# An unbounded intensity model for point processes[*]

Kim Christensen[†]    Aleksey Kolokolov[‡]

May, 2024

## Abstract

We develop a model for point processes on the real line, where the intensity can be locally unbounded without inducing an explosion. In contrast to an orderly point process, for which the probability of observing more than one event over a short time interval is negligible, the bursting intensity causes an extreme clustering of events around the singularity. We propose a nonparametric approach to detect such bursts in the intensity. It relies on a heavy traffic condition, which admits inference for point processes over a finite time interval. With Monte Carlo evidence, we show that our testing procedure exhibits size control under the null, whereas it has high rejection rates under the alternative. We implement our approach on high-frequency data for the EUR/USD spot exchange rate, where the test statistic captures abnormal surges in trading activity. We detect a nontrivial amount of intensity bursts in these data and describe their basic properties. Trading activity during an intensity burst is positively related to volatility, illiquidity, and the probability of observing a drift burst. The latter effect is reinforced if the order flow is imbalanced or the price elasticity of the limit order book is large.

**JEL Classification**: C10; C80.

**Keywords**: Cox process; heavy traffic; high-frequency data; intensity burst; order imbalance; point process; slope of the order book; trading activity.

---

[†]Aarhus University, Department of Economics and Business Economics, Denmark. KIM@ECON.AU.DK.

[‡]University of Manchester, Alliance Manchester Business School, United Kingdom. ALEKSEY.KOLOKOLOV@MANCHESTER.AC.UK.

# 1 Introduction

Trading activity is an important variable in financial economics.[1] In the mixture of distribution hypothesis (e.g. Clark, 1973; Epps and Epps, 1976; Tauchen and Pitts, 1983), trades are driven by a latent stochastic process, interpreted as the flow of news. If the unobservable arrival of news is random and persistent, then so is the observable trading activity. In market microstructure, a number of theories suggest that optimal execution strategies induce time-varying trading activity (e.g. Kyle, 1985; Admati and Pfleiderer, 1988; Almgren and Chriss, 2001). Moreover, the increased volatility typically observed during distressed market conditions often coincide with abnormal increases in trading activity.

The literature has therefore devoted a lot of attention to build models of trading activity (e.g. Engle and Russell, 1998, and references therein and thereto). In practice, the sequence of trades constitute a time series of irregularly spaced high-frequency data, so point processes are a natural starting point. Trading activity can be described by a point process with random—and possibly persistent—intensity, such as an inhomogeneous Poisson process (e.g. Cox, 1955). The rate of trade arrivals can also depend on other observable characteristics, even the history of the process itself as in Hawkes (1971). In the latter class of self-exciting processes, each event increases the likelihood of a new event for a short while. However, the assumption in all this literature is that the intensity process remains locally bounded.

In this paper, we study a point process, where the intensity is, potentially, unbounded over short time intervals. We coin this an intensity burst. We adopt a theoretical foundation, where the random intensity is decomposed into a base intensity, describing "normal" trading activity, and an exploding intensity, describing "abnormal" trading activity. In the normal state, the intensity is locally bounded and follows a doubly stochastic Poisson process. This implies that the probability of observing more than one trade in a short time interval is negligible, so none of the point clusters observed in a fixed window (interpreted as, e.g., a trading day) are substantially larger than the others. In the abnormal state, the intensity is unbounded in the vicinity of a stopping time, which permits the model to produce an extreme concentration of trades in a neighbourhood of that instant. However, the integrated intensity (i.e. the compensator) remains bounded, which ensures that the point process is non-explosive and well-defined.

We propose a framework for intensity burst detection by studying a pointwise test statistic, that is testing for an intensity burst near a single candidate time. We further refine this technique such that a burst in intensity can be separated from a jump in intensity. In contrast to standard inference for point processes, which normally proceeds in a "long-span" setting with time going to infinity, we attempt to detect abnormal surges in trading activity over short time intervals. Hence, we study the "infill" limit, where the time interval is fixed. To permit inference under the null

---

[1]Trading activity is a measure of how much a given financial asset has traded over a period of time. It can be taken as the either the number of transactions, the number of shares traded, or the dollar volume (price times quantity). In this paper, we mainly employ the transaction count.

hypothesis of no intensity burst, we appeal to a novel heavy traffic condition, where independent copies of the point process are stacked to generate an increasing numbers of events over a fixed time interval (e.g. Kingman, 1961). Clinet and Potiron (2018) propose a related inference for the doubly stochastic self-exciting point process (see also Potiron and Mykland, 2020), but they do not exploit it for intensity burst detection.

The heavy traffic assumption is a logical precursor for infill asymptotic theory widely employed in financial econometrics, such as nonparametric estimation of the volatility of arbitrage-free price processes (see, e.g., Andersen and Bollerslev, 1998; Barndorff-Nielsen and Shephard, 2002; Jacod and Protter, 2012). In that setting, the number of observations over a fixed (or shrinking) time interval increases to infinity with the mesh going to zero. This fits directly into our framework. Indeed, a realization of our point process belongs to the class of stochastic sampling times analyzed in Hayashi, Jacod, and Yoshida (2011). We contribute to this field by adapting the observed asymptotic variance of Mykland and Zhang (2017) to local estimation.

Christensen, Oomen, and Renò (2022) suggest that the drift of an asset price (an Itô semi-martingale driven by a Brownian motion) can diverge without inducing a divergence in the price. We extend their concept of a burst to the intensity, which is a natural measure of drift for point processes. However, the construction of our test is much different due to some theoretical subtleties. Moreover, while drift burst detection relies on ultra high-frequency data of the price, our test statistic is based solely on the arrival times. It does not require a further mark in the form of a transaction price or midquote. This has the distinct advantage that our test statistic is unaffected by microstructure noise, which tends to impede high-frequency estimation of the drift and volatility (e.g. Hansen and Lunde, 2006).[2]

Our framework allows to independently screen financial high-frequency data in real-time for abnormal increases in trading activity that may—or may not—be associated with a drift burst or pockets of extreme return persistence, see Andersen, Li, Todorov, and Zhou (2023). To illustrate, in Panel A of Figure 1 we plot the spot EUR/USD exchange rate on May 14, 2019 from 1:00pm to 1:30pm Central European time (CET). The exchange rate drops sharply around 1:15pm. This is detected as a significant event by the drift burst test statistic. In Panel B of Figure 1, we report a nonparametric estimator of the time-varying trade intensity. It too accelerates, before reverting back toward a normal level. Interestingly, there are several short-lived spikes in sell-side activity around the time the exchange rate starts to depreciate, which is indicated by the order imbalance (buyer- minus seller-initiated trades). The figure also plots the intensity burst test statistic. As evident, our approach also identifies this as a significant event. In the empirical application, we shed further light on the dependence between drift and intensity, and we also show how the demand for and supply of liquidity affect this relationship.

---

[2]In fact, a branch of the high-frequency volatility estimation literature has advocated duration-based measures extracted from the point process generating the noisy transaction or quotation data to circumvent this problem (e.g., Andersen, Dobrev, and Schaumburg, 2008; Hong, Nolte, Taylor, and Zhao, 2021).

