points of the counting process on $(a, b]$, and δ_n is a bandwidth parameter. $\widehat{\lambda}_t$ is the "realized intensity," or average number of counts per time unit over an interval of length δ_n. In a setup with no heavy traffic (i.e. n fixed), it follows that $\widehat{\lambda}_t \xrightarrow{p} 0$. However, heavy traffic changes the stochastic limit of $\widehat{\lambda}_t$.

Lemma 1. *Suppose that Assumption 1 holds. Then, under $\mathscr{H}_0$, as $n \to \infty$ and $\delta_n \to 0$ such that $n\delta_n \to \infty$, it holds for all fixed $t \in [0, T]$ that*

$$\widehat{\lambda}_t \xrightarrow{p} \lambda_t. \tag{10}$$

Moreover, under $\mathscr{H}_1$,

$$\widehat{\lambda}_{\tau_{\text{ib}}} = O_p\big(\delta_n^{-\alpha}\big). \tag{11}$$

Lemma 1 shows that, under the null hypothesis and very general assumptions about the intensity process, the intensity estimator measures the local intensity at time t. The result is rather intuitive. To generate a sufficient number of observations inside the estimation window $[t - \delta_n, t]$, we need $\delta_n n \to \infty$. As $n \to \infty$, the summation of independent copies of the Cox process in the heavy traffic limit ensures that, on average, the accumulation of points close to t correspond approximately to the instantaneous arrival rate. A law of large numbers then shows that $\widehat{\lambda}_t$ converges in probability and, as $\delta_n \to 0$, the limit is λ_t. On the other hand, under the alternative, the abnormal arrival rate of the bursting process causes the estimator to explode around τ_{ib}.

The lemma highlights an important difference between the estimation of the local intensity of point processes and the estimation of the local drift of Brownian semimartingales. In the latter, the drift estimator is asymptotically unbiased but inconsistent (Kristensen, 2010), because the increments of a Brownian motion, although mean zero, exhibit so much variation over short time intervals that any information about the drift is lost. This is exploited by Christensen, Oomen, and Renò (2022) to construct a drift burst test statistic, which relies on the different rates of divergence of the drift estimator under the null and alternative. By contrast, under heavy traffic $\widehat{\lambda}_t$ is consistent for the local intensity of the point process under the null hypothesis. Hence, for intensity burst detection we propose a slightly different test statistic compared to Christensen, Oomen, and Renò (2022). In particular, we compare two local intensity estimators calculated from the nearest lagged block of observations. That is, our test is based on the difference

$$\nabla\widehat{\lambda}_t = \widehat{\lambda}_t - \widehat{\lambda}_{t-\delta_n}. \tag{12}$$

Under $\mathscr{H}_0$, $\nabla\widehat{\lambda}_t$ converges in probability to zero, but it is unbounded in probability under $\mathscr{H}_1$, as we prove below. Hence, one can derive an asymptotic confidence interval for $\nabla\widehat{\lambda}_t$ under $\mathscr{H}_0$ and reject $\mathscr{H}_0$ if $\nabla\widehat{\lambda}_t$ is outside of it. This suggests a standard test statistic of the form:

$$\phi_t^{\text{ib}} = \frac{\nabla\widehat{\lambda}_t}{\widehat{\text{std}}(\nabla\widehat{\lambda}_t)}, \tag{13}$$

where $\widehat{\text{std}}(\nabla\widehat{\lambda}_t)$ is an estimator of $\text{std}(\nabla\widehat{\lambda}_t)$.

It turns out, however, that the asymptotic variance of $\widehat{\lambda}_t$ depends crucially on the interplay between δ_n and n, which follows from a central limit theorem (CLT) under $\mathscr{H}_0$. To derive this CLT, we need a regularity condition on the baseline intensity μ of N.

Assumption 2. *μ is a continuous Itô semimartingale:*

$$\mu_t = \mu_0 + \int_0^t a_s \mathrm{d}s + \int_0^t \nu_s \mathrm{d}W_s, \tag{14}$$

where μ_0 is $\mathcal{F}_0$-measurable, $a = (a_t)_{t\geq 0}$ and $\nu = (\nu_t)_{t\geq 0}$ are adapted, càdlàg stochastic processes, and W is a standard Brownian motion.

Assumption 2 is common in the high-frequency literature (e.g. Jacod and Protter, 2012). It allows to apply standard estimates for semimartingales to control the discretization error in the proofs. In particular, for a constant $C > 0$, it follows that:

$$|\mathbb{E}[\mu_t - \mu_{t+\Delta} \mid \mathcal{F}_t]| \leq C\Delta \qquad \text{and} \qquad \mathbb{E}[|\mu_t - \mu_{t+\Delta}|^r \mid \mathcal{F}_t] \leq C\Delta^{r/2}, \tag{15}$$

for some $r > 0$. This assumption can be replaced, for example, by an appropriate smoothness conditions, such as assuming the paths of μ are Hölder continuous up to some order.

Theorem 1. *Suppose that Assumptions 1 and 2 hold. Then, under $\mathscr{H}_0$, as $n \to \infty$ and $\delta_n \to 0$ such that $n\delta_n \to \infty$, it holds for all fixed $t \in [0, T]$ that*

(i) if $\delta_n\sqrt{n} \to 0$,

$$\sqrt{n\delta_n}\left(\widehat{\lambda}_t - \mu_t\right) \xrightarrow{\mathfrak{D}_s} \sqrt{\mu_t}Z, \tag{16}$$

and

$$\sqrt{n\delta_n}\nabla\widehat{\lambda}_t \xrightarrow{\mathfrak{D}_s} \sqrt{2\mu_t}Z, \tag{17}$$

(ii) if $\delta_n\sqrt{n} \to \theta > 0$,

$$\frac{1}{\sqrt{\delta_n}}\left(\widehat{\lambda}_t - \mu_t\right) \xrightarrow{\mathfrak{D}_s} \sqrt{\frac{1}{\theta^2}\mu_t + \frac{1}{3}\nu_t^2}Z, \tag{18}$$

and

$$\frac{1}{\sqrt{\delta_n}}\nabla\widehat{\lambda}_t \xrightarrow{\mathfrak{D}_s} \sqrt{\frac{2}{\theta^2}\mu_t + \frac{8}{3}\nu_t^2}Z, \tag{19}$$

(iii) if $\delta_n\sqrt{n} \to \infty$,

$$\frac{1}{\sqrt{\delta_n}}\left(\widehat{\lambda}_t - \mu_t\right) \xrightarrow{\mathfrak{D}_s} \sqrt{\frac{1}{3}\nu_t^2}Z, \tag{20}$$

and

$$\frac{1}{\sqrt{\delta_n}}\nabla\widehat{\lambda}_t \xrightarrow{\mathfrak{D}_s} \sqrt{\frac{8}{3}\nu_t^2}Z, \tag{21}$$

where Z is a standard normal random variable, which is defined on an extension of $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and independent of $\mathcal{F}_t$, and $\xrightarrow{\mathfrak{D}_s}$ is stable convergence in law (see, e.g., Jacod and Protter, 2012).

Theorem 1 shows that the asymptotic distribution of our estimator is always mixed normal, but the convergence rate and limiting variance depends on the order of δ_n and n. On the one hand, if

$\delta_n\sqrt{n} \to 0$, the localization dominates and the variation of the intensity parameter along its sample path is immaterial. Hence, the variance is the mean. On the other hand, if $\delta_n\sqrt{n} \to \infty$, the roles are reversed, and heavy traffic dominates. In that case, the variation of the stochastic intensity parameter controls the asymptotic variance. In both cases, the rate of convergence can be rather slow. The optimal convergence rate, $n^{-1/4}$, is achieved with $\delta_n \asymp n^{-1/2}$, where the opposing forces are balanced.

Note that the standard errors of $\nabla\widehat{\lambda}_t$ appearing in the second part of (ii) and (iii) are larger than one would expect when comparing with (i). The explanation is that in the latter settings, n diverges sufficiently fast compared to the vanishing of δ_n, so the lagged estimator $\widehat{\lambda}_{t-\delta_n}$ is much more imprecise compared to $\widehat{\lambda}_t$. This effect is not present in part (i) of the theorem.

3.2 Observed asymptotic local variance

In practice, the choice of bandwidth—and picking the correct estimator of the asymptotic variance—is made difficult, because n is not observed. To construct a feasible test statistic, we follow Mykland and Zhang (2017) by alluding to the notion of the observed asymptotic variance.[1] We set $\delta_n = \ell_n\Delta_n$, where ℓ_n is a deterministic sequence of natural numbers and $\Delta_n > 0$ represents a small time interval. We further impose that $\Delta_n = n^{-1}$, but this is merely done for notational convenience. Nothing changes if $\Delta_n = O(n^{-1})$, except we introduce an additional tuning parameter.

$\widehat{\lambda}_t$ can then be rewritten:

$$\widehat{\lambda}_t = \frac{1}{\ell_n} \sum_{i \in \mathcal{D}^n_{t-}} \Delta_i N^n, \tag{22}$$

where $\Delta_i N^n = N^n(i\Delta_n, (i-1)\Delta_n)$ is the increment of N^n_t over a short time interval and $\mathcal{D}^n_{t-} = \{tn - \ell_n + 1, tn - \ell_n + 2, \dots, tn\}$ is a set of time indexes. Equation (22) shows that the intensity estimator can be expressed as a local average of increments of the discretized point process N^n_t, which is a more convenient formulation for developing our asymptotic variance estimator.[2] With this notation, Theorem 1 can be reformulated as follows.

Corollary 1. *Suppose that Assumptions 1 and 2 hold. Then, under $\mathscr{H}_0$, as $n \to \infty$ and $\ell_n \to \infty$ such that $\ell_n\Delta_n \to 0$, it holds for all fixed $t \in [0,T]$ that*

$$\sqrt{\rho_n}\nabla\widehat{\lambda}_t \xrightarrow{\mathfrak{D}_s} \sqrt{\mathsf{avar}(\nabla\widehat{\lambda}_t)}Z, \tag{23}$$

[1] Our automatic inference procedure can also exploit the subsampling approach of Politis, Romano, and Wolf (1999), which was adapted to the high-frequency setting in Kalnina (2011) and extended in Christensen, Podolskij, Thamrongrat, and Veliyev (2017).

[2] In our framework, N^n is observed in continuous-time. However, many existing datasets do not reveal the exact location of all points of N^n within $[0,T]$, but only report discrete observations. For example, the number of trades in every 10-second interval may be available. In such cases, the data is structured in the form of a discretized version of N^n used in equation (22). Thus, Corollary 1 and subsequent theorems remains applicable to such data.

where

$$\left(\rho_n, \mathsf{avar}\big(\nabla\widehat{\lambda}_t\big)\right) = \begin{cases} \left(\ell_n,\ 2\mu_t\right), & \text{if } \ell_n\sqrt{\Delta_n} \to 0, \\ \left((\ell_n\Delta_n)^{-1},\ \frac{2}{\theta^2}\mu_t + \frac{8}{3}\nu_t^2\right), & \text{if } \ell_n\sqrt{\Delta_n} \to \theta > 0, \\ \left((\ell_n\Delta_n)^{-1},\ \frac{8}{3}\nu_t^2\right), & \text{if } \ell_n\sqrt{\Delta_n} \to \infty. \end{cases} \tag{24}$$

The observed asymptotic local variance of $\nabla\widehat{\lambda}_t$ is then defined as:

$$\widetilde{\mathsf{avar}}\big(\nabla\widehat{\lambda}_t\big) = \frac{\rho_n}{K_n}\sum_{j=0}^{K_n-1}\left(\widehat{\lambda}_{t-2j\ell_n\Delta_n} - \widehat{\lambda}_{t-(\ell_n+2j\ell_n)\Delta_n}\right)^2, \tag{25}$$

where K_n is another sequence of natural numbers.

$\widetilde{\mathsf{avar}}\big(\nabla\widehat{\lambda}_t\big)$ is the sample variance of K_n estimates computed over non-overlapping blocks consisting of $2\ell_n$ observations. When K_n is sufficiently large and the observation blocks are not too distant from t, $\widetilde{\mathsf{avar}}\big(\nabla\widehat{\lambda}_t\big)$ is consistent for the asymptotic local variance $\mathsf{avar}\big(\nabla\widehat{\lambda}_t\big)$.

Theorem 2. *Suppose that Assumptions 1 and 2 hold. Then, under $\mathscr{H}_0$, as $n \to \infty$, $\ell_n \to \infty$, $K_n \to \infty$ such that $\ell_n\Delta_n \to 0$ and $\ell_n K_n \Delta_n \to 0$, it holds for all fixed $t \in [0, T]$ that*

$$\widetilde{\mathsf{avar}}\big(\nabla\widehat{\lambda}_t\big) \xrightarrow{p} \mathsf{avar}\big(\nabla\widehat{\lambda}_t\big). \tag{26}$$

That is, $\widetilde{\mathsf{avar}}\big(\nabla\widehat{\lambda}_t\big)$ converges in probability to $\mathsf{avar}\big(\nabla\widehat{\lambda}_t\big)$ for any limiting behaviour of ρ_n allowed in Corollary 1.