Figure 1: Intensity burst in the EUR/USD with a drift burst.
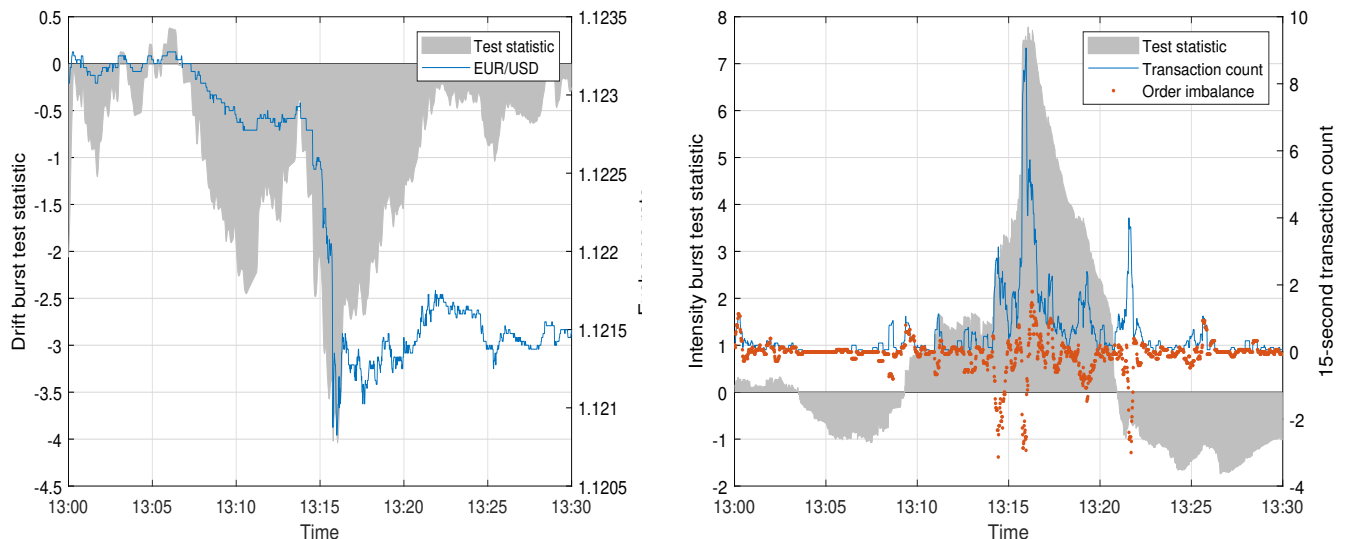


*Note.* This figure shows the EUR/USD spot exchange rate on May 14, 2019, where we detect an intensity burst and a drift burst. In the left panel, we plot the exchange rate from 1:00am to 1:30am CET along with the drift burst test statistic proposed in Christensen, Oomen, and Renò (2022). In the right panel, we plot a nonparametric estimator of the time-varying trading intensity (as proxied by the number of transactions in a 15-second window) along with the intensity burst test statistic and a measure of order imbalance.

Rambaldi, Pennesi, and Lillo (2015) and Rambaldi, Filimonov, and Lillo (2018) study a parametric self-exciting point process in order to explain clusters of events driven by an acceleration in trading activity. In their model, an "intensity burst" is an exogenous (potentially random) number of points added to the counting process at a single (potentially random) time. However, the intensity remains locally bounded, so none of the clusters can describe the enormous concentration of points observed during an intensity burst. Moreover, our approach is nonparametric and the number of burst points can be endogenous.

Bollerslev, Li, and Xue (2018) also study spot trading intensity using high-frequency data based on the nonparametric state-space model of Li and Xiu (2016). They conduct an event study to examine the relationship between jumps in trading volume and return volatility around the release of macroeconomic information, as motivated by the differences of opinion literature (e.g., Kandel and Pearson, 1995). As in our paper, a main ingredient for the analysis is the change in a local intensity estimator at different time points (in their case an ex-ante and ex-post measure before and after the announcement). In contrast to Bollerslev, Li, and Xue (2018), we propose a test statistic for detecting a burst in the intensity process, as formalized in Assumption 1. We show that our test statistic is robust to Lévy-style jump processes and roughness in the intensity. Furthermore, we develop a separate test procedure that builds on the "change-of-frequency" approach often employed in the volatility literature (e.g., Aït-Sahalia and Jacod, 2009; Todorov and Tauchen, 2010). The latter can be applied to separate a jump in intensity from a burst in intensity. In this

framework, the detection of either an intensity burst or an intensity jump is essentially a two-sample mean problem. This is tightly related to the sharp regression discontinuity design (RDD) of Thistlethwaite and Campbell (1960), which has a prominent history in econometric analysis. In our empirical application, there is overwhelming evidence toward our intensity burst test statistic not being driven by a jump component.

The roadmap of the paper is as follows. Section 2 introduces the theoretical foundation of point processes and describes the intensity burst model. Section 3 develops an inference strategy for the proposed test statistic. We also propose an estimator of the asymptotic variance of our test statistic. Section 4 includes an extensive simulation study to demonstrate the finite sample properties of our procedure. An empirical application is conducted in Section 5, while we conclude in Section 6. The proofs are deferred to the Appendix.

# 2  Intensity burst of a point process

We suppose a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ describes a random configuration of ordered points observed on the interval $[0, T]$, i.e. $0 < t_1 < \cdots < t_m < T$. The sequence of random variables $(t_i)_{i=1}^m$ is assumed to be a realization of a simple point process with associated counting process $N = (N_t)_{t\geq 0}$, which is defined as $N_t = \sum_{i\leq t} \mathbb{1}(t_i \leq t)$, where $\mathbb{1}(\cdot)$ is the indicator function. In this paper, we assume $N$ to be a doubly stochastic Poisson—or Cox (1955)—process with associated random intensity—or rate—process $\lambda = (\lambda_t)_{t\geq 0}$, where $\lambda$ is an adapted and strictly positive real-valued stochastic process. That is, conditionally on $\lambda$, $N$ is an inhomogeneous Poisson process with rate $\lambda$, i.e. the conditional characteristic function of the increment $N_{t+\Delta} - N_t$ is given by

$$\varphi_{N_{t+\Delta}-N_t}(u) = \mathbb{E}\Big(e^{iu(N_{t+\Delta}-N_t)} \mid \mathcal{F}_t^\lambda\Big) = \exp\left((e^{iu} - 1)\int_t^{t+\Delta} \lambda_s \mathrm{d}s\right), \tag{2.1}$$

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

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

$$\lambda_t = \lim_{\Delta\to 0} \frac{\mathbb{E}\big[N_{t+\Delta} - N_t \mid \mathcal{F}_t^\lambda\big]}{\Delta}. \tag{2.2}$$

Hence, when $\lambda_t$ is locally bounded, the instantaneous expectation of the number of points is finite. In order to define intensity bursts, we therefore suppose that $\lambda_t$ 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.** *$\lambda_t$ can be decomposed as follows:*

$$\lambda_t = \mu_t + \beta_t, \tag{2.3}$$

4

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

$$\beta_t = \frac{\sigma_t}{|\tau_{\mathrm{ib}} - t|^\alpha}, \tag{2.4}$$

for a stopping time $\tau_{\mathrm{ib}}$, where $\sigma = (\sigma_t)_{t \geq 0}$ is a non-negative locally bounded and stochastically continuous random 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 therefore extends the notion of a drift burst in the asset log-price from Christensen, Oomen, and Renò (2022) to the intensity process generating the discretely observed log-price record. $\mu$ is then the arrival rate of trades during "normal" market conditions, whereas $\beta$ represents the arrival rate of trades during "distressed" market conditions. The stopping time $\tau_{\mathrm{ib}}$ is called the intensity burst time.