The condition $K_n\ell_n\Delta_n \to 0$ in Theorem 2 is somewhat restrictive. However, it can be loosened with overlapping blocks in the definition of the observed asymptotic local variance. Hence, an alternative version of the estimator is the following:

$$\widehat{\mathsf{avar}}\big(\nabla\widehat{\lambda}_t\big) = \frac{\rho_n}{K_n}\sum_{j=0}^{K_n-1}\left(\widehat{\lambda}_{t-j\Delta_n} - \widehat{\lambda}_{t-(\ell_n+j)\Delta_n}\right)^2. \tag{27}$$

Theorem 3. *Suppose that Assumptions 1 and 2 hold. Then, under $\mathscr{H}_0$, as $n \to \infty$, $\ell_n \to \infty$, $K_n \to \infty$ such that $\ell_n\Delta_n \to 0$, $K_n\Delta_n \to 0$ and $\ell_n/K_n \to 0$, it holds for all fixed $t \in [0, T]$ that*

$$\widehat{\mathsf{avar}}\big(\nabla\widehat{\lambda}_t\big) \xrightarrow{p} \mathsf{avar}\big(\nabla\widehat{\lambda}_t\big). \tag{28}$$

The condition $\ell_n/K_n \to 0$ in Theorem 3 indicates that the number of differences used to estimate the local asymptotic variance of $\nabla\widehat{\lambda}_t$ should increase faster than the number of observations used to compute the local intensity estimate. The conditions $K_n\Delta_n \to 0$ and $\ell_n\Delta_n \to 0$ ensure that all observations remain close to t. Thus, the test statistic in (13) is computed as:

$$\phi_t^{\text{ib}} = \frac{\sqrt{\rho_n}\nabla\widehat{\lambda}_t}{\sqrt{\widehat{\mathsf{avar}}\big(\nabla\widehat{\lambda}_t\big)}}. \tag{29}$$

Theorem 4. *Suppose that the conditions of Theorems 1 and 3 hold. Then,*

$$\begin{cases} \phi_t^{\text{ib}} \xrightarrow{\mathfrak{D}} N(0,1), & \textit{conditional on } \mathscr{H}_0, \\ \phi_{\tau_{\text{ib}}}^{\text{ib}} \xrightarrow{p} \infty, & \textit{conditional on } \mathscr{H}_1. \end{cases} \tag{30}$$

Hence, we reject $\mathscr{H}_0$ if ϕ_t^{ib} exceeds a quantile in the right-hand tail of the standard normal distribution corresponding to a given significance level α. This ensures the test has size control under $\mathscr{H}_0$. On the other hand, under $\mathscr{H}_1$, ϕ_t^{ib} is unbounded in probability as $t \to \tau_{\text{ib}}$, so the test is also consistent.[3]

It is important to emphasize that we do not need to choose ρ_n for computing ϕ_t^{ib}. This follows by direct insertion, showing the test statistic can be reexpressed as:

$$\phi_t^{\text{ib}} = \frac{N^n(t-\delta_n, t) - N^n(t-2\delta_n, t-\delta_n)}{\sqrt{\dfrac{1}{K_n}\displaystyle\sum_{j=0}^{K_n-1}\big(N^n(t-j\Delta_n-\delta_n, t-j\Delta_n) - N^n(t-j\Delta_n-2\delta_n, t-j\Delta_n-\delta_n)\big)^2}}. \tag{31}$$

In practice, we take Δ_n to be one second by convention, although this is immaterial. The choice of K_n and ℓ_n are studied in the Monte Carlo analysis presented in Section 4.

3.3 Maximum test

The pointwise test statistic explodes when the test time approaches an intensity burst. In practice, this suggests that identification of intensity bursts consists of computing ϕ_t^{ib} over a grid of points $0 \leq t_1^* < t_2^* < \cdots < t_m^* \leq T$, where m eventually diverges (at a slow enough rate) such that $\max_{1\leq i\leq m} t_i^* - t_{i-1}^* \to 0$. However, this has an important drawback, because the multiple testing leads to size distortions if $\{\phi_{t_i^*}^{\text{ib}}\}_{i=1}^m$ are evaluated against the critical values from the standard normal distribution.

Set against this backdrop, as in Christensen, Oomen, and Renò (2022) we propose to look at a single statistic from the sequence $\{\phi_{t_i^*}^{\text{ib}}\}_{i=1}^m$. In particular, we follow Lee and Mykland (2008) and extract the maximum $\phi^{\text{ib},\star} = \max_{1\leq i\leq m}\phi_{t_i^*}^{\text{ib}}$. Observe that in the absence of an intensity burst, $\{\phi_{t_i^*}^{\text{ib}}\}_{i=1}^m$ is asymptotically a centered, unit variance Gaussian process with unknown correlation structure. However, the serial dependence of the intensity burst test statistic is complicated, even if the underlying point process is a homogeneous Poisson process. As as remedy, we estimate the sample autocovariance function of $\{\phi_{t_i^*}^{\text{ib}}\}_{i=1}^m$ up to lag $2\ell_n$ and then simulate a Gaussian process with the estimated covariance structure by circulant embedding. In each simulation, we record the maximum value of the process. Repeating this procedure a large number of times, we calculate the critical value for our test statistic based on the distribution of the simulated maxima.

4 Simulation study

In this section, we investigate the small sample size and power of our intensity burst test statistic by a Monte Carlo approach. On the interval $[0, 1]$, we simulate a continuous-time realization of a convolution of independent counting processes representing the "normal" and "burst" intensity,

[3] Theorem 4 continues to hold if $\widehat{\mathsf{avar}}(\nabla\widehat{\lambda}_t)$ is replaced with $\widetilde{\mathsf{avar}}(\nabla\widehat{\lambda}_t)$ in the definition of ϕ_t^{ib}, provided the conditions of Theorem 2 hold.

$N_t = N_t^{\text{normal}} + N_t^{\text{burst}}$. We interpret the unit interval as a standard 6.5 hour trading day on a U.S. stock exchange, which we partition with a discretization step of $\mathrm{d}t = 1/(23{,}400 \times 100)$, corresponding to one-hundredth of a second.

To generate a challenging data-generating process under the null, we assume the intensity of N_t^{normal} is an exponential Hawkes process:

$$\mu_t = \lambda_0 + \int_{-\infty}^{t} \theta e^{-\kappa(t-s)} \mathrm{d}N_s^{\text{normal}}, \tag{32}$$

where λ_0, θ and κ are parameters. λ_0 is called the background intensity, which is a lower bound on μ_t, while $f(\tau) = \theta e^{-\kappa\tau}$ is the excitation function. Our choice of exponential kernel follows the original article by Hawkes (1971).[4]

A Hawkes process is self-exiting and capable of generating event clusters. That is, after the arrival of an event the intensity of N_t^{normal} inclines by θ. Hence, the probability of an increment in the next time interval $t + \mathrm{d}t$ increases. Tempering by the excitation function helps to pull the intensity back toward its baseline level λ_0, until further events occur. κ controls the rate of memory decay. If $\theta/\kappa < 1$ the self-excitation is held in check by the mean reversion, and the Hawkes process is non-explosive and stationary with unconditional mean:

$$\mathbb{E}(\mu_t) = \frac{\lambda_0}{1 - \theta/\kappa}. \tag{33}$$

Meanwhile, none of the clusters from the Hawkes process are as large as one expects during an intensity burst. In the latter setting, the intensity diverges, whereas it remains bounded, almost surely, in a Hawkes process. However, the point clusters generated by N_t^{normal} may nevertheless be confused by our test statistic as an intensity burst. The purpose is to see if this distorts our test statistic in finite samples.

In our simulations, we set $\lambda_0 = 1/3$, $\theta = 3/10$ and $\kappa = 9/10$, so that on average there is one event every second. This is broadly representative of the daily transaction count observed in our empirical high-frequency data.

The intensity of N_t^{burst} is defined as:

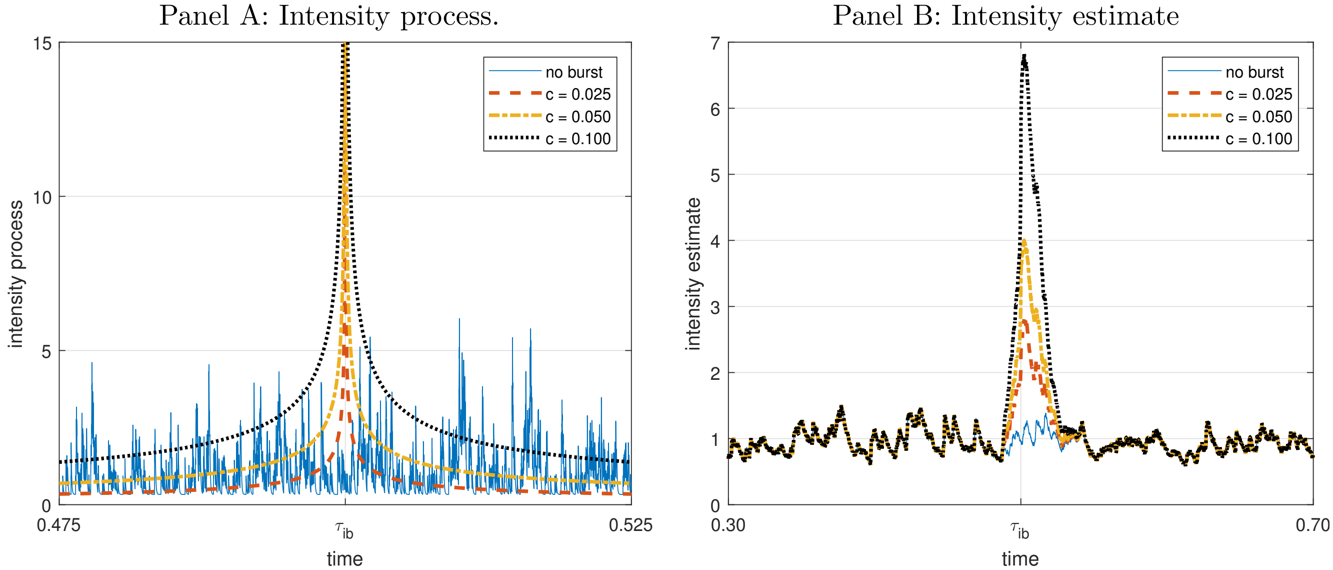
$$\beta_t = \frac{\sigma}{|\tau_{\text{ib}} - t|^{\alpha}}, \quad \text{for } t \in [0.475, 0.525], \tag{34}$$

and $\beta_t = 0$ otherwise. Here, σ and α are constant. We position $\tau_{\text{ib}} = 0.5$ at the center of the interval $[0, 1]$ and make the duration of the intensity burst cover a 10-minute window. To generate intensity bursts of varying magnitude, we take $\alpha \in \{0.25, 0.50, 0.75\}$ and calibrate σ such that N_t^{burst} generates on average a fraction c of the points produced by N_t^{normal}, where $c = 0.000, 0.025, 0.050, 0.100$.[5] We refer to this as no, small, medium, or large intensity burst.

[4]The intensity of the Hawkes process in (32) does not fulfill the definition in (1). It is also not continuous and falls outside Assumption 2. Hence, the simulations are also meant to illustrate the robustness of our testing procedure.

[5]Solving for σ in $c \int_0^1 \mu_t \mathrm{d}t = \int_{\tau_{\text{ib}}-\Delta}^{\tau_{\text{ib}}+\Delta} \beta_t \mathrm{d}t$, where Δ is the duration of the intensity burst and $\beta_t = \frac{\sigma}{|\tau_{\text{ib}} - t|^{\alpha}}$ as in (34), we get $\sigma = c \frac{1+\alpha}{2\Delta^{1+\alpha}} \int_0^1 \mu_t \mathrm{d}t$, such that $c = 0$ corresponds to no burst in the intensity.

Figure 2: Example of a simulated intensity burst.



Note. This figure shows the true (in Panel A) and estimated (in Panel B) intensity of the simulated counting process N_t with no burst, and a small, medium and large intensity burst.

We compute ϕ_t^{ib} at a one-second grid with $\ell_n \in \{60, 300, 600\}$ seconds for the calculation of $\widehat{\lambda}_t$ (corresponding to a 1, 5, and 10-minute block) and $K_n = 5\ell_n$ (corresponding to a 5, 25, and 50-minute block) for the asymptotic variance estimation.

In Panel A of Figure 2, we illustrate the simulated intensity of the underlying Hawkes process of N_t^{normal} (no burst, or $c = 0.000$) together with a small ($c = 0.025$), medium ($c = 0.050$) and large ($c = 0.100$) intensity burst of N_t^{burst} with $\alpha = 0.50$. In Panel B, we show the associated intensity estimate for $\ell_n = 300$. Across the burst states, the maximal local intensity estimate is roughly 3, 4, and 7 times larger than the average rate of the Hawkes process. The intensity burst displayed in Figure 1 is much larger, hence our simulated intensity bursts are conservative relative to those observed in empirical data.

In Table 1, we collect the results of the Monte Carlo analysis. The left-hand side reports the rejection rates for the pointwise test applied at $t = 0.5$, which also corresponds to the explosion point τ_{ib} under the alternative. Here, the test statistic is evaluated against the usual critical values from the standard normal distribution z_p, i.e. the pth percentile. The right-hand side shows the associated rejection rates of the maximum test evaluated against the simulated critical values q_p, as explained in Section 3.3.