If $\sigma_t = 0$ almost surely for all $t$, the explosive term $\beta_t$ is absent and $N_t$ evolves as a simple Cox process. In particular, $N$ is orderly. That is, for any time $t$ it verifies the condition

$$\lim_{\Delta \to 0} \frac{\mathbb{P}(N_{t+\Delta} - N_t > 1)}{\Delta} = 0, \tag{2.5}$$

which implies that the probability of observing more than one event over a short time interval is negligible.

If $\sigma_t > 0$ almost surely, in a neighbourhood of $\tau_{\mathrm{ib}}$,

$$\int_{\tau_{\mathrm{ib}} - \Delta}^{\tau_{\mathrm{ib}} + \Delta} \mu_s \mathrm{d}s = O_p(\Delta) \qquad \text{and} \qquad \int_{\tau_{\mathrm{ib}} - \Delta}^{\tau_{\mathrm{ib}} + \Delta} \beta_s \mathrm{d}s = O_p(\Delta^{1-\alpha}), \tag{2.6}$$

as $\Delta \to 0$. The last condition implies that the left-hand side of equation (2.5) diverges. Hence, $N_t$ is not orderly. This implies that $N$ can generate a large cluster of points near each other in a small neighbourhood of $\tau_{\mathrm{ib}}$. This property, which is crucial for generating extreme point clusters under the heavy traffic condition, cannot be reproduced by classical point processes (for example, a stationary Hawkes process), because any simple and stationary point process is orderly (see, e.g., Section 3.3.V–IV in Daley and Vere-Jones, 2003). However, due to the integrability condition on $\beta_t$ in Assumption 1, the compensator $\Lambda_T = \int_0^T \lambda_s \mathrm{d}s < \infty$ and $\mathbb{P}(N_{t+\Delta} - N_t > 1) = o(\Delta^{1-\alpha})$ as $\Delta \to 0$, 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, to do inference we impose an alternative "heavy traffic" assumption (e.g., Kingman, 1961), in which the intensity of the underlying point process diverges under the null. 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 $n = 1, 2, \ldots$. For a fixed $n$, the rate function of $N_t^n$ is equal to $n\lambda_t$, and

the observed configuration of points is 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:

$$N_t^n = \sum_{i=1}^{n} N_{t,i}, \tag{2.7}$$

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

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 (e.g. 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 $t_0, t_1, \ldots, t_m$ 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 $n \to \infty$. 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 alluded to in high-frequency 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 econometric analysis of financial high-frequency data.

Note that our theoretical exposition does not include self-exiting Hawkes (1971)-type processes, where the stochastic intensity process depends on the realization of $N$. In contrast to a Cox process, stacking independent self-exiting processes to generate an increasing number of events over a fixed time interval result in an poor model, because the correlation between increments of $N^n$ remain independent of $n$, which prevents inference of the local intensity. However, self-excitation can be embedded into our heavy traffic framework in several ways. Clinet and Potiron (2018) study an exponential doubly-stochastic Hawkes process, where the base intensity and excitation function depend on $n$, so the correlation between increments of $N^n$ declines as $n$ increases. Alternatively, one may follow the stacking construction in (2.7) without the independence assumption. We postpone a theoretical treatment of this problem for future research. On a practical side, we explore the Hawkes processes as a model for the "normal" time event flow in the Monte Carlo section.

# 3  Identification

The detection of intensity bursts amounts to the following hypothesis:

$$\mathcal{H}_0 : \omega \in \Omega_0 \qquad \text{and} \qquad \mathcal{H}_1 : \omega \in \Omega_1, \tag{3.1}$$

where $\Omega_0$ and $\Omega_1$ are complementary subsets of $\Omega$:

$$\Omega_0 = \left\{ \omega \in \Omega : \int_0^T \beta_t^2 \mathrm{d}t = 0 \right\} \qquad \text{and} \qquad \Omega_1 = \left\{ \omega \in \Omega : \int_0^T \beta_t^2 \mathrm{d}t > 0 \right\}. \tag{3.2}$$

In this section, we propose a formal testing procedure to figure out which subset the realization of $N^n$ belongs to. In particular, we propose a test that conducts a statistical inference for the presence of an intensity burst near a single candidate time instance.

## 3.1 Test statistic

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

$$\widehat{\lambda}_t = \frac{N^n(t - \delta_n, t)}{n\delta_n}, \tag{3.3}$$

where $N^n(a, b) = N_b^n - N_a^n$ is the number of points of the counting process on $(a, b]$, and $\delta_n$ is a bandwidth parameter.[3]

$\widehat{\lambda}_t$ is nothing more than the realized intensity, i.e. the number of counts per time unit over an interval of length $\delta_n$, averaged across Cox processes. In a setup with light traffic (i.e. $n$ fixed), it follows that $\widehat{\lambda}_t \xrightarrow{p} 0$. However, heavy traffic (i.e. $n \to \infty$) changes the stochastic properties 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*

$$\widehat{\lambda}_t \xrightarrow{p} \mu_{t-}. \tag{3.4}$$

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

$$\widehat{\lambda}_{\tau_{\mathrm{ib}}} = O_p\big(\delta_n^{-\alpha}\big). \tag{3.5}$$

The lemma shows that, under general assumptions about the intensity process, the intensity estimator converges to the left-limit of the local intensity at time $t$ under the null hypothesis. 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. To generate a sufficient number of points inside the estimation window $[t - \delta_n, t]$, we need $\delta_n n \to \infty$. A law of large numbers then shows that $\widehat{\lambda}_t$ converges in probability and, as $\delta_n \to 0$, the limit is $\mu_{t-}$.

On the other hand, under the alternative, the arrival rate of the bursting process causes the estimator to explode around $\tau_{\mathrm{ib}}$.

Lemma 1 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 the 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)

---

[3]The presence of $n$ in (3.3) is inconvenient, since the parameter is not observed. However, it is only needed to show convergence of $\widehat{\lambda}_t$. As we show below, $n$ drops out for the construction of the intensity burst test statistic.

to construct a drift burst test statistic, which relies on the different rates of divergence of the drift estimator under the null and alternative. Their test statistic has the convenient side effect that it knocks out the true drift coefficient asymptotically. By contrast, under the null hypothesis $\widehat{\lambda}_t$ is consistent for the local intensity of the point process in the heavy traffic limit. Hence, for intensity burst detection we propose a slightly different test statistic compared to Christensen, Oomen, and Renò (2022). In particular, to center the statistic we compare two local intensity estimators calculated from the nearest lagged block of observations:

$$\nabla\widehat{\lambda}_t = \widehat{\lambda}_t - \widehat{\lambda}_{t-\delta_n}. \tag{3.6}$$

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

$$\phi_t^{\mathrm{ib}} = \frac{\nabla\widehat{\lambda}_t}{\widehat{\mathrm{std}}(\nabla\widehat{\lambda}_t)}, \tag{3.7}$$

where $\widehat{\mathrm{std}}(\nabla\widehat{\lambda}_t)$ is an estimator of $\mathrm{std}(\nabla\widehat{\lambda}_t)$ under $\mathscr{H}_0$.