Looking at Panel A ($c = 0.000$), the finite sample sizes of both tests are close to the theoretical nominal levels. A mild inflation is observed for the pointwise test and also for the maximum test with the shortest bandwidth. As noted above, this is due to the self-excitation of the Hawkes intensity under $\mathscr{H}_0$. However, the effect diminishes once we smooth the intensity estimate over a slightly longer window. The power of the local test in Panel B – D ($c = 0.025 - 0.100$) is nearly perfect. This is not surprising, since the test statistic is calculated at the peak of the intensity burst under $\mathscr{H}_1$, although detection of small bursts remains challenging. The rejection rate of the maximum test is smaller. However, it suffices to detect the majority of medium and large intensity bursts, whereas small intensity bursts with a slow rate of divergence may go unnoticed. This is not a cause of concern in practice, since we are primarily interested in the behavior of large surges in trading activity.

Table 1: Rejection rate of intensity burst test statistic.

	pointwise test									maximum test								
	$\ell_n = 60$			$\ell_n = 300$			$\ell_n = 600$			$\ell_n = 60$			$\ell_n = 300$			$\ell_n = 600$		
	$z_{0.950}$	$z_{0.975}$	$z_{0.995}$	$z_{0.950}$	$z_{0.975}$	$z_{0.995}$	$z_{0.950}$	$z_{0.975}$	$z_{0.995}$	$q_{0.950}$	$q_{0.975}$	$q_{0.995}$	$q_{0.950}$	$q_{0.975}$	$q_{0.995}$	$q_{0.950}$	$q_{0.975}$	$q_{0.995}$
Panel A: size ($c = 0.000$, *no burst*)																		
	8.8	5.4	2.2	6.2	3.6	1.4	6.6	3.6	0.6	23.6	10.6	8.2	3.2	0.4	0.0	0.4	0.0	0.0
Panel B: power ($c = 0.025$, *small burst*)																		
$\alpha = 0.25$	28.0	19.6	8.4	98.0	94.8	80.8	87.2	80.0	60.8	24.2	10.6	7.4	5.8	1.2	0.6	8.4	1.4	0.6
$\alpha = 0.50$	85.2	79.6	64.4	99.2	97.6	89.4	90.2	84.2	67.6	37.4	18.4	14.8	19.2	6.2	4.0	15.2	4.0	2.4
$\alpha = 0.75$	100.0	100.0	99.6	99.8	99.4	97.8	95.4	92.2	81.2	93.4	88.4	84.8	60.4	45.2	39.4	30.4	17.8	13.4
Panel C: power ($c = 0.050$, *medium burst*)																		
$\alpha = 0.25$	48.6	37.4	18.0	100.0	100.0	99.8	100.0	100.0	99.2	26.6	12.2	8.0	14.4	3.4	1.0	29.8	9.6	5.8
$\alpha = 0.50$	99.8	99.6	97.6	100.0	100.0	100.0	100.0	100.0	99.4	77.6	61.2	52.6	57.0	34.6	26.2	59.0	38.4	30.0
$\alpha = 0.75$	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.2	97.6	96.2	88.6	80.4	75.4
Panel D: power ($c = 0.100$, *large burst*)																		
$\alpha = 0.25$	76.4	63.6	30.6	100.0	100.0	100.0	100.0	100.0	100.0	28.6	11.6	9.2	23.8	7.2	3.0	53.4	28.2	19.6
$\alpha = 0.50$	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.0	96.2	94.8	79.4	52.6	43.8	91.6	80.8	73.6
$\alpha = 0.75$	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.6	99.4

Note. We simulate a point process as the sum of an exponential Hawkes process ("normal") and an intensity burst process ("burst"), $N_t = N_t^{\text{normal}} + N_t^{\text{burst}}$. We test the hypothesis $\mathscr{H}_0 : N_t^n(\omega) \in \Omega^0$. The table reports rejection rates of the intensity burst statistic both for the pointwise test (left side) and the maximum test (right side). The pointwise test is calculated at $t = 0.5$, which is equal to τ_{ib} under the alternative. ℓ_n is the bandwidth (in seconds) for the intensity estimator. The bandwidth for the observed local asymptotic variance is $K_n = 5\ell_n$. α is the explosion rate of the bursting process in (34), whereas c is the proportion of the total transaction count induced by the intensity burst. z_p is the critical value from the standard normal distribution for the pointwise test, while q_p is the associated simulated critical value for the maximum statistic.

5 Empirical application

We examine a continuous record of high-frequency data for futures contracts traded on the Chicago Mercantile Exchange (CME) electronic Globex system, namely CL (crude oil), EC (foreign exchange), ED (short-term interest rates), ES (equities), GC (gold), and TY (long-term interest rates), thus covering all major financial asset classes. The instruments are very liquid and generally trade around the clock five days a week with only a short daily maintenance break. The data were acquired from Tick Data (tickdata.com) for the sample period January 4, 2010 to October 14, 2021; or about 3,000 days in total. Timestamps are in milliseconds. A summary of the data is presented in Table 2. We follow Christensen, Oomen, and Renò (2022) and restrict attention to the most active trading hours from 1:00am to 3:15pm Chicago time, covering the main European and American sessions. As readily seen, this removes only a small fraction of observations.

Table 2: CME futures data.

Contract	Market	Venue	# days	# trades	Retained	Duration
CL	Crude Oil	NYMEX	3,041	493,879,290	94.36%	0.45
EC	Euro FX	CME	3,045	310,413,762	90.02%	0.85
ED	Eurodollar	CME	3,044	25,754,111	94.88%	15.42
ES	E-mini S&P 500	CME	3,045	1,006,304,505	93.43%	0.28
GC	Gold	COMEX	3,041	284,900,979	87.02%	0.85
TY	10-Year T-Note	CME	3,044	268,045,084	92.81%	1.05

Note. The table reports for each futures contract, its ticker symbol (or contract name), the underlying asset, the trading venue, the number of days in the sample, the total number of transactions and the average trade duration (in seconds). The sample period is January 2010–October 2021. We restrict attention to the most active trading hours from 1:00am–3:15pm Chicago time. The fraction of volume retained after removing the most illiquid parts of the day is reported in the penultimate column.

To avoid identifying spurious bursts in liquidity associated with inflated trading volume around pre-scheduled releases of key market information, such as the macroeconomic news announcements, we compute an asset-specific nonparametric estimator of the pointwise time-of-the-day mean intensity. This is calculated as the sample average of the number of transactions in each 15-second bucket over the whole day, where the average is taken over the days in the sample. We compute such a curve separately for FOMC announcement days, since they feature a highly unusual trading pattern compared to the rest of the sample. The associated diurnal curves are reported in Figure B.1 in the appendix. We correct the estimated local intensity in (9) by the diurnal pattern before computing the intensity burst test statistic, which we calculate on a one-second grid throughout the retained part of the trading session.

In Table 3, we report the number of identified events at various critical values of the maximum intensity burst test statistic. The critical values are intentionally set so high that the expected number of false positives is close to zero. Judging by the middle column $\phi^{\mathrm{ib},\star} > 7$, we see that intensity bursts happen frequently, averaging about one event every five–ten days depending on the market. A couple of intensity burst examples are plotted in Figure 3. The events are associated with large increases in the trading intensity over short time intervals leading to a significant intensity burst test statistic. Howeever, they are not necessarily associated with a big change in the transaction price. Hence, only the Treasury futures event is also flagged as a significant drift burst by the test statistic of Christensen, Oomen, and Renò (2022).

Table 3: Intensity burst summary statistics.

contract	market	$\phi^{\text{ib},\star} > 6.0$		$\phi^{\text{ib},\star} > 7.0$		$\phi^{\text{ib},\star} > 8.0$	
		#	%	#	%	#	%
CL	Crude Oil	1407	46.27%	680	22.36%	349	11.48%
EC	Euro FX	1194	39.21%	615	20.20%	342	11.23%
ED	Eurodollar	560	18.40%	335	11.01%	222	7.29%
ES	E-mini S&P 500	786	25.81%	279	9.16%	128	4.20%
GC	Gold	1446	47.55%	799	26.27%	442	14.53%
TY	10-Year T-Note	576	18.92%	258	8.48%	133	4.37%

Note. We report for each futures contract the number of days with intensity bursts identified for critical values ranging between 6 and 8 (percentage of total sample in parenthesis). The number of false positives we expect, which can be computed using the technique described in Section 3.3, is virtually zero.

6 Conclusion

In this paper, we propose a model with unbounded intensity for point processes. We allow the spot intensity to explode such that the integrated intensity is finite, ensuring the point process is non-explosive. The model is capable of generating extreme clusters of observations over small time intervals that are far more concentrated than what a standard doubly stochastic Poisson or self-exciting Hawkes processes with bounded intensity can produce.

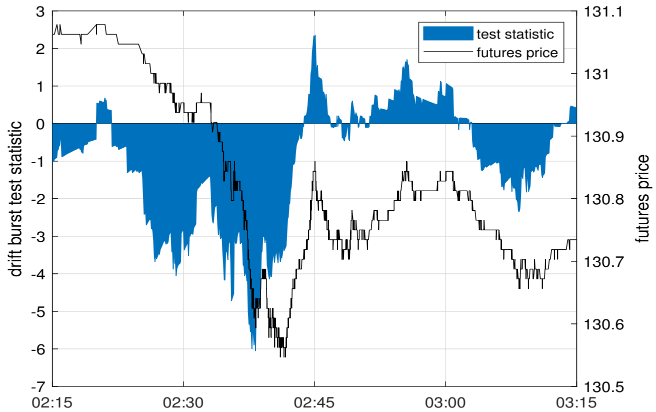
We develop an inference strategy for detecting an intensity burst. In its absence (null hypothesis), the asymptotic distribution of our test statistic is standard normal. The theory relies on a novel heavy traffic assumption, which allows to consistently estimate and draw inference about the spot intensity of a point process over a finite time interval. It resembles the standard in-fill condition for asymptotic theory of realized variance. The rate of convergence depends on the speed with which we accumulate points induced by the heavy traffic condition, so we base the feasible implementation of our test statistic on an automatic inference procedure. In particular, we adapt the observed asymptotic variance of Mykland and Zhang (2017) to spot estimation. Conditional on an intensity burst (alternative hypothesis), our test statistic diverges. Hence, the testing procedure has power converging to one.

A simulation study shows that the test statistic has good finite sample properties, even in a more general setting than the theoretical framework permits. We implement the statistic on high-frequency data from several liquid futures contracts. We detect a nontrivial amount of intensity bursts in the data and describe some of their basic properties.

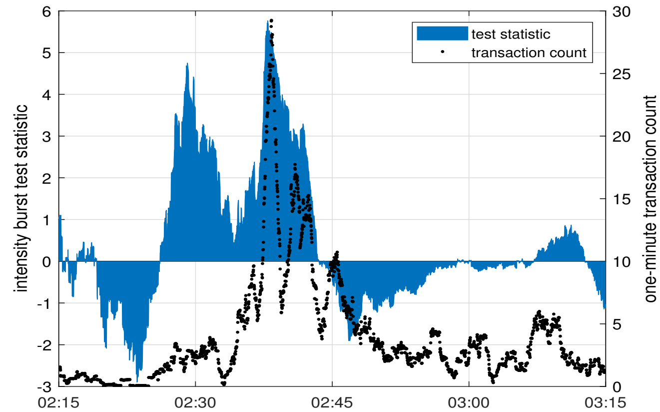
Figure 3: Further examples of intensity bursts in various markets.

TY, 02/05/2020

Panel A: Futures price.

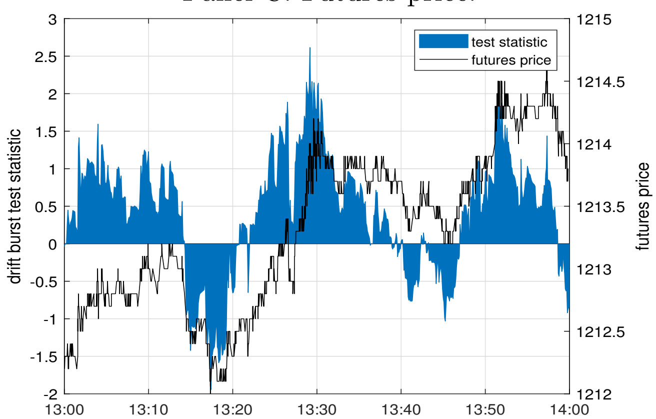


Panel B: Transaction count.

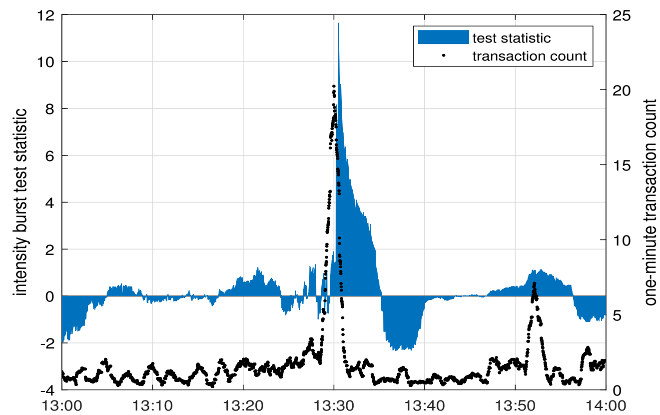


GC, 04/28/2015

Panel C: Futures price.