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

**Assumption 2.** *$\mu$ is an Itô semimartingale of the form:*

$$\mu_t = \mu_0 + \int_0^t a_s \mathrm{d}s + \int_0^t \nu_s \mathrm{d}W_s + \int_0^t \int_{|\delta_s(x)| \leq 1} \delta_s(x) \left(\mathfrak{p}(\mathrm{d}s, \mathrm{d}x) - \mathfrak{q}(\mathrm{d}s, \mathrm{d}x)\right) + \int_0^t \int_{|\delta_s(x)| > 1} \delta_s(x)\mathfrak{p}(\mathrm{d}s, \mathrm{d}x), \tag{3.8}$$

*where $\mu_0$ is $\mathcal{F}_0$-measurable, $a = (a_t)_{t\geq 0}$ and $\nu = (\nu_t)_{t\geq 0}$ are adapted, càdlàg stochastic processes, $W$ is a standard Brownian motion, $\mathfrak{p}$ is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}$ with compensator $\mathfrak{q}(\mathrm{d}s, \mathrm{d}x) = \mathrm{d}sF(\mathrm{d}x)$, where $F$ is a $\sigma$-finite measure on $\mathbb{R}$, and $\delta_s(x)$ is a predictable function such that there exists a sequence of $\mathcal{F}_t$-stopping times $(\tau_n)_{n=1}^\infty$ with $\tau_n \to \infty$ and, for each $n$, a deterministic and non-negative function $\Gamma_n$ with $\min\left(|\delta_t(x)|, 1\right) \leq \Gamma_n(x)$ for $t \leq \tau_n$, $\int_{\mathbb{R}} \Gamma_n(x)^2 F(\mathrm{d}x) < \infty$, and $\int_{\{x:\Gamma_n(x)\leq\kappa\}} \Gamma_n(x)F(\mathrm{d}x) < \infty$, for all $(t, x)$ and $\kappa \in (0, 1)$.*

Assumption 2 is common in high-frequency analysis (e.g. Jacod and Protter, 2012), since it allows to apply estimates for semimartingales to control the discretization error. It is a standard tool often used to describe arbitrage-free price processes, since these must be semimartingale (Delbaen and Schachermayer, 1994). In that context, the assumption is—apart from the integrability condition on the driving processes—nonparametric, as it does not restrict the dynamic of the price and allows it to exhibit time-varying drift, stochastic volatility, and a more or less unrestricted jump component, which can be both infinitely active and of infinite variation. Moreover, the various terms can be arbitrarily dependent.

However, since we are looking at an intensity process and not the price of a tradable asset, we

are not bound by the no free lunch principle. We therefore also inspect an alternative framework below, which replaces the above structural assumption by a general smoothness condition.

**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 $\delta_n\sqrt{n} \to 0$,*

$$\sqrt{n\delta_n}\left(\widehat{\lambda}_t - \mu_{t-}\right) \xrightarrow{\mathfrak{D}_s} \sqrt{\mu_{t-}}\mathcal{Z}, \tag{3.9}$$

*and*

$$\sqrt{n\delta_n}\nabla\widehat{\lambda}_t \xrightarrow{\mathfrak{D}_s} \sqrt{2\mu_{t-}}\mathcal{Z}, \tag{3.10}$$

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

$$\frac{1}{\sqrt{\delta_n}}\left(\widehat{\lambda}_t - \mu_{t-}\right) \xrightarrow{\mathfrak{D}_s} \sqrt{\frac{1}{\theta^2}\mu_{t-} + \frac{1}{3}\nu_t^2}\mathcal{Z}, \tag{3.11}$$

*and*

$$\frac{1}{\sqrt{\delta_n}}\nabla\widehat{\lambda}_t \xrightarrow{\mathfrak{D}_s} \sqrt{\frac{2}{\theta^2}\mu_{t-} + \frac{8}{3}\nu_t^2}\mathcal{Z}, \tag{3.12}$$

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

$$\frac{1}{\sqrt{\delta_n}}\left(\widehat{\lambda}_t - \mu_{t-}\right) \xrightarrow{\mathfrak{D}_s} \sqrt{\frac{1}{3}\nu_t^2}\mathcal{Z}, \tag{3.13}$$

*and*

$$\frac{1}{\sqrt{\delta_n}}\nabla\widehat{\lambda}_t \xrightarrow{\mathfrak{D}_s} \sqrt{\frac{8}{3}\nu_t^2}\mathcal{Z}, \tag{3.14}$$

*where $\mathcal{Z}$ 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 $\mathcal{F}$, 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 mixed normal, but the convergence rate and limiting variance depends on the order of $\delta_n$. 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, by the Poisson distribution, the variance is equal to the mean. On the other hand, if $\delta_n\sqrt{n} \to \infty$, the roles are reversed, and the variation of the 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 and both affect the limiting distribution.

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 to (i). The explanation is that $n$ diverges so fast compared to the shrinking of $\delta_n$ that 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.

Assumption 2 prevents the baseline intensity to be described by a non-semimartingale, such as a fractional Brownian motion. The latter is a popular model in the realm of asset return variation, since it can generate roughness and long memory (e.g. Comte and Renault, 1998; Gatheral, Jaisson, and Rosenl 2018). To the extent that trading activity and return volatility are connected, which holds empirically and is also a standard implication of the information flow interpretation of the mixture of distribution hypothesis (e.g., Andersen, 1996; Clark, 1973; Tauchen and Pitts, 1983), we should feel uncomfortable with that restriction. It can, however, be replaced by an appropriate smoothness condition, such as assuming the sample paths of $\mu$ are Hölder continuous.

**Assumption 3.** *There exists $H^\mu > 0$, such that*

$$\mathbb{E}\left[\sup_{s \le t \le s+\Delta} |\mu_s - \mu_t|\right] \le C\Delta^{H^\mu}, \tag{3.15}$$

*for some constant $C > 0$ and all $0 \le s \le t \le T$.*

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

$$\sqrt{n\delta_n}\left(\widehat{\lambda}_t - \mu_{t-}\right) \xrightarrow{\mathfrak{D}_s} \sqrt{\mu_{t-}}\mathcal{Z}, \tag{3.16}$$

Theorem 2 shows our spot intensity estimator is consistent and asymptotically normal under Assumption 3 with an unaltered convergence rate. However, it only covers the short bandwidth setting in part (i) of Theorem 1, where the local variation of the intensity process does not affect the asymptotic conditional variance of the estimator. This is not true for part (ii) and (iii), and that is where the structure of Assumption 2 is helpful. Without such knowledge, it is difficult to extend Theorem 2 in that direction.

The split points in Theorem 1 are dictated by the behavior of $\delta_n\sqrt{n}$, since for Brownian motion (the leading term is Assumption 2), the probabilistic order of its increment is $O_p(\sqrt{\delta_n})$. This condition is replaced in Theorem 2 by $\delta_n n^{1-\frac{2H^\mu}{3}}$. Thus, if $\delta_n n^{1-\frac{2H^\mu}{3}} \to 0$, we are again in the short bandwidth setting vis-à-vis the Hölder regularity of the process. However, even for $H^\mu = 1/2$ (the index of Brownian motion), the rate condition reduces to $\delta_n n^{2/3} \to 0$ and is more restrictive than before. This is because in the proof of Theorem 1, we also exploit that Brownian motion is a martingale.

At last, we should stress that the dual rate conditions enforced by Theorem 2 are in a direct conflict of interest. On the one hand, it should hold that $n\delta_n \to \infty$, so that there are a sufficient number of observations inside the estimation window. One the other hand, for $H^\mu$ very small we more or less also need $n\delta_n \to 0$. This is because the smoothness condition imposed in kernel-based nonparametric analysis requires a shorter bandwidth for rough processes.[4]

---

[4]If the localization window is minuscule, there may be too few observations inside it to get the mechanics of the central limit theorem working, risking that the asymptotic theory is a poor approximation to the finite sampling distribution of the estimator. Bugni, Li, and Li (2023) also study non-price high-frequency data employing the

## 3.2 Observed asymptotic local variance

In practice, picking the correct estimator of the asymptotic variance is 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.[5] We set $\delta_n = \ell_n \Delta_n$, where $\ell_n$ 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:

$$\widehat{\lambda}_t = \frac{1}{\ell_n} \sum_{i \in \mathcal{D}_{t-}^n} \Delta_i N^n, \tag{3.17}$$

where $\Delta_i N^n = N^n(i\Delta_n, (i-1)\Delta_n)$ is the increment of $N_t^n$ over a short time interval and $\mathcal{D}_{t-}^n = \{tn - \ell_n + 1, tn - \ell_n + 2, \ldots, tn\}$.

The formulation in (3.17) follows the traditional high-frequency approach. It shows that the intensity estimator can be expressed as a local average of increments to the discretized point process $N_t^n$, which is a more convenient formulation for developing our asymptotic variance estimator.[6] With this convention, Theorem 1 can be stated as follows.

**Corollary 1.** *Suppose that Assumptions 1 and 2 hold. Then, under $\mathcal{H}_0$, 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*

$$\sqrt{\rho_n} \nabla \widehat{\lambda}_t \xrightarrow{\mathfrak{D}_s} \sqrt{\mathsf{avar}(\nabla \widehat{\lambda}_t)} \mathcal{Z}, \tag{3.18}$$

*where*

$$\left( \rho_n, \mathsf{avar}\left(\nabla \widehat{\lambda}_t\right) \right) = \begin{cases} (\ell_n, 2\mu_{t-}), & \text{if } \ell_n \sqrt{\Delta_n} \to 0, \\ ((\ell_n \Delta_n)^{-1}, \frac{2}{\theta^2} \mu_{t-} + \frac{8}{3} \nu_t^2), & \text{if } \ell_n \sqrt{\Delta_n} \to \theta > 0, \\ ((\ell_n \Delta_n)^{-1}, \frac{8}{3} \nu_t^2), & \text{if } \ell_n \sqrt{\Delta_n} \to \infty. \end{cases} \tag{3.19}$$

Then, we propose to set

$$\widetilde{\mathsf{avar}}\left(\nabla \widehat{\lambda}_t\right) = \frac{\rho_n}{K_n} \sum_{j=0}^{K_n - 1} \left( \widehat{\lambda}_{t - 2j\ell_n \Delta_n} - \widehat{\lambda}_{t - (\ell_n + 2j\ell_n)\Delta_n} \right)^2, \tag{3.20}$$

---

nonparametric state-space model of Li and Xiu (2016). They propose a permutation-based test to detect general distributional discontinuities, which includes the local two-sample mean problem. The advantage of their framework is its finite sample validity, which addresses potential concerns caused by roughness. However, developing such a permutation-based test is beyond the scope of the paper.

[5]An alternative inference procedure is subsampling (e.g. Politis, Romano, and Wolf, 1999), which was adapted to high-frequency data in Kalnina (2011) and Christensen, Podolskij, Thamrongrat, and Veliyev (2017).

[6]In our framework, $N^n$ is observed in continuous-time. However, many existing datasets do not reveal the exact location of all events from a point process, but only report discrete observations. For example, the number of trades in every 1-second slot may be available. In such cases, the data is structured in the form of a discretized version of $N^n$ used in equation (3.17). Thus, Corollary 1 and subsequent theorems remains applicable to such data.

where $K_n$ is another sequence of natural numbers.

The observed asymptotic local variance, $\widetilde{\mathsf{avar}}(\nabla\widehat{\lambda}_t)$, 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 near $t$, $\widetilde{\mathsf{avar}}(\nabla\widehat{\lambda}_t)$ is consistent for $\mathsf{avar}(\nabla\widehat{\lambda}_t)$.

**Theorem 3.** *Suppose that Assumptions 1 and 2 hold. Then, under $\mathcal{H}_0$, 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*

$$\widetilde{\mathsf{avar}}(\nabla\widehat{\lambda}_t) \xrightarrow{p} \mathsf{avar}(\nabla\widehat{\lambda}_t). \tag{3.21}$$

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

The condition $K_n\ell_n\Delta_n \to 0$ in Theorem 3 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:

$$\widehat{\mathsf{avar}}(\nabla\widehat{\lambda}_t) = \frac{\rho_n}{K_n} \sum_{j=0}^{K_n-1} \left(\widehat{\lambda}_{t-j\Delta_n} - \widehat{\lambda}_{t-(\ell_n+j)\Delta_n}\right)^2. \tag{3.22}$$

**Theorem 4.** *Suppose that Assumptions 1 and 2 hold. Then, under $\mathcal{H}_0$, 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*

$$\widehat{\mathsf{avar}}(\nabla\widehat{\lambda}_t) \xrightarrow{p} \mathsf{avar}(\nabla\widehat{\lambda}_t). \tag{3.23}$$

The condition $\ell_n/K_n \to 0$ in Theorem 4 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 (3.7) becomes:

$$\phi_t^{\text{ib}} = \frac{\sqrt{\rho_n}\nabla\widehat{\lambda}_t}{\sqrt{\widehat{\mathsf{avar}}(\nabla\widehat{\lambda}_t)}}. \tag{3.24}$$

**Theorem 5.** *Suppose that the conditions of Theorems 1 and 4 hold. Then,*

$$\begin{cases} \phi_t^{\text{ib}} \xrightarrow{\mathfrak{D}} N(0,1), & \text{conditional on } \mathcal{H}_0, \\ \phi_{\tau_{\text{ib}}}^{\text{ib}} \xrightarrow{p} \infty, & \text{conditional on } \mathcal{H}_1. \end{cases} \tag{3.25}$$

Hence, we reject $\mathcal{H}_0$ if $\phi_t^{\text{ib}}$ exceeds a quantile in the right-hand tail of the standard normal distribution corresponding to a given significance level $\varsigma$. This ensures the test has size control under $\mathcal{H}_0$. On the other hand, under $\mathcal{H}_1$, $\phi_t^{\text{ib}}$ is unbounded in probability as $t \to \tau_{\text{ib}}$, so the test is also consistent.[7]

---

[7]Theorem 5 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 $\phi_t^{\text{ib}}$, provided the conditions of Theorem 3 hold.

It is important to emphasize that we do not need to choose $\rho_n$ to compute $\phi_t^{\text{ib}}$. This follows by direct insertion, showing the test statistic can be expressed as:

$$\phi_t^{\text{ib}} = \frac{N^n(t - \delta_n, t) - N^n(t - 2\delta_n, t - \delta_n)}{\sqrt{\dfrac{1}{K_n} \displaystyle\sum_{j=0}^{K_n - 1} \left(N^n(t - j\Delta_n - \delta_n, t - j\Delta_n) - N^n(t - j\Delta_n - 2\delta_n, t - j\Delta_n - \delta_n)\right)^2}}. \tag{3.26}$$

In practice, we take $\Delta_n$ to be one second, although this is merely a convenient choice. The selection of $\ell_n$ and $K_n$ are studied in the Monte Carlo analysis presented in Section 4.