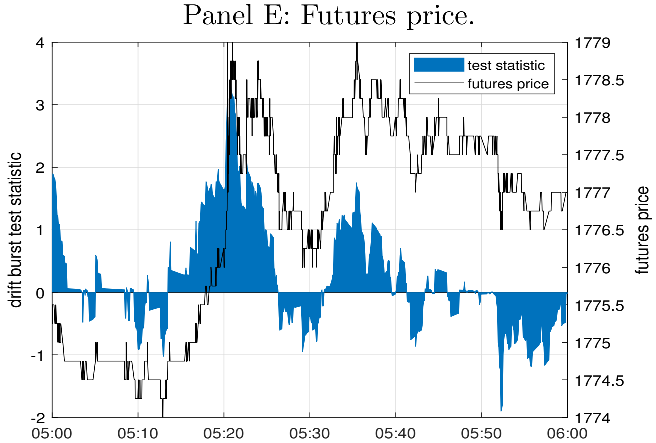


Panel D: Transaction count.

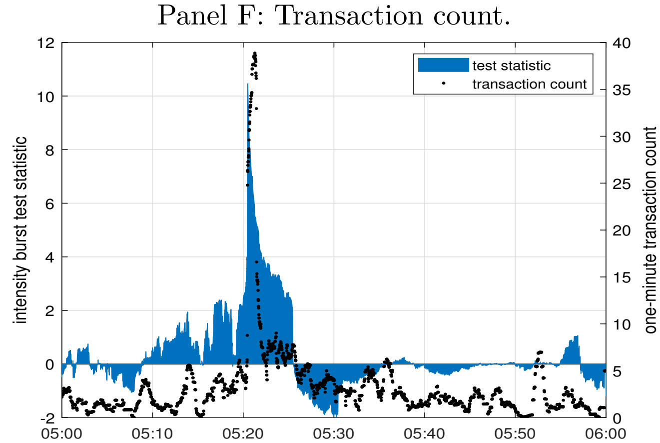


ES, 01/30/2014

Panel E: Futures price.



Panel F: Transaction count.



Note. In the left-hand side of the figure, we plot the futures price along with the drift burst test statistic of Christensen, Oomen, and Renò (2022) for the selected days of the futures contracts (name of contract and MM/DD/YYYY date appears in caption). In the right-hand side, we plot the associated one-minute transaction count along with the intensity burst test statistic.

A Proofs

We here derive the theoretical results listed in the main text. We note that under Assumption 1 – 2, we can appeal to the localization procedure described in Jacod and Protter (2012, Section 4.4.1) to bound various processes.

To establish stable convergence of the local intensity estimator in (9), we need an auxiliary result, which is a reproduction of Alvarez, Panloup, Pontier, and Savy (2012, Lemma 8).

Lemma 2. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a probability space. For each $n \geq 1$, suppose $\zeta_2^n, \zeta_3^n, \dots, \zeta_{k_n}^n$ are martingale increments with respect to the sub-σ-fields of $\mathcal{F}$: $\mathcal{F}_{n,1} \subseteq \mathcal{F}_{n,2} \subseteq \cdots \subseteq \mathcal{F}_{n,k_n}$. Set $S_n = \sum_{i=2}^{k_n} \zeta_i^n$ and $\mathcal{G} = \cap_{n \geq 1} \mathcal{F}_{n,1}$. Assume that $n \mapsto \mathcal{F}_{n,k_n}$ is a non-increasing sequence of σ-fields such that $\cap_{n \geq 1} \mathcal{F}_{n,k_n} = \mathcal{G}$. Then, if the following conditions hold:*

(i) There exists a $\mathcal{G}$-measurable random variable η such that, as $n \to \infty$,

$$\sum_{i=2}^{k_n} \mathbb{E}\Big[(\zeta_i^n)^2 \mid \mathcal{F}_{n,i-1}\Big] \xrightarrow{p} \eta,$$

(ii) For every $\epsilon > 0$,

$$\sum_{i=2}^{k_n} \mathbb{E}\Big[(\zeta_i^n)^2 \mathbf{1}_{\{(\zeta_i^n)^2 \geq \epsilon\}} \mid \mathcal{F}_{n,i-1}\Big] \xrightarrow{p} 0,$$

then $S_n \xrightarrow{\mathfrak{D}_s} S$, defined on an extension of $(\Omega, \mathcal{F}, \mathbb{P})$, such that conditionally on $\mathcal{F}$, the distribution of S is a centered Gaussian with variance η.

A.1 Proof of Lemma 1

Theorem 1 implies that, under $\mathscr{H}_0$,

$$\widehat{\lambda}_t - \lambda_t = o_p(1), \tag{35}$$

which shows the first statement.

Next, under $\mathscr{H}_1$, we notice that the observed process can be decomposed as

$$N_t^n = N_t^{n,\mu} + N_t^{n,\beta},$$

where $N_t^{n,\mu}$ and $N_t^{n,\beta}$ are inhomogeneous Poisson processes with rates μ_t and β_t.

Therefore, at time τ,

$$\widehat{\lambda}_\tau = \frac{N^{n,\mu}(\tau - \delta_n, \tau)}{n\delta_n} + \frac{N^{n,\beta}(\tau - \delta_n, \tau)}{n\delta_n} = \widehat{\mu}_\tau + \widehat{\beta}_\tau.$$

It follows from (35) that

$$\widehat{\mu}_\tau = \mu_\tau + o_p(1),$$

so it is enough to look at $\widehat{\beta}_\tau$. To prove that $\widehat{\beta}_\tau = O_p(\delta_n^{-\alpha})$, by the definition of stochastic orders in probability, it suffices to show that for every $\epsilon > 0$ there exists a $0 < \Delta_\epsilon < \infty$ and $N_\epsilon \in \mathbb{N}$ such that for all $n \geq N_\epsilon : \mathbb{P}\big(\delta_n^\alpha \widehat{\beta}_\tau > \Delta_\epsilon\big) < \epsilon$.

By Markov's inequality:

$$\mathbb{P}\Big(\widehat{\beta}_\tau > \delta_n^{-\alpha}\Delta_\epsilon\Big) \leq \frac{\delta_n^\alpha \mathbb{E}(\widehat{\beta}_\tau)}{\Delta_\epsilon}.$$

Next, conditioning on σ and employing the law of iterated expectations:

$$\mathbb{E}\big(\widehat{\beta}_\tau\big) = \mathbb{E}\left[\frac{N^{n,\beta}(\tau-\delta_n,\tau)}{n\delta_n}\right] = \mathbb{E}\left[\mathbb{E}\left(\frac{N^{n,\beta}(\tau-\delta_n,\tau)}{n\delta_n} \mid \mathcal{F}^\sigma\right)\right] = \mathbb{E}\left[\frac{1}{\delta_n}\int_{\tau-\delta_n}^{\tau} \sigma_u|\tau-u|^{-\alpha}\mathrm{d}u\right]$$
$$\leq \frac{C}{\delta_n}\int_{\tau-\delta_n}^{\tau} |\tau-u|^{-\alpha}\mathrm{d}u = \frac{C}{1-\alpha}\delta_n^{-\alpha},$$

where $\mathcal{F}^\sigma$ is the σ-algebra generated by σ and $C > 0$ is a constant that bounds the process from above in light of the localization procedure. Thus,

$$\mathbb{P}\Big(\widehat{\beta}_\tau > \delta_n^{-\alpha}\Delta_\epsilon\Big) \leq \frac{C}{\Delta_\epsilon(1-\alpha)}.$$

Thus, for every $\epsilon > 0$, we can choose $\Delta_\varepsilon > \left(\frac{C}{\epsilon(1-\alpha)}\right)$ to make $\mathbb{P}\big(\delta_n^\alpha \widehat{\beta}_\tau > \Delta_\epsilon\big) < \epsilon$. ■

A.2 Proof of Theorem 1

We show the univariate statement in (i), (ii), and (iii). The second half follows from the calculations in the proof of Theorem 2.

To this end, we define the sequence ρ_n as follows:

$$\rho_n = \begin{cases} \delta_n n, & \text{if } \delta_n\sqrt{n} \to 0, \\ \delta_n^{-1}, & \text{if } \delta_n\sqrt{n} \to \theta \text{ or } \delta_n\sqrt{n} \to \infty, \end{cases}$$

and study the difference $\sqrt{\rho_n}(\widehat{\lambda}_t - \lambda_t)$.

We assume δ_n can be expressed as $\delta_n = \ell_n \Delta_n$, where $\Delta_n = \frac{1}{n}$ and ℓ_n is a deterministic sequence of positive integers. Then, $\widehat{\lambda}_t$ can be expressed as

$$\widehat{\lambda}_t = \frac{1}{\ell_n}\sum_{i\in\mathcal{D}_{t-}^n} \Delta_i N^n,$$

where $\Delta_i N^n = N^n(0, i\Delta_n) - N^n(0,(i-1)\Delta_n)$ are the increments of the process N^n over the short time intervals of length Δ_n, and

$$\mathcal{D}_{t-}^n = \{tn - \ell_n + 1, tn - \ell_n + 2, \ldots, tn\}.$$

We approximate $\Delta_i N^n$ by the increments of an inhomogeneous Poisson process $\widetilde{N}^n$ with piecewise constant intensity, $\Delta_i \widetilde{N}^n$. This is done as follows. By the random time-change theorem for point processes (see, e.g., Daley and Vere-Jones, 2003, Theorem 7.4.I), there exists a homogeneous unit intensity Poisson process $N^{(\star),(i)}$, such that

$$\Delta_i N^n = N^{(\star),(i)}\left(n \int_{(i-1)\Delta_n}^{i\Delta_n} \mu_s \mathrm{d}s\right).$$

We set $\Delta_i \widetilde{N}^n = N^{(\star),(i)}\left(n \int_{(i-1)\Delta_n}^{i\Delta_n} \mu_{(i-1)\Delta_n} \mathrm{d}s\right)$. Hence, $\widetilde{N}^n(k\Delta_n) = \sum_{i \leq k} \Delta_i \widetilde{N}^n$ is an inhomogeneous Poisson process with piecewise constant intensity, such that

$$\mathbb{E}\Big[\Delta_i \widetilde{N}^n \mid \mathcal{F}_{(i-1)\Delta_n}\Big] = n \int_{(i-1)\Delta_n}^{i\Delta_n} \mu_{(i-1)\Delta_n} \mathrm{d}s = \mu_{(i-1)\Delta_n}.$$

Moreover, the absolute value of the approximation error $|\Delta_i N^n - \Delta_i \widetilde{N}^n|$ can be expressed as an increment of the process $N^{(\star),(i)}$:

$$|\Delta_i N^n - \Delta_i \widetilde{N}^n| = N^{(\star),(i)}(\underline{t}, \overline{t}),$$

where

$$\underline{t} = n\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \mu_s \mathrm{d}s \wedge \int_{(i-1)\Delta_n}^{i\Delta_n} \mu_{(i-1)\Delta_n} \mathrm{d}s\right) \quad \text{and} \quad \overline{t} = n\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \mu_s \mathrm{d}s \vee \int_{(i-1)\Delta_n}^{i\Delta_n} \mu_{(i-1)\Delta_n} \mathrm{d}s\right).$$

Then, we can write

$$\widehat{\lambda}_t - \mu_t = \underbrace{\frac{1}{\ell_n} \sum_{i \in \mathcal{D}_{t-}^n} \left(\Delta_i N^n - \Delta_i \widetilde{N}^n\right)}_{\text{(I)}} + \underbrace{\frac{1}{\ell_n} \sum_{i \in \mathcal{D}_{t-}^n} \left(\Delta_i \widetilde{N}^n - \mu_{(i-1)\Delta_n}\right)}_{\text{(II)}} + \underbrace{\frac{1}{\ell_n} \sum_{i \in \mathcal{D}_{t-}^n} \left(\mu_{(i-1)\Delta_n} - \mu_t\right)}_{\text{(III)}}.$$

(I) arises by approximating the observed process with a point process that has a locally constant intensity, (II) is the deviation of the approximating process from its conditional expectation, and (III) denotes the error due to the variation in the rate process.

(I) can be further decomposed as follows:

$$\begin{aligned}\text{(I)} = {} & \frac{1}{\ell_n} \sum_{i \in \mathcal{D}_{t-}^n} \left(\Delta_i N^n - \Delta_i \widetilde{N}^n - \mathbb{E}\big[\Delta_i N^n - \Delta_i \widetilde{N}^n \mid \mathcal{F}_{(i-1)\Delta_n}\big]\right) \\ & + \frac{1}{\ell_n} \sum_{i \in \mathcal{D}_{t-}^n} \mathbb{E}\big[\Delta_i N^n - \Delta_i \widetilde{N}^n \mid \mathcal{F}_{(i-1)\Delta_n}\big].\end{aligned}$$

Moreover, due to Assumption 2, (III) can be split into a drift and volatility part:

$$\text{(III)} = \frac{1}{\ell_n} \sum_{i \in \mathcal{D}_{t-}^n} (tn - \ell_n - i)(A_{i\Delta_n} - A_{(i-1)\Delta_n}) + \frac{1}{\ell_n} \sum_{i \in \mathcal{D}_{t-}^n} (tn - \ell_n - i)(M_{i\Delta_n} - M_{(i-1)\Delta_n}),$$

where $A_t = \int_0^t a_s \mathrm{d}s$ and $M_t = \int_0^t \nu_s \mathrm{d}W_s$.