**Remark 1.** *The local intensity estimator defined in* (3.17) *can be rewritten as a standard kernel-based estimator of the form*

$$\widehat{\lambda}_t = \frac{1}{\ell_n} \sum_{i=1}^{\lfloor T\Delta_n \rfloor} \mathcal{K}\left(\frac{i\Delta_n - t}{\ell_n}\right) \Delta_i N^n, \tag{3.27}$$

*where $\ell_n$ is the bandwidth and $\mathcal{K}$ is the indicator kernel $\mathcal{K}(x) = \mathbf{1}_{[-1,0]}(x)$. It is of course possible to select a different kernel. While the full development of a corresponding theory is outside of the scope of our paper, we conjecture that Theorem 5 remains valid under suitable regularity conditions for the kernel and bandwidth (e.g. Kristensen, 2010; Christensen, Oomen, and Renò, 2022). We investigate the exponential kernel in the simulation analysis and empirical application.*

## 3.3 How to separate a burst from a jump?

Theorem 1 and Theorem 5 show that—in the limit—our test statistic is robust to jumps in the intensity process at the test time $t$. The intuition is that the main term in the numerator is $\widehat{\lambda}_t - \widehat{\lambda}_{t-\delta_n}$. With a fixed $t$ and a sufficiently small $\delta_n$, both $[t - \delta_n, t]$ and $[t - 2\delta_n, t - \delta_n]$ are close to $t$, so any jumps (of vanishing size as $\delta_n \to 0$) from a left neighbourhood of $t$ fall outside the estimation windows, so $\widehat{\lambda}_t$ and $\widehat{\lambda}_{t-\delta_n}$ converge to the left-limit $\mu_{t-}$, and the latter is not affected by the behaviour of the intensity process at time $t$.

In finite samples, $\delta_n$ is fixed. In that case, $\widehat{\lambda}_t - \widehat{\lambda}_{t-\delta_n}$ may numerically coincide with the difference between a forward- and backward-looking intensity estimator:

$$\widehat{\lambda}_t^{(-)} = \frac{1}{\ell_n} \sum_{i \in \mathcal{D}_{t-}^n} \Delta_i N^n, \qquad \text{and} \qquad \widehat{\lambda}_t^{(+)} = \frac{1}{\ell_n} \sum_{i \in \mathcal{D}_{t+}^n} \Delta_i N^n, \tag{3.28}$$

where $\mathcal{D}_{t-}^n = \{tn - \ell_n + 1, tn - \ell_n + 2, \ldots, tn\}$ and $\mathcal{D}_{t+}^n = \{tn + 1, tn + 2, \ldots, tn + \ell_n\}$. $\widehat{\lambda}^{(-)}$ is equivalent to $\widehat{\lambda}_t$ and thus has probability limit $\mu_{t-}$, whereas $\widehat{\lambda}_t^{(+)}$ converges to the right-limit $\mu_{t+} = \mu_t$. As such, one may suspect that the real cause of the rejection was in fact due to a jump in the intensity process, i.e. $\mu_t \neq \mu_{t-}$.

In this section, we dive into this issue and propose a framework that can separate a burst from a jump. We assume to know the existence of a time instant $\theta$, where there is a change-point in the intensity process–either a burst or a jump. $\theta$ can be a stopping time, a known point of pre-scheduled

news announcements, or perhaps the center of an intensity burst that has already been detected with our testing procedure.

The crest of our approach is to study the limiting behaviour of a variant of our intensity estimator around the change point. We build on the "change-of-frequency" approach often employed in the volatility literature (e.g., Aït-Sahalia and Jacod, 2009; Todorov and Tauchen, 2010) and propose a two-sided local intensity estimator:

$$\widetilde{\lambda}_t(\delta_n) = \frac{1}{2\ell_n} \sum_{i=tn-\ell_n+1}^{tn+\ell_n} \Delta_i N = \frac{N(t-\delta_n, t+\delta_n)}{2n\delta_n}. \tag{3.29}$$

We highlight the dependence of $\widetilde{\lambda}_t$ on the bandwidth, because discrimination between a burst and a jump relies on sampling the estimator with different $\delta_n$.

To state a more concrete result, we let $\Omega_0^H \subseteq \Omega_0$ be the subset of the sample space, where the intensity process fulfills Assumption 3:

$$\Omega_0^H = \left\{ \omega \in \Omega_0 : \sup_{s \leq t \leq s+\Delta} |\mu_s - \mu_t| < L^\mu \Delta^{H^\mu} \right\}, \tag{3.30}$$

where $L^\mu$ is an almost surely finite positive random variable. Note that there is no intensity burst on $\Omega_0^H$. Next, define $\Omega_1^J \subseteq \Omega$ such that:

$$\Omega_1^J = \left\{ \omega \in \Omega : \exists \theta \in (0,T) \, \exists C > 0 : |\Delta\mu_\theta| > C \quad \text{and} \quad \exists \epsilon > 0 : (\mu_t)_{t \in [\theta-\epsilon, \theta+\epsilon] \setminus \{\theta\}} \in \Omega_0^H \right\}, \tag{3.31}$$

where $\Delta\mu_t = \mu_t - \mu_{t-}$. The interpretation is that on $\Omega_1^J$, there is a jump in $\mu$ of size $\Delta\mu_\theta$ at time $\theta$, but otherwise the intensity process is smooth (and, in particular, there is no intensity burst) in a closed ball around this point. At last, we consider the set $\Omega_1^B \subseteq \Omega_1$ such that there is an intensity burst with center at $\tau_{\text{ib}} = \theta$.

As we show below, it turns out that the pathwise properties of $\widetilde{\lambda}_t(\delta_n)$ are different on $\Omega_1^J$ and $\Omega_1^B$. This can be exploited to discriminate an intensity burst from an intensity jump. However, to develop an asymptotic theory for $\widetilde{\lambda}_t(\delta_n)$ in the presence of an intensity burst, we need a smoothness condition on $\sigma$ (the volatility of the burst process) vis-á-vis the above.

**Assumption 4.** *There exists $H^\sigma > \alpha/2$, such that*

$$\sup_{s \leq t \leq s+\Delta} |\sigma_s - \sigma_t| < L^\sigma \Delta^{H^\sigma}, \tag{3.32}$$

*where $L^\sigma$ is an almost surely finite positive random variable, and $\alpha$ is the explosion rate of the intensity burst from Assumption 1.*

Now, we are able to establish the following lemma.

**Lemma 2.** *Suppose that Assumptions 1 and 4 hold and let $k > 0$ be a positive number. As $n \to \infty$ and $\delta_n \to 0$ such that $n\delta_n \to \infty$ and $\delta_n n^{1-\frac{2H^\mu}{3}} \to 0$, it holds that*

(i) Conditional on $\Omega_1^B$,

$$\delta_n^{\alpha/2}\left(\widetilde{\lambda}_\theta(k\delta_n) - \sigma_\theta\left(2k\delta_n\right)^{-\alpha}\right) \xrightarrow{p} 0 \qquad \text{and} \qquad \frac{\widetilde{\lambda}_\theta(k\delta_n)}{\widetilde{\lambda}_\theta(\delta_n)} \xrightarrow{p} k^{-\alpha}. \tag{3.33}$$

(ii) Conditional on $\Omega_1^J$,

$$\widetilde{\lambda}_\theta(k\delta_n) \xrightarrow{p} \mu_{\theta-} + \frac{\Delta\mu_\theta}{2} \qquad \text{and} \qquad \frac{\widetilde{\lambda}_\theta(k\delta_n)}{\widetilde{\lambda}_\theta(\delta_n)} \xrightarrow{p} 1. \tag{3.34}$$

The first part of Lemma 2 suggests that on $\Omega_1^B$ we can construct a direct estimator of the explosion rate of an intensity burst, $\alpha$, via a logarithmic transformation:

$$\widehat{\alpha} = -\frac{\log\left(\widetilde{\lambda}_\theta(k\delta_n)\right) - \log\left(\widetilde{\lambda}_\theta(\delta_n)\right)}{\log k} \xrightarrow{p} \alpha. \tag{3.35}$$

On $\Omega_1^J$, this estimator converges to zero in probability as it should be in the absence of an intensity burst, i.e. $\hat{\alpha} \xrightarrow{p} 0$.

A more efficient estimator of $\alpha$ can be retrieved by calculating $\log\left(\widetilde{\lambda}_\theta(k\delta_n)\right)$ for a range of $k$ and regressing these against $\log(k)$. The negative of the slope estimate is then $\hat{\alpha}$. We explore this possibility in the empirical application.

The testing procedure to separate the competing hypothesis of a jump or a burst in the intensity process can now be conducted with either $\Omega_1^B$ or $\Omega_1^J$ as the null, but it requires a formal derivation of the asymptotic distribution of the ratio statistic in that state. If the prior is that intensity has jumped, then a statistical test to reject this conviction can be deduced from the following result.

**Theorem 6.** *Suppose that Assumption 1 holds. Conditional on $\Omega_1^J$, as $n \to \infty$ and $\delta_n \to 0$ such that $n\delta_n \to \infty$ and $\delta_n n^{1-\frac{2H^\mu}{3}} \to 0$, it follows that*

$$\sqrt{n\delta_n}\left(\frac{\widetilde{\lambda}_\theta(k\delta_n)}{\widetilde{\lambda}_\theta(\delta_n)} - 1\right) \xrightarrow{\mathfrak{D}_s} \sqrt{\mathrm{avar}_k}\mathcal{Z}, \tag{3.36}$$

*where $\mathcal{Z}$ is as in Theorem 1, and*

$$\mathrm{avar}_k = \left(\mu_{\theta-} + \frac{\Delta\mu_\theta}{2}\right)^{-1}\left(\frac{1}{2} + \frac{1}{2k} - \left[k\left(\mu_{\theta-} + \frac{\Delta\mu_\theta}{2}\right)\right]^{-1}\right). \tag{3.37}$$

Conversely, $\frac{\widetilde{\lambda}_\theta(k\delta_n)}{\widetilde{\lambda}_\theta(\delta_n)} \xrightarrow{p} k^{-\alpha}$ conditional on $\Omega_1^B$ such that $\sqrt{n\delta_n}\left(\frac{\widetilde{\lambda}_\theta(k\delta_n)}{\widetilde{\lambda}_\theta(\delta_n)} - 1\right) \to -\infty$.

# 4 Simulation study

We investigate the small sample properties of our intensity burst test statistic by a Monte Carlo approach. We set $T = 1$ and interpret [0,1] 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.

We simulate the aggregate intensity as a multiplicative process of the form $\mu_t = \mu_{s,t}\mu_{d,t}$, where the first term is a stationary process and the second term is a recurrent diurnal component. We assume the stationary part is a continuous-time realization of independent point processes representing the "normal" and "burst" intensity, $N_{s,t} = N_{s,t}^{\text{normal}} + N_{s,t}^{\text{burst}}$, where the latter is only active under the alternative.

To gauge the robustness of our approach, we inspect several choices of $N_{s,t}^{\text{normal}}$. We start with a standard homogenous Poisson process that has constant intensity $\mu_{s,t} = 1$ (measured on a second-by-second basis).

The second extends the above to an inhomogenous Poisson process with stochastic intensity of the Cox, Ingersoll, and Ross (1985) form:

$$\mu_{s,t} = \lambda(1 - e^{-\kappa t}) + \mu_{s,0}e^{-\kappa t} + \gamma \int_0^t e^{-\kappa(u-s)}\sqrt{\mu_{s,u}}\,\mathrm{d}W_u. \tag{4.1}$$

Here, the intensity is time-varying and mean-reverting. If the current value of the process, $\mu_{s,t}$, is above its steady state level, $\lambda$, the intensity has a tendency to go down, and vice versa. $\kappa$ controls the speed of mean reversion. $\gamma$ is the volatility. The scaling with $\sqrt{\mu_{s,t}}$ inside the stochastic integral prevents the process from being negative. This model has been applied to describe the default intensity in a credit risk framework by Duffie (2005) and time-varying volatility in Heston (1993). We select the parameter vector as $(\lambda, \kappa, \gamma) = (1.00, 0.03, 0.20)$ following Li, Todorov, and Tauchen (2013). To ensure the process is stationary with $\mathbb{E}(\mu_{s,t}) = 1$, we draw $\mu_{s,0}$ at random from $\text{Gamma}(2\lambda\kappa\gamma^{-2}, 2\kappa\gamma^{-2})$.

To generate a more challenging data-generating process under the null, as our last choice we assume the intensity of $N_{s,t}^{\text{normal}}$ is an exponential Hawkes process:[8]

$$\mu_{s,t} = \lambda_0 + \int_0^t \theta e^{-\kappa(u-s)}\mathrm{d}N_{s,u}^{\text{normal}}, \tag{4.2}$$

where $\lambda_0$, $\theta$ and $\kappa$ are parameters. $\lambda_0$ is called the background intensity, which is a lower bound on $\mu_{s,t}$, while $f(\tau) = \theta e^{-\kappa\tau}$ is the excitation function. Our choice of kernel follows the original article by Hawkes (1971).

A Hawkes process is self-exciting. That is, after the arrival of an event the intensity of $N_{s,t}^{\text{normal}}$ inclines by $\theta$. Hence, the probability of an increment in the next time interval $t + \mathrm{d}t$ increases. Tempering by the excitation function pulls the intensity back toward its baseline level $\lambda_0$ until further events occur. $\kappa$ controls the rate of 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

$$\mathbb{E}(\mu_t) = \frac{\lambda_0}{1 - \theta/\kappa}. \tag{4.3}$$

The Hawkes process is also capable of generating event clusters. However, none of them are as large

---

[8]Although the intensity of the Hawkes process fulfills Assumption 2, it falls outside of the Cox framework in (2.1). This part of the simulations is therefore also meant to illustrate the robustness of our testing procedure.

as one expects during an intensity burst. In the former, the intensity remains bounded, almost surely, whereas it diverges in the latter. Nevertheless, this may be confused by our test statistic as an intensity burst. The purpose is to see if this distorts our test statistic. As in the diffusion model, the mean reversion is $\kappa = 0.030$. We further select $\lambda_0 = 0.500$ and $\theta = (1 - \lambda_0)\kappa = 0.015$, which again ensures that $\mathbb{E}(\mu_{s,t}) = 1$.