Consequently,

$$\sqrt{\rho_n}(\widehat{\lambda}_t - \mu_t) = \Lambda_1^n(t) + \Lambda_2^n(t) + \Lambda_3^n(t) + \Lambda_4^n(t) + \Lambda_5^n(t),$$

where

$$\Lambda_1^n(t) = \frac{\sqrt{\rho_n}}{\ell_n} \sum_{i\in\mathcal{D}_{t-}^n} \big(\Delta_i N^n - \Delta_i \widetilde{N}^n - \mathbb{E}\big[\Delta_i N^n - \Delta_i \widetilde{N}^n \mid \mathcal{F}_{(i-1)\Delta_n}\big]\big),$$

$$\Lambda_2^n(t) = \frac{\sqrt{\rho_n}}{\ell_n} \sum_{i\in\mathcal{D}_{t-}^n} \mathbb{E}\big[\Delta_i N^n - \Delta_i \widetilde{N}^n \mid \mathcal{F}_{(i-1)\Delta_n}\big],$$

$$\Lambda_3^n(t) = \frac{\sqrt{\rho_n}}{\ell_n} \sum_{i\in\mathcal{D}_{t-}^n} \big(\Delta_i \widetilde{N}^n - \mu_{(i-1)\Delta_n}\big),$$

$$\Lambda_4^n(t) = \frac{\sqrt{\rho_n}}{\ell_n} \sum_{i\in\mathcal{D}_{t-}^n} (tn-\ell_n-i)(A_{i\Delta_n} - A_{(i-1)\Delta_n}).$$

$$\Lambda_5^n(t) = \frac{\sqrt{\rho_n}}{\ell_n} \sum_{i\in\mathcal{D}_{t-}^n} (tn-\ell_n-i)(M_{i\Delta_n} - M_{(i-1)\Delta_n}),$$

Now, we show that the sum $\Lambda_3^n(t) + \Lambda_5^n(t)$ converges stably in law, while the other terms ($\Lambda_1^n(t)$, $\Lambda_2^n(t)$, and $\Lambda_4^n(t)$) are asymptotically negligible.

Set

$$\Lambda_3^n(t) + \Lambda_5^n(t) = \sum_{i\in\mathcal{D}_{t-}^n} \zeta_i^n(3) + \zeta_i^n(5),$$

where

$$\zeta_i^n(3) = \frac{\sqrt{\rho_n}}{\ell_n}\big(\Delta_i \widetilde{N}^n - \mu_{(i-1)\Delta_n}\big) \quad \text{and} \quad \zeta_i^n(5) = \frac{\sqrt{\rho_n}}{\ell_n}(tn-\ell_n-i)(M_{i\Delta_n} - M_{(i-1)\Delta_n}).$$

By construction, $\zeta_i^n(3)$ and $\zeta_i^n(5)$ are uncorrelated $\mathcal{F}_{(i-1)\Delta_n}$-martingale differences:

$$\mathbb{E}\big[\zeta_i^n(3) \mid \mathcal{F}_{(i-1)\Delta_n}\big] = \mathbb{E}\big[\zeta_i^n(5) \mid \mathcal{F}_{(i-1)\Delta_n}\big] = \mathbb{E}\big[\zeta_i^n(3)\zeta_i^n(5) \mid \mathcal{F}_{(i-1)\Delta_n}\big] = 0.$$

As a result,

$$\sum_{i\in\mathcal{D}_{t-}^n} \mathbb{E}\Big[\big(\zeta_i^n(3) + \zeta_i^n(5)\big)^2 \mid \mathcal{F}_{(i-1)\Delta_n}\Big] = \sum_{i\in\mathcal{D}_{t-}^n} \mathbb{E}\Big[\big(\zeta_i^n(3)\big)^2 \mid \mathcal{F}_{(i-1)\Delta_n}\Big] + \sum_{i\in\mathcal{D}_{t-}^n} \mathbb{E}\Big[\big(\zeta_i^n(5)\big)^2 \mid \mathcal{F}_{(i-1)\Delta_n}\Big].$$

Since $\Delta_i \widetilde{N}^n$ is Poisson distributed, we deduce that:

$$\mathbb{E}\Big[\big(\zeta_i^n(3)\big)^2 \mid \mathcal{F}_{(i-1)\Delta_n}\Big] = \frac{\rho_n}{\ell_n^2}\mu_{(i-1)\Delta_n},$$

and because μ is continuous, as $\delta_n \to 0$,

$$\frac{1}{\ell_n} \sum_{i\in\mathcal{D}_{t-}^n} \mu_{(i-1)\Delta_n} \xrightarrow{a.s} \mu_t.$$

From the definition of ρ_n:

$$\frac{\rho_n}{\ell_n} \to \begin{cases} 1, & \text{if } \delta_n\sqrt{n} \to 0, \\ \dfrac{1}{\theta^2}, & \text{if } \delta_n\sqrt{n} \to \theta, \\ 0, & \text{if } \delta_n\sqrt{n} \to \infty. \end{cases}$$

Therefore,

$$\sum_{i\in\mathcal{D}_{t-}^n} \mathbb{E}\Big[\big(\zeta_i^n(3)\big)^2 \mid \mathcal{F}_{(i-1)\Delta_n}\Big] \xrightarrow{p} \begin{cases} \mu_t, & \text{if } \delta_n\sqrt{n} \to 0, \\ \dfrac{1}{\theta^2}\mu_t, & \text{if } \delta_n\sqrt{n} \to \theta, \\ 0, & \text{if } \delta_n\sqrt{n} \to \infty. \end{cases}$$

Next,

$$\mathbb{E}\Big[\big(\zeta_i^n(5)\big)^2 \mid \mathcal{F}_{(i-1)\Delta_n}\Big] = \frac{\rho_n}{\ell_n^2}(tn-\ell_n-i)^2 \int_{(i-1)\Delta_n}^{i\Delta_n} \mathbb{E}\big[\nu_s^2 \mid \mathcal{F}_{(i-1)\Delta_n}\big]\mathrm{d}s.$$

Assumption 2 implies that, as $n\to\infty$,

$$\mathbb{E}\Bigg[\sup_{s\in[t-\delta_n,t]} |\nu_s^2-\nu_t^2|\Bigg] \to 0,$$

which further means that

$$\sum_{i\in\mathcal{D}_{t-}^n}\Bigg(\mathbb{E}\Big[\big(\zeta_i^n(5)\big)^2 \mid \mathcal{F}_{(i-1)\Delta_n}\Big] - \frac{\rho_n}{\ell_n^2}(tn-\ell_n-i)^2\Delta_n\nu_t^2\Bigg) \xrightarrow{p} 0.$$

On the other hand, multiplying and dividing by δ_n,

$$\sum_{i\in\mathcal{D}_{t-}^n} \frac{\rho_n}{\ell_n^2}(tn-\ell_n-i)^2\Delta_n\nu_t^2 = \rho_n\delta_n\nu_t^2\frac{1}{\ell_n^3}\sum_{i\in\mathcal{D}_{t-}^n}(tn-\ell_n-i)^2.$$

The last part is convergent with limit

$$\frac{1}{\ell_n^3}\sum_{i\in\mathcal{D}_{t-}^n}(tn-\ell_n-i)^2 = \frac{1}{\ell_n^3}\sum_{j=1}^{\ell_n} j^2 = \frac{1}{\ell_n^3}\left[\frac{\ell_n(\ell_n+1)(2\ell_n+1)}{6}\right] \to \frac{1}{3}.$$

Moreover,

$$\rho_n\delta_n = \begin{cases} \delta_n^2 n \to 0, & \text{if } \delta_n\sqrt{n}\to 0, \\ 1, & \text{if } \delta_n\sqrt{n}\to\theta \text{ or } \delta_n\sqrt{n}\to\infty. \end{cases}$$

Putting it together,

$$\sum_{i\in\mathcal{D}_{t-}^n} \mathbb{E}\Big[\big(\zeta_i^n(5)\big)^2 \mid \mathcal{F}_{(i-1)\Delta_n}\Big] \xrightarrow{p} \begin{cases} 0, & \text{if } \delta_n\sqrt{n}\to 0, \\ \dfrac{1}{3}\nu_t^2 & \text{if } \delta_n\sqrt{n}\to\theta \text{ or } \delta_n\sqrt{n}\to\infty, \end{cases}$$

which implies that

$$\sum_{i\in\mathcal{D}_{t-}^n}\mathbb{E}\Big[\big(\zeta_i^n(3)\big)^2+\big(\zeta_i^n(5)\big)^2\mid\mathcal{F}_{(i-1)\Delta_n}\Big]\overset{p}{\longrightarrow}\begin{cases}\mu_t, & \text{if } \delta_n\sqrt{n}\to 0,\\ \dfrac{1}{\theta^2}\mu_t+\dfrac{1}{3}\nu_t^2 & \text{if } \delta_n\sqrt{n}\to\theta,\\ \dfrac{1}{3}\nu_t^2 & \text{if } \delta_n\sqrt{n}\to\infty.\end{cases}$$

To establish the asymptotic distribution, we prove a Lindeberg condition of the form:

$$\sum_{i\in\mathcal{D}_{t-}^n}\mathbb{E}\Big[\big(\zeta_i^n(3)+\zeta_i^n(5)\big)^2\mathbf{1}_{\left\{(\zeta_i n(3)+\zeta_i^n(5))^2\geq\epsilon\right\}}\mid\mathcal{F}_{(i-1)\Delta_n}\Big]\overset{a.s.}{\longrightarrow}0,\quad\forall\epsilon>0.$$

By the Cauchy-Schwarz and Chebyshev inequalities,

$$\begin{aligned}&\mathbb{E}\Big[\big(\zeta_i^n(3)+\zeta_i^n(5)\big)^2\mathbf{1}_{\left\{(\zeta_i^n(3)+\zeta_i^n(5))^2\geq\epsilon\right\}}\mid\mathcal{F}_{(i-1)\Delta_n}\Big]\\ &\qquad\leq\sqrt{\mathbb{E}\Big[\big(\zeta_i^n(3)+\zeta_i^n(5)\big)^4\mid\mathcal{F}_{(i-1)\Delta_n}\Big]\mathbb{E}\Big[\mathbf{1}_{\left\{(\zeta_i^n(3)+\zeta_i^n(5))^4\geq\epsilon\right\}}\mid\mathcal{F}_{(i-1)\Delta_n}\Big]}\\ &\qquad=\sqrt{\mathbb{E}\Big[\big(\zeta_i^n(3)+\zeta_i^n(5)\big)^4\mid\mathcal{F}_{(i-1)\Delta_n}\Big]\mathbb{P}\Big(\big(\zeta_i^n(3)+\zeta_i^n(5)\big)^2\geq\epsilon\mid\mathcal{F}_{(i-1)\Delta_n}\Big)}\\ &\qquad\leq\frac{1}{\epsilon}\left(\mathbb{E}\Big[\big(\zeta_i^n(3)\big)^4\mid\mathcal{F}_{(i-1)\Delta_n}\Big]+\mathbb{E}\Big[\big(\zeta_i^n(5)\big)^4\mid\mathcal{F}_{(i-1)\Delta_n}\Big]\right).\end{aligned}$$

Since $\Delta_i\widetilde{N}^n$ follows a Poisson distribution with bounded intensity:

$$\begin{aligned}\sum_{i\in\mathcal{D}_{t-}^n}\mathbb{E}\Big[\big(\zeta_i^n(3)\big)^4\mid\mathcal{F}_{(i-1)\Delta_n}\Big]&=\sum_{i\in\mathcal{D}_{t-}^n}\frac{\rho_n^2}{\ell_n^4}\mathbb{E}\Big[\Delta_i\widetilde{N}^n-\mu_{(i-1)\Delta_n}^4\mid\mathcal{F}_{(i-1)\Delta_n}\Big]\\ &=\sum_{i\in\mathcal{D}_{t-}^n}\frac{\rho_n^2}{\ell_n^4}\mu_{(i-1)\Delta_n}(1+3\mu_{(i-1)\Delta_n})\\ &\leq C\frac{\rho_n^2}{\ell_n^3}\longrightarrow 0,\end{aligned}$$

for both choices of ρ_n.

On the other hand, using the boundedness of ν,

$$\begin{aligned}\sum_{i\in\mathcal{D}_{t-}^n}\mathbb{E}\Big[\big(\zeta_i^n(4)\big)^4\mid\mathcal{F}_{(i-1)\Delta_n}\Big]&=\sum_{i\in\mathcal{D}_{t-}^n}\frac{\rho_n^2}{\ell_n^4}(tn-\ell_n-i)^4\;\mathbb{E}\Big[\big(M_{i\Delta_n}-M_{(i-1)\Delta_n}\big)^4\mid\mathcal{F}_{(i-1)\Delta_n}\Big]\\ &=3\frac{\rho_n^2}{\ell_n^4}\sum_{i\in\mathcal{D}_{t-}^n}(tn-\ell_n-i)^4\;\mathbb{E}\left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n}\nu_s^2\mathrm{d}s\right)^2\mid\mathcal{F}_{(i-1)\Delta_n}\right]\\ &\leq C\frac{\rho_n^2\Delta_n^2}{\ell_n^4}\sum_{i\in\mathcal{D}_{t-}^n}(tn-\ell_n-i)^4=C\frac{\rho_n^2\Delta_n^2}{\ell_n^4}\frac{6\ell_n^5+15\ell_n^4+10\ell_n^3-\ell_n}{30}\to 0,\end{aligned}$$

again for both choices of ρ_n. Hence, the Lindeberg condition holds.

By Lemma 2, we therefore conclude that

$$\Lambda_3^n(t) + \Lambda_5^n(t) \xrightarrow{\mathcal{D}_s} \begin{cases} \sqrt{\mu_t} Z, & \text{if } \delta_n \sqrt{n} \to 0, \\ \sqrt{\dfrac{1}{\theta^2}\mu_t + \dfrac{1}{3}\nu_t^2} Z, & \text{if } \delta_n \sqrt{n} \to \theta, \\ \sqrt{\dfrac{1}{3}\nu_t^2} Z, & \text{if } \delta_n \sqrt{n} \to \infty, \end{cases}$$

where $Z \sim N(0,1)$ independent of $\mathcal{F}$.

To end the proof, we next demonstrate asymptotic negligibility of the remaining terms. We start with $\Lambda_1^n(t)$, which we express as follows:

$$\Lambda_1^n(t) = \sum_{i \in \mathcal{D}_{t-}^n} \zeta_i^n(1),$$

where

$$\zeta_i^n(1) = \frac{\sqrt{\rho_n}}{\ell_n} \Big(\Delta_i N^n - \Delta_i \widetilde{N}^n - \mathbb{E}\Big[\Delta_i N^n - \Delta_i \widetilde{N}^n \mid \mathcal{F}_{(i-1)\Delta_n} \Big] \Big)$$

is an $\mathcal{F}_{(i-1)\Delta_n}$-martingale difference sequence by design. Hence,

$$\begin{aligned} \mathbb{E}\big[|\Lambda_1^n(t)|^2 \big] &= \frac{\rho_n}{\ell_n^2} \sum_{i \in \mathcal{D}_{t-}^n} \mathbb{E}\Big[\Big(\Delta_i N^n - \Delta_i \widetilde{N}^n - \mathbb{E}\big[\Delta_i N^n - \Delta_i \widetilde{N}^n \mid \mathcal{F}_{(i-1)\Delta_n} \big] \Big)^2 \Big] \\ &= \frac{\rho_n}{\ell_n^2} \sum_{i \in \mathcal{D}_{t-}^n} \mathbb{E}\Big[n \int_{(i-1)\Delta_n}^{i\Delta_n} \mu_s - \mu_{(i-1)\Delta_n} \mathrm{d}s \Big] \\ &\leq \frac{\rho_n n}{\ell_n^2} \sum_{i \in \mathcal{D}_{t-}^n} \int_{(i-1)\Delta_n}^{i\Delta_n} \mathbb{E}\big[|\mu_s - \mu_{(i-1)\Delta_n}| \big] \mathrm{d}s, \end{aligned}$$

where the tower property of conditional expectation was used. Now, Assumption 2 and standard estimates for semimartingales imply the existence of a constant $C > 0$, such that

$$\mathbb{E}\big[|\mu_s - \mu_{(i-1)\Delta_n}| \big] \leq C \sqrt{\Delta_n}. \tag{36}$$

Thus, for any definition of ρ_n,

$$\mathbb{E}\big[|\Lambda_1^n(t)|^2 \big] \leq \frac{\rho_n C \sqrt{\Delta_n}}{\ell_n} \longrightarrow 0,$$

which implies the asymptotic negligibility of $\Lambda_1^n(t)$.

Next,

$$\Lambda_2^n(t) = \sum_{i \in \mathcal{D}_{t-}^n} \zeta_i^n(2),$$

where

$$\zeta_i^n(2) = \frac{\sqrt{\rho_n}}{\ell_n} \mathbb{E}\Big[\Delta_i N^n - \Delta_i \widetilde{N}^n \mid \mathcal{F}_{(i-1)\Delta_n} \Big].$$

Employing the estimate in (36),

$$\mathbb{E}\big[|\Lambda_2^n(t)|\big] \le \frac{\sqrt{\rho_n}}{\ell_n} \sum_{i\in\mathcal{D}_{t-}^n} \mathbb{E}\Big[|n\int_{(i-1)\Delta_n}^{i\Delta_n} \mu_s - \mu_{(i-1)\Delta_n}\mathrm{d}s| \mid \mathcal{F}_{(i-1)\Delta_n}\Big] \le C\sqrt{\rho_n\Delta_n} \longrightarrow 0,$$

so $\Lambda_2^n(t)$ also vanishes.

$\Lambda_4^n(t)$ is asymptotically negligible following the proof from Alvarez, Panloup, Pontier, and Savy (2012) for the local volatility estimator. We omit the details for brevity. ■

A.3 Proof of Theorem 2

The observed asymptotic local variance can be expressed as

$$\widetilde{\mathsf{avar}}(\nabla\widehat{\lambda}_t) = \frac{\rho_n}{K_n}\sum_{j=0}^{K_n-1}\big(\widehat{\lambda}_{t_j} - \widehat{\lambda}_{t_j-\ell_n\Delta_n}\big)^2.$$

where $t_j = t - 2j\ell_n\Delta_n$.

We denote by $\mathcal{D}_{t_j}^n$ the union of $\mathcal{D}_{t_j-}^n$ and $\mathcal{D}_{(t_j-\ell_n\Delta_n)-}^n$. Then, for every t_j, as in the proof of Theorem 1 the local intensity estimator can be decomposed as:

$$\sqrt{\rho_n}\big(\widehat{\lambda}_{t_j} - \widehat{\lambda}_{t_j-\ell_n\Delta_n}\big) = \Lambda_1^n(t_j) + \Lambda_2^n(t_j) + \Lambda_3^n(t_j) + \Lambda_4^n(t_j) + \Lambda_5^n(t_j),$$

where

$$\Lambda_1^n(t_j) = \frac{\sqrt{\rho_n}}{\ell_n}\sum_{i\in\mathcal{D}_{t_j}^n}(-1)^{\mathbb{1}\{i<t_j-\ell_n\Delta_n\}}\big(\Delta_i N^n - \Delta_i\widetilde{N}^n - \mathbb{E}\big[\Delta_i N^n - \Delta_i\widetilde{N}^n \mid \mathcal{F}_{(i-1)\Delta_n}\big]\big),$$

$$\Lambda_2^n(t_j) = \frac{\sqrt{\rho_n}}{\ell_n}\sum_{i\in\mathcal{D}_{t_j}^n}(-1)^{\mathbb{1}\{i<t_j-\ell_n\Delta_n\}}\mathbb{E}\big[\Delta_i N^n - \Delta_i\widetilde{N}^n \mid \mathcal{F}_{(i-1)\Delta_n}\big],$$

$$\Lambda_3^n(t_j) = \frac{\sqrt{\rho_n}}{\ell_n}\sum_{i\in\mathcal{D}_{t_j}^n}(-1)^{\mathbb{1}\{i<t_j-\ell_n\Delta_n\}}\big(\Delta_i\widetilde{N}^n - \mu_{(i-1)\Delta_n}\big),$$

$$\Lambda_4^n(t_j) = \frac{\sqrt{\rho_n}}{\ell_n}\sum_{i\in\mathcal{D}_{t_j}^n}(-1)^{\mathbb{1}\{i<t_j-\ell_n\Delta_n\}}(t_j n - 2\ell_n - i)(A_{i\Delta_n} - A_{(i-1)\Delta_n}).$$

$$\Lambda_5^n(t_j) = \frac{\sqrt{\rho_n}}{\ell_n}\sum_{i\in\mathcal{D}_{t_j}^n}(-1)^{\mathbb{1}\{i<t_j-\ell_n\Delta_n\}}(t_j n - 2\ell_n - i)(M_{i\Delta_n} - M_{(i-1)\Delta_n}),$$

with N_t, M_t, and A_t defined as in Theorem 1.

Then,

$$\widetilde{\mathsf{avar}}(\nabla\widehat{\lambda}_t) = \mathcal{E}_1^n(t) + \mathcal{E}_2^n(t) + \mathcal{R}^n(t),$$

where

$$\mathcal{E}_1^n(t) = \frac{1}{K_n}\sum_{j=0}^{K_n-1}\big(\Lambda_3^n(t_j)\big)^2,$$

$$\mathcal{E}_2^n(t) = \frac{1}{K_n} \sum_{j=0}^{K_n-1} \big(\Lambda_5^n(t_j)\big)^2,$$

$$\mathcal{R}^n(t) = \widetilde{\mathsf{avar}}(\nabla \widehat{\lambda}_t) - \mathcal{E}_1^n(t) - \mathcal{E}_2^n(t).$$

What is left amounts to showing that $\mathcal{E}_1^n(t)$ and $\mathcal{E}_2^n(t)$ converge to the first and the second term in the true asymptotic variance, whereas $\mathcal{R}^n(t) \xrightarrow{p} 0$.

For the first term, we observe that

$$\big(\Lambda_3^n(t_j)\big)^2 = \sum_{i \in \mathcal{D}_{t_j}^n} \big(\zeta_i^n(3)\big)^2 + 2 \sum_{s,i \in \mathcal{D}_{t_j}^n : s>i} \zeta_i^n(3)\zeta_s^n(3)(-1)^{\mathbb{1}_{\{i<t_j-\ell_n\Delta_n\}}+\mathbb{1}_{\{s<t_j-\ell_n\Delta_n\}}},$$

where $\zeta_i^n(3) = \frac{\sqrt{\rho_n}}{\ell_n}\big(\Delta_i \widetilde{N}^n - \mu_{(i-1)\Delta_n}\big)$ as above. Since $\zeta_i^n(3)$ is a $\mathcal{F}_{i\Delta_n}$-martingale and

$$\mathbb{E}\Big[\big(\zeta_i^n(3)\big)^2 \mid \mathcal{F}_{(i-1)\Delta_n}\Big] = \frac{\rho_n}{\ell_n^2}\mu_{(i-1)\Delta_n},$$

it follows that

$$\mathbb{E}\Big[\big(\Lambda_3^n(t_j)\big)^2 - \frac{\rho_n}{\ell_n^2} \sum_{i \in \mathcal{D}_{t_j}^n} \mu_{(i-1)\Delta_n}\Big] = 0.$$

Then, we decompose $\mathcal{E}_1^n(t)$ into

$$\mathcal{E}_1^n(t) = \mathcal{E}_1'^n(t) + \mathcal{E}_1''^n(t),$$

where

$$\mathcal{E}_1'^n(t) = \frac{1}{K_n} \sum_{j=0}^{K_n-1} \Big(\big(\Lambda_3^n(t_j)\big)^2 - \frac{\rho_n}{\ell_n^2} \sum_{i \in \mathcal{D}_{t_j}^n} \mu_{(i-1)\Delta_n}\Big),$$

$$\mathcal{E}_1''^n(t) = \frac{1}{K_n} \frac{\rho_n}{\ell_n^2} \sum_{j=0}^{K_n-1} \sum_{i \in \mathcal{D}_{t_j}^n} \mu_{(i-1)\Delta_n}.$$

So $\mathbb{E}\big[\mathcal{E}_1'^n(t)\big] = 0$, and because the $\Lambda_3^n(t_j)$'s are based on non-overlapping blocks of observations, the variance of $\mathcal{E}_1'^n(t)$ has the following form:

$$\mathbb{E}\Big[\big(\mathcal{E}_1'^n(t)\big)^2\Big] = \frac{1}{K_n^2} \sum_{j=0}^{K_n-1} \mathbb{E}\Big[\Big(\big(\Lambda_3^n(t_j)\big)^2 - \frac{\rho_n}{\ell_n^2} \sum_{i \in \mathcal{D}_{t_j}^n} \mu_{(i-1)\Delta_n}\Big)^2\Big] \leq \frac{C}{K_n^2} \sum_{j=0}^{K_n-1} \mathbb{E}\Big[\big(\Lambda_3^n(t_j)\big)^4\Big].$$

Next,

$$\mathbb{E}\Big[\big(\Lambda_3^n(t_j)\big)^4\Big] = \sum_{i \in \mathcal{D}_{t_j}^n} \mathbb{E}\Big[\big(\zeta_i^n(3)\big)^4\Big] + 4 \sum_{s,i \in \mathcal{D}_{t_j}^n : s>i} \mathbb{E}\Big[\big(\zeta_i^n(3)\zeta_s^n(3)\big)^2\Big].$$

Alluding to the boundedness of μ, we get the estimates:

$$\mathbb{E}\Big[\big(\zeta_i^n(3)\big)^2\Big] \leq C\frac{\rho_n}{\ell_n^2} \quad \text{and} \quad \mathbb{E}\Big[\big(\zeta_i^n(3)\big)^4\Big] \leq C\frac{\rho_n^2}{\ell_n^4}.$$

Based on this, we deduce that:

$$\mathbb{E}\Big[\big(\Lambda_3^n(t_j)\big)^4\Big] \le \frac{\rho_n^2}{\ell_n^4}\left(\sum_{i=1}^{2\ell_n} C + 4\sum_{i=1}^{2\ell_n-1}\sum_{s=i+1}^{2\ell_n} C\right) = C\frac{\rho_n^2(2\ell_n + 4\ell_n(2\ell_n-1))}{\ell_n^4}. \tag{37}$$

Ergo,

$$\mathbb{E}\Big[\big(\mathcal{E}_1'^n(t)\big)^2\Big] \le \frac{C}{K_n^2}\sum_{j=0}^{K_n-1}\frac{\rho_n^2(2\ell_n+4\ell_n(2\ell_n-1))}{\ell_n^4} = O\Big(K_n^{-1}\ell_n^{-2}\rho_n^2\Big),$$

so that $\mathcal{E}_1'^n(t) \xrightarrow{p} 0$.

The second term, $\mathcal{E}_1''^n(t)$, can be represented as:

$$\mathcal{E}_1''^n(t) = \frac{\rho_n}{\ell_n^2 K_n}\sum_{s=1}^{2K_n\ell_n}\mu_{t-s\Delta_n} = 2\frac{\rho_n}{\ell_n}\frac{1}{2K_n\ell_n}\sum_{s=1}^{2K_n\ell_n}\mu_{t-s\Delta_n}.$$

Since $K_n\ell_n\Delta_n \to 0$, following the train of thought in the proof of Theorem 1 implies that

$$\mathcal{E}_1^n(t) \xrightarrow{p} \begin{cases} 2\mu_t, & \text{if } \ell_n\sqrt{\Delta_n} \to 0, \\ \dfrac{2}{\theta^2}\mu_t, & \text{if } \ell_n\sqrt{\Delta_n} \to \theta, \\ 0, & \text{if } \ell_n\sqrt{\Delta_n} \to \infty. \end{cases}$$

Now, moving on to $\mathcal{E}_2^n(t)$, we again observe that

$$\big(\Lambda_5^n(t_j)\big)^2 = \sum_{i\in\mathcal{D}_{t_j}^n}\big(\zeta_i^n(5)\big)^2 + 2\sum_{s,i\in\mathcal{D}_{t_j}^n: s>i}\zeta_i^n(5)\zeta_s^n(5)(-1)^{\mathbb{1}\{i<t_j-\ell_n\Delta_n\}+\mathbb{1}\{s<t_j-\ell_n\Delta_n\}},$$

where $\zeta_i^n(5) = \frac{\sqrt{\rho_n}}{\ell_n}(t_j n - 2\ell_n - i)(M_{i\Delta_n} - M_{(i-1)\Delta_n})$. Since $\zeta_i^n(5)$ is $\mathcal{F}_{i\Delta_n}$-martingale and

$$\mathbb{E}\Big[\big(\zeta_i^n(5)\big)^2 \mid \mathcal{F}_{(i-1)\Delta_n}\Big] = \frac{\rho_n}{\ell_n^2}(t_j n - 2\ell_n - i)^2\int_{(i-1)\Delta_n}^{i\Delta_n}\mathbb{E}\big[\nu_s^2 \mid \mathcal{F}_{(i-1)\Delta_n}\big]\mathrm{d}s,$$

we have:

$$\mathbb{E}\left[\big(\Lambda_5^n(t_j)\big)^2 - \sum_{i\in\mathcal{D}_{t_j}^n}\frac{\rho_n}{\ell_n^2}(t_j n - 2\ell_n - i)^2\int_{(i-1)\Delta_n}^{i\Delta_n}\mathbb{E}\big[\nu_s^2 \mid \mathcal{F}_{(i-1)\Delta_n}\big]\mathrm{d}s\right] = 0.$$

Then, we decompose $\mathcal{E}_2^n(t)$ as the sum of four terms:

$$\mathcal{E}_2^n(t) = \mathcal{E}_2'^n(t) + \mathcal{E}_2''^n(t) + \mathcal{E}_2'''^n(t) + \mathcal{E}_2''''^n(t),$$

where

$$\mathcal{E}_2'^n(t) = \frac{1}{K_n}\sum_{j=0}^{K_n-1}\left(\big(\Lambda_5^n(t_j)\big)^2 - \frac{\rho_n}{\ell_n^2}\sum_{i\in\mathcal{D}_{t_j}^n}(t_j n - 2\ell_n - i)^2\int_{(i-1)\Delta_n}^{i\Delta_n}\mathbb{E}\big[\nu_s^2 \mid \mathcal{F}_{(i-1)\Delta_n}\big]\mathrm{d}s\right),$$

$$\mathcal{E}_2''^n(t) = \frac{\rho_n}{\ell_n^2 K_n}\sum_{j=0}^{K_n-1}\sum_{i\in\mathcal{D}_{t_j}^n}(t_j n - 2\ell_n - i)^2\int_{(i-1)\Delta_n}^{i\Delta_n}\mathbb{E}\big[\nu_s^2 - \nu_{(i-1)\Delta_n}^2 \mid \mathcal{F}_{(i-1)\Delta_n}\big]\mathrm{d}s,$$

$$\mathcal{E}_2'''^n(t) = \frac{\rho_n}{\ell_n^2 K_n} \sum_{j=0}^{K_n-1} \sum_{i \in \mathcal{D}_{t_j}^n} (t_j n - 2\ell_n - i)^2 \big(\nu_{(i-1)\Delta_n}^2 - \nu_t^2\big) \Delta_n,$$

$$\mathcal{E}_2''''^n(t) = \frac{\rho_n}{\ell_n^2 K_n} \sum_{j=0}^{K_n-1} \sum_{i \in \mathcal{D}_{t_j}^n} (t_j n - 2\ell_n - i)^2 \nu_t^2 \Delta_n.$$

By construction, $\mathbb{E}[\mathcal{E}_2'^n(t)] = 0$ and, as in the proof for $\mathcal{E}_1'^n(t)$,

$$\mathbb{E}\Big[\big(\mathcal{E}_2'^n(t)\big)^2\Big] \leq \frac{C}{K_n} \sum_{j=0}^{K_n-1} \mathbb{E}\Big[\big(\Lambda_5^n(t_j)\big)^4\Big],$$

where, since $\zeta_i^n(5)$ in the definition of $\Lambda_5^n(t_j)$ is an $\mathcal{F}_{i\Delta_n}$-martingale,

$$\mathbb{E}\Big[\big(\Lambda_5^n(t_j)\big)^4\Big] = \sum_{i \in \mathcal{D}_{t_j}^n} \mathbb{E}\Big[\big(\zeta_i^n(5)\big)^4\Big] + 4 \sum_{s,i \in \mathcal{D}_{t_j}^n : s > i} \mathbb{E}\Big[\big(\zeta_i^n(5)\zeta_s^n(5)\big)^2\Big].$$

From the proof of Theorem 1:

$$\mathbb{E}\Big[\big(\zeta_i^n(5)\big)^4\Big] \leq \frac{C\rho_n^2\Delta_n^2}{\ell_n^4} \sum_{i \in \mathcal{D}_{t_j}^n} (t_j n - 2\ell_n - i)^4 \quad \text{and} \quad \mathbb{E}\Big[\big(\zeta_i^n(5)\big)^2\Big] \leq \frac{C\rho_n\Delta_n}{\ell_n^2} \sum_{i \in \mathcal{D}_{t_j}^n} (t_j n - 2\ell_n - i)^2.$$

Standard formulas for calculating the sum of powers of integers yield

$$\sum_{i \in \mathcal{D}_{t_j}^n} (t_j n - 2\ell_n - i)^4 = \frac{96\ell_n^5 + 120\ell_n^4 + 40\ell_n^3 - \ell_n}{15},$$

and

$$\sum_{i \in \mathcal{D}_{t_j}^n} (t_j n - 2\ell_n - i)^2 = \frac{8\ell_n^3 + 6\ell_n^2 + \ell_n}{3}. \tag{38}$$

Hence,

$$\mathbb{E}\Big[\big(\zeta_i^n(5)\big)^4\Big] = O(\rho_n^2 \Delta_n^2 \ell_n),$$

and it follows that $\mathcal{E}_2'^n(t) = o_p(1)$.

For the second term,

$$\mathbb{E}\big[|\mathcal{E}_2''^n(t)|\big] \leq \overline{M} \frac{\rho_n \Delta_n}{\ell_n^2 K_n} \sum_{j=0}^{K_n-1} \sum_{i \in \mathcal{D}_{t_j}^n} (t_j n - 2\ell_n - i)^2,$$

where $\overline{M} = \mathbb{E}\Big[\sup_{s \in [t - K_n \ell_n \Delta_n, t]} |\nu_s^2 - \nu_t^2|\Big]$. Since ν is càdlàg and $K_n \ell_n \Delta_n \to 0$, $\overline{M} = o_p(1)$. The sum was computed in (38), so combining the terms shows that

$$\frac{\rho_n \Delta_n}{\ell_n^2 K_n} \sum_{j=0}^{K_n-1} \sum_{i \in \mathcal{D}_{t_j}^n} (t_j n - 2\ell_n - i)^2 = \frac{8}{3} \rho_n \ell_n \Delta_n + o(1),$$

where

$$\frac{8}{3}\rho_n \ell_n \Delta_n \to \begin{cases} 0, & \text{if } \ell_n \sqrt{\Delta_n} \to 0, \\ \frac{8}{3}, & \text{if } \ell_n \sqrt{\Delta_n} \to \theta > 0 \text{ or } \ell_n \sqrt{\Delta_n} \to \infty. \end{cases}$$

As such, $\mathcal{E}_2^{\prime\prime n}(t) = o_p(1)$.

Following the above steps, we also deduce that

$$\mathcal{E}_2^{\prime\prime\prime n}(t) \xrightarrow{p} 0.$$

The last term can be written as

$$\mathcal{E}_2^{\prime\prime\prime\prime n}(t) = \nu_t^2 \left(\frac{\rho_n \Delta_n}{\ell_n^2 K_n} \sum_{j=0}^{K_n - 1} \sum_{i \in \mathcal{D}_{t_j}^n} (t_j n - 2\ell_n - i)^2 \right),$$

where the sum on the right-hand side is given by (38). As a consequence,

$$\mathcal{E}_2^{\prime\prime\prime\prime n}(t) \xrightarrow{p} \begin{cases} 0, & \text{if } \ell_n \sqrt{\Delta_n} \to 0, \\ \frac{8}{3}\nu_t^2, & \text{if } \ell_n \sqrt{\Delta_n} \to \theta > 0 \text{ or } \ell_n \sqrt{\Delta_n} \to \infty, \end{cases}$$

which handles the analysis of $\mathcal{E}_2^n(t)$.

To finalize the proof up, we notice that $\mathcal{R}^n(t)$ is an average of terms that all converge in probability to zero, and therefore it also converges in probability to zero. ■

A.4 Proof of Theorem 3

We largely copy from the proof of Theorem 2. Therefore, many repeated details are omitted and we concentrate on explaining the main differences.

The observed asymptotic local variance can again be expressed as follows:

$$\widehat{\mathsf{avar}}(\nabla \widehat{\lambda}_t) = \frac{\rho_n}{K_n} \sum_{j=0}^{K_n - 1} \left(\widehat{\lambda}_{t_j} - \widehat{\lambda}_{t_j - \ell_n \Delta_n} \right)^2,$$

where $t_j = t - j\Delta_n$.

This can further be split into

$$\widehat{\mathsf{avar}}(\nabla \widehat{\lambda}_t) = \mathcal{E}_1^n(t) + \mathcal{E}_2^n(t) + \mathcal{R}^n(t),$$

where

$$\mathcal{E}_1^n(t) = \frac{1}{K_n} \sum_{j=0}^{K_n - 1} \left(\Lambda_3^n(t_j) \right)^2,$$

$$\mathcal{E}_2^n(t) = \frac{1}{K_n} \sum_{j=0}^{K_n - 1} \left(\Lambda_5^n(t_j) \right)^2,$$

$$\mathcal{R}^n(t) = \widehat{\mathsf{avar}}(\nabla \widehat{\lambda}_t) - \mathcal{E}_1^n(t) - \mathcal{E}_2^n(t).$$

The only difference here compared to the proof of Theorem 2 is the definition of t_j. It therefore follows that $\mathcal{R}^n(t) = o_p(1)$. So, to complete the proof it suffices to establish the convergence of $\mathcal{E}_1^n(t)$ and $\mathcal{E}_2^n(t)$.

As above, we add and substract $\mu_{(i-1)\Delta_n}$ terms to write

$$\mathcal{E}_1^n(t) = \mathcal{E}_1'^n(t) + \mathcal{E}_1''^n(t),$$

with

$$\mathcal{E}_1'^n(t) = \frac{1}{K_n}\sum_{j=0}^{K_n-1}\left(\left(\Lambda_3^n(t_j)\right)^2 - \frac{\rho_n}{\ell_n^2}\sum_{i\in\mathcal{D}_{t_j}^n}\mu_{(i-1)\Delta_n}\right) \equiv \frac{1}{K_n}\sum_{j=0}^{K_n-1}\epsilon(t_j),$$

$$\mathcal{E}_1''^n(t) = \frac{1}{K_n}\frac{\rho_n}{\ell_n^2}\sum_{j=0}^{K_n-1}\sum_{i\in\mathcal{D}_{t_j}^n}\mu_{(i-1)\Delta_n} = 2\frac{\rho_n}{\ell_n}\frac{1}{2K_n\ell_n}\sum_{j=0}^{K_n-1}\sum_{i=1}^{2\ell_n}\mu_{t-(i+j)\Delta_n}.$$

where $\epsilon(t_j) = \left(\left(\Lambda_3^n(t_j)\right)^2 - \frac{\rho_n}{\ell_n^2}\sum_{i\in\mathcal{D}_{t_j}^n}\mu_{(i-1)\Delta_n}\right)$. As before, $\mathbb{E}\left[\mathcal{E}_1'^n(t)\right] = 0$, but now the variance of $\mathcal{E}_1'^n(t)$ has a more complicated structure due to the overlapping sampling:

$$\mathbb{E}\left[\left(\mathcal{E}_1'^n(t)\right)^2\right] = \frac{1}{K_n^2}\sum_{j=0}^{K_n-1}\mathbb{E}\left[\left(\epsilon(t_j)\right)^2\right] + \frac{2}{K_n^2}\sum_{j=0}^{K_n-2}\sum_{s=j+1}^{K_n-1}\mathbb{E}\left[\epsilon(t_j)\epsilon(t_s)\right].$$

Notice that, for $s > 2\ell_n j$, $\mathcal{D}_{t_j}^n(t) \cap \mathcal{D}_{t_s}^n(t) = \emptyset$, which implies that $\mathbb{E}\left[\epsilon(t_j)\epsilon(t_s)\right] = 0$. On the other hand, for every j and s, $\left|\mathbb{E}\left[\epsilon(t_j)\epsilon(t_s)\right]\right| \leq \mathbb{E}\left[\left(\epsilon(t_j)\right)^2\right]$, since the covariance is always smaller than the variance in absolute value.

As a consequence,

$$\mathbb{E}\left[\left(\mathcal{E}_1'^n(t)\right)^2\right] \leq \frac{1+4\ell_n}{K_n^2}\sum_{j=0}^{K_n-1}\mathbb{E}\left[\left(\epsilon(t_j)\right)^2\right] \leq C\frac{\ell_n}{K_n^2}\sum_{j=0}^{K_n-1}\mathbb{E}\left[\left(\Lambda_3^n(t_j)\right)^4\right],$$

where the second inequality follows from the arguments of Theorem 2. Now, employing the estimate in (37), we deduce that

$$\mathbb{E}\left[\left(\mathcal{E}_1'^n(t)\right)^2\right] = O(K_n^{-1}\ell_n^{-1}\rho_n^2).$$

Now, by the assumptions made in the theorem, it follows that $\mathbb{E}\left[\left(\mathcal{E}_1'^n(t)\right)^2\right] = o(1)$. Hence, the above inequality implies $\mathcal{E}_1'^n(t) \xrightarrow{p} 0$.

In the $\mathcal{E}_1''^n(t)$ term, since by assumption $\ell_n/K_n \to 0$ and $K_n\Delta_n \to 0$, $(K_n + 2\ell_n)\Delta_n \to 0$. This shows that $t-(i+j)\Delta_n \to t$ for every i and j in the sum. Since μ is càdlàg, as $\Delta_n \to 0$,

$$\mathcal{E}_1''^n(t) \xrightarrow{p} \begin{cases} 2\mu_t, & \text{if } \rho_n/\ell_n \to 1, \\ \frac{2}{\theta^2}\mu_t, & \text{if } \rho_n/\ell_n \to \frac{1}{\theta^2}, \\ 0, & \text{if } \rho_n/\ell_n \to 0. \end{cases}$$

Thus, alluding to the definition of ρ_n, we see that

$$\mathcal{E}_1^n(t) \xrightarrow{p} \begin{cases} 2\mu_t, & \text{if } \ell_n\sqrt{\Delta_n} \to 0, \\ \dfrac{2}{\theta^2}\mu_t, & \text{if } \ell_n\sqrt{\Delta_n} \to \theta, \\ 0, & \text{if } \ell_n\sqrt{\Delta_n} \to \infty. \end{cases}$$

■

A.5 Proof of Theorem 4

The statement of the theorem under $\mathscr{H}_0$ follows immediately by combining Theorem 1 and Theorem 3 with Slutsky's theorem.

To complete the proof, we next look at the test statistic under $\mathscr{H}_1$, and we use the version based on the non-overlapping observed asymptotic local variance in (25), $\widetilde{\mathsf{avar}}\big(\nabla\widehat{\lambda}_t\big)$. In that case, the test statistic can be expressed as (writing τ_{ib} as τ):

$$\phi_\tau^{\text{ib}} = \frac{\widetilde{D}_{\tau,0}}{\sqrt{\dfrac{1}{K_n}\displaystyle\sum_{j=0}^{K_n-1}\widetilde{D}_{\tau,j}^2}},$$

where

$$\widetilde{D}_{\tau,j} = \frac{N^n(\tau-(1+2j)\delta_n, \tau-2j\delta_n) - N^n(\tau-(2+2j)\delta_n, \tau-(1+2j)\delta_n)}{n\delta_n},$$

for $j = 0, 1, \ldots, K_n - 1$.

As in the proof of Lemma 1, we derive the stochastic order of the numerator and denominator. Note that it again suffices to explore the diverging part:

$$D_{\tau,j} = \frac{N^{n,\beta}(\tau-(1+2j)\delta_n, \tau-2j\delta_n) - N^{n,\beta}(\tau-(2+2j)\delta_n, \tau-(1+2j)\delta_n)}{n\delta_n},$$

for which

$$\begin{aligned}
\mathbb{E}\big[|D_{\tau,j}|\big] &\leq \mathbb{E}\left[\frac{N^{n,\beta}(\tau-(1+2j)\delta_n, \tau-2j\delta_n) + N^{n,\beta}(\tau-(2+2j)\delta_n, \tau-(1+2j)\delta_n)}{n\delta_n}\right] \\
&= \mathbb{E}\left[\frac{N^{n,\beta}(\tau-(2+2j)\delta_n, \tau-2j\delta_n)}{n\delta_n}\right] = \mathbb{E}\left[\frac{1}{\delta_n}\int_{\tau-(2+2j)\delta_n}^{\tau-2j\delta_n}\sigma_u|\tau-u|^{-\alpha}\mathrm{d}u\right] \\
&\leq \frac{C}{(1-\alpha)\delta_n}\Big(((j+1)\delta_n)^{(1-\alpha)} - (j\delta_n)^{(1-\alpha)}\Big) = O_p(\delta_n^{-\alpha}),
\end{aligned}$$

by the proof of Lemma 1.

Regarding the denominator,

$$\frac{1}{K_n}\sum_{j=0}^{K_n-1}\widetilde{D}_{\tau,j}^2 = O_p\bigg(K_n^{-1}\Big(\big((K_n)^{1-\alpha} - (K_n-1)^{1-\alpha}\big)\delta_n^{-\alpha}\Big)^2\bigg) = O_p(\delta_n^{2\alpha}K_n^{-2\alpha-1}),$$

where the last inequality follows, since

$$\lim_{K_n \to \infty} \frac{K_n^{1-\alpha} - (K_n - 1)^{1-\alpha}}{K_n^{-\alpha}} = C,$$

for a constant C.

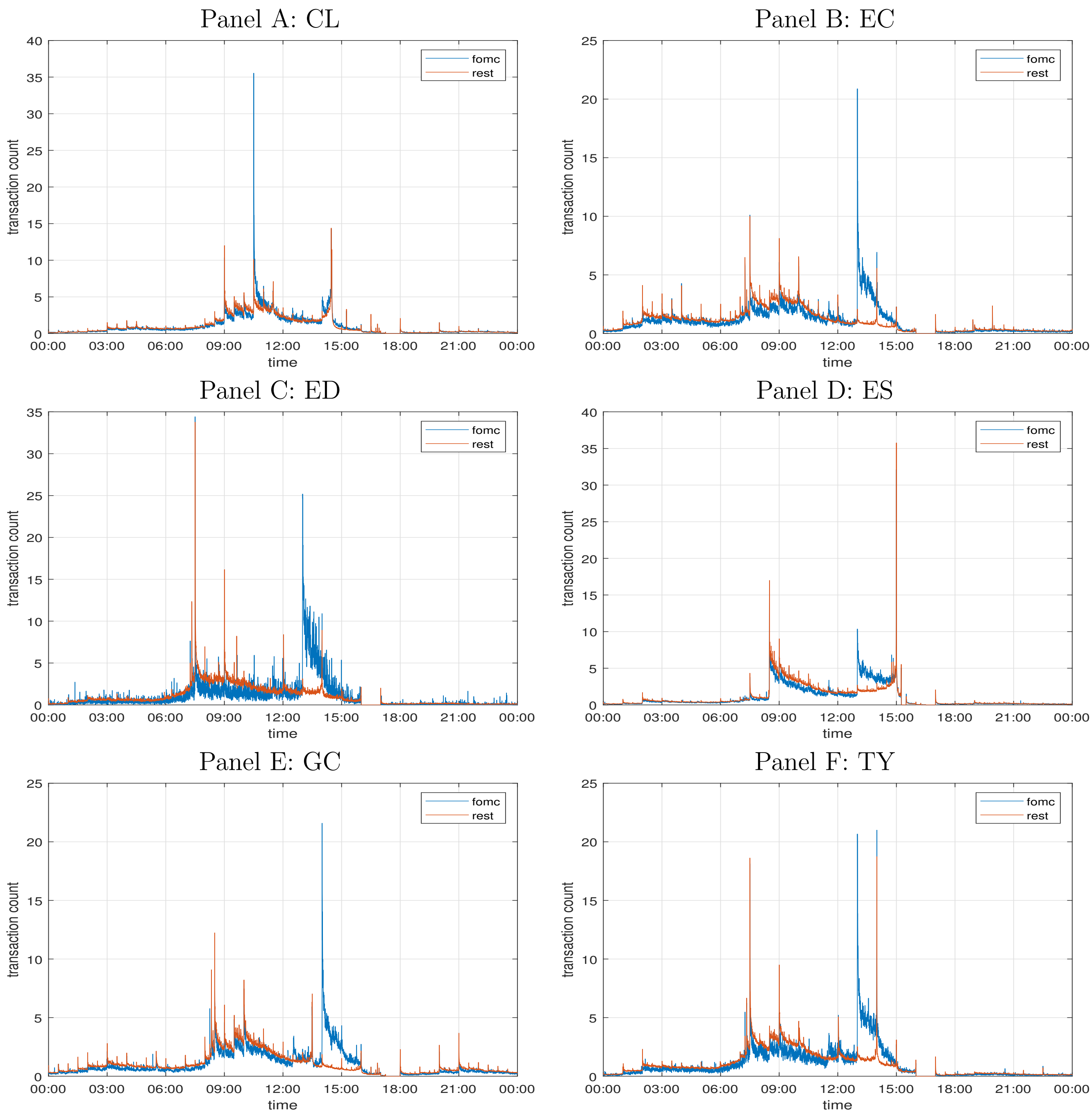
In conclusion,

$$\phi_\tau^{\mathrm{ib}} = O_p\left(K_n^{\frac{1}{2}+\alpha}\right),$$

■

B Diurnal variation in trading intensity.

Figure B.1: Nonparametric estimate of periodicity in intraday trading activity.



Note. We plot a nonparametric estimate of the pointwise time-of-day mean intensity of the transaction count process for each futures contract included in our empirical application. First, we count the number of transactions in each 15-second bucket over a day. Second, we average the transaction count across days in our sample. In doing so, we split the sample into days with and without FOMC announcements. Third, we normalize the estimates so the curve integrates to one.

References

Alvarez, A., F. Panloup, M. Pontier, and N. Savy, 2012, "Estimation of the instantaneous volatility," *Statistical Inference for Stochastic Processes*, 15(1), 27–59.

Andersen, T. G., and T. Bollerslev, 1998, "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts," *International Economic Review*, 39(4), 885–905.

Andersen, T. G., D. Dobrev, and E. Schaumburg, 2008, "Duration-based volatility estimation," Working paper, Northwestern University.

Barndorff-Nielsen, O. E., and N. Shephard, 2002, "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society: Series B*, 64(2), 253–280.

Christensen, K., R. C. A. Oomen, and R. Renò, 2022, "The drift burst hypothesis," *Journal of Econometrics*, 227(2), 461–497.

Christensen, K., M. Podolskij, N. Thamrongrat, and B. Veliyev, 2017, "Inference from high-frequency data: A subsampling approach," *Journal of Econometrics*, 197(2), 245–272.

Clinet, S., and Y. Potiron, 2018, "Statistical inference for the doubly stochastic self-exciting process," *Bernoulli*, 24(4B), 3469–3493.

Daley, D. J., and D. Vere-Jones, 2003, *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods*. Springer-Verlag, Berlin, 2nd edn.

Engle, R. F., and J. R. Russell, 1998, "Autoregressive conditional duration: A new model for irregular spaced transaction data," *Econometrica*, 66(5), 1127–1162.

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

Hayashi, T., J. Jacod, and N. Yoshida, 2011, "Irregular sampling and central limit theorems for power variations : the continuous case," *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 47(4), 1197–1218.

Hong, S. Y., I. Nolte, S. J. Taylor, and X. Zhao, 2021, "Volatility estimation and forecasts based on price durations," *Journal of Financial Econometrics*, Forthcoming, https://doi.org/10.1093/jjfinec/nbab006.

Jacod, J., and P. E. Protter, 2012, *Discretization of Processes*. Springer, Berlin, 2nd edn.

Kalnina, I., 2011, "Subsampling high frequency data," *Journal of Econometrics*, 161(2), 262–283.

Kingman, J. F. C., 1961, "The single server queue in heavy traffic," *Mathematical Proceedings of the Cambridge Philosophical Society*, 57(4), 902–904.

Kristensen, D., 2010, "Nonparametric filtering of the realised spot volatility: A kernel-based approach," *Econometric Theory*, 26(1), 60–93.

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

Mykland, P. A., and L. Zhang, 2017, "Assessment of uncertainty in high frequency data: The observed asymptotic variance," *Econometrica*, 85(1), 197–231.

Politis, D. N., J. P. Romano, and M. Wolf, 1999, *Subsampling*. Springer, Berlin, 1st edn.

Potiron, Y., and P. A. Mykland, 2020, "Local parametric estimation in high frequency data," *Journal of Business and Economic Statistics*, 38(3), 679–692.

Rambaldi, M., V. Filimonov, and F. Lillo, 2018, "Detection of intensity bursts using Hawkes processes: An application to high-frequency financial data," *Physical Review E*, 97(032318), 1–14.

Rambaldi, M., P. Pennesi, and F. Lillo, 2015, "Modeling foreign exchange market activity around macroeconomic news: Hawkes-process approach," *Physical Review E*, 91(012819), 1–15.