In practice, trading intensity (and return volatility) follows a deterministic U- or inverted J-shape at the intraday horizon. There is heightened activity in financial markets near the opening and closing of trading.[9] At a theoretical level, this can be explained in models such as Hong and Wang (2000). To capture this effect, we borrow from the volatility literature (e.g., Andersen, Dobrev, and Schaumburg, 2012; Hasbrouck, 1999) and assume that

$$\mu_{d,t} = C + Ae^{-a_1 t} + Be^{-a_2(1-t)}, \tag{4.4}$$

where $A = 0.75$, $B = 0.25$, $C = 0.89998744$ and $a_1 = a_2 = 10$. The choice of $C$ ensures that $\int_0^1 \mu_{d,u} \mathrm{d}u = 1$.

Prior to the calculation of the test statistic, we deflate the increment of the counting process using a nonparametric estimator of the intraday intensity curve with a direct adaption of Taylor and Xu (1997) from the volatility setting. We base the latter on 518 randomly selected Monte Carlo replica out of the 1,000 in total. This corresponds to the sample size in our empirical part and is therefore intended to ensure that the size of the measurement error corresponds to what we face in the application.

The above configurations imply that there is on average one event per second in the normal state, which is slightly higher compared to the daily transaction count observed in our empirical high-frequency data.

The intensity of $N_{s,t}^{\text{burst}}$ is defined as:

$$\beta_t = \frac{\sigma}{|\tau_{\text{ib}} - t|^\alpha}, \quad \text{for } t \in [\tau_{\text{ib}} - 0.05, \tau_{\text{ib}} + 0.05], \tag{4.5}$$

and $\beta_t = 0$ otherwise. $\sigma$ and $\alpha$ are constant. We position $\tau_{\text{ib}}$ at random in the interval $[0.05, 0.95]$ and make the duration of the intensity burst cover a 10-minute window. To get events of varying magnitude, we take $\alpha \in \{0.25, 0.50, 0.75\}$ and calibrate $\sigma$ such that $N_{s,t}^{\text{burst}}$ generates on average a fraction $c$ of the points produced by $N_{s,t}^{\text{normal}}$, where $c \in \{0.000, 0.025, 0.050, 0.100\}$.[10] This is called no, small, medium, or large intensity burst.

We compute $\phi_t^{\text{ib}}$ at $t = \tau_{\text{ib}}$, which is also the intensity burst time under the alternative. The bandwidth is $\ell_n \in \{60, 300, 600\}$ seconds for the calculation of $\widehat{\lambda}_t$ (corresponding to a 1, 5, and 10-minute block) and $K_n = 10\ell_n$ (corresponding to a 5, 25, and 50-minute block) for $\widehat{\mathrm{avar}}(\nabla\widehat{\lambda}_t)$.

---

[9]In global markets, where trading is spread out across the world (such as foreign exchange rates), there can be multiple intraday curves aligned with the start and end of trading in the various geographical regions (such as Asia, Europe, and the US).

[10]Solving for $\sigma$ 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 $\Delta$ is the duration of the intensity burst and $\beta_t = \frac{\sigma}{|\tau_{\text{ib}}-t|^\alpha}$, 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 intensity burst.

Figure 2: Example of a simulated intensity burst.
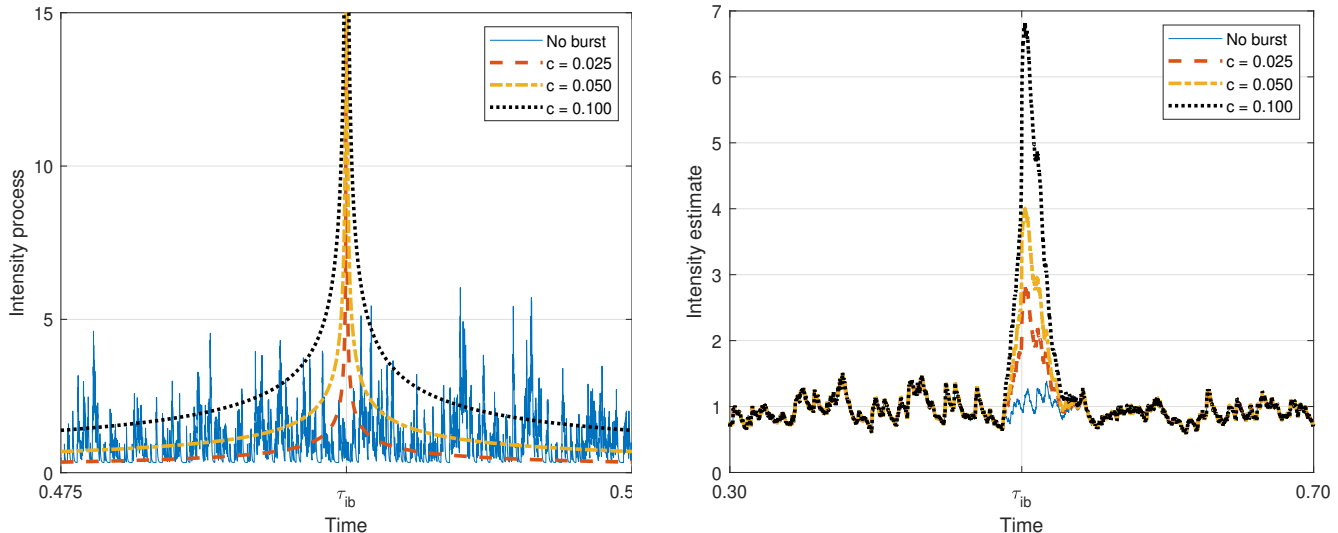


*Note.* This figure shows the true (in Panel A) and estimated (in Panel B) intensity of the simulated Hawkes process with no burst, and a small, medium and large intensity burst.

We experiment both with the indicator kernel $\mathcal{K}(x) = \mathbf{1}_{[-1,0]}(x)$, as implicit in (3.17), and the exponential kernel $\mathcal{K}(x) = \exp(x)$, for $x \leq 0$.

In Panel A of Figure 2, we illustrate the simulated intensity of the Hawkes process for $N_{s,t}^{\mathrm{normal}}$ (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 $N_{s,t}^{\mathrm{burst}}$ with $\alpha = 0.50$. In Panel B, we show the associated intensity estimate for $\ell_n = 300$. Across burst regimes, the maximal local intensity estimate is roughly 3, 4, and 7 times larger than the average rate of the Hawkes process. In comparison, the intensity burst displayed in Figure 1 is much larger. Hence, our simulated intensity bursts are conservative relative to many of those observed in empirical data.

Tables 1 – 3 collects the results of the Monte Carlo analysis by reporting rejection rates for the test statistic across 1,000 independent trials. The left-hand side is for the indicator kernel, while the right-hand side is for the exponential kernel. Throughout, the test statistic is evaluated against critical values from the standard normal distribution, i.e. the $p$th percentile $z_p = \Phi^{-1}(p)$ for $p \in \{0.950, 0.975, 0.990\}$.

Looking at Panel A in Tables 1 – 3, we observe that in the absence of an intensity burst the rejection rates are close to the nominal level across bandwidth choices. Stochastic intensity creates a minuscule size distortion, in particular for the shortest bandwidth and exponential kernel. However, the effect vanishes once we calculate the intensity estimator over a slightly longer window. What is also interesting to note is that the self-excitation of the Hawkes intensity is not a separate source of additional size distortion.

Turning to the alternative in Panels B – D, the rejection rates of the test are nearly perfect

for a large intensity burst, irrespective of the dynamic of $N_t^{\text{normal}}$. This is as expected, since the test statistic is calculated at the center of the burst process. In contrast, detection of small bursts remains challenging if they account for an insignificant fraction of the total transaction count—as measured by a smaller value of $c$—or are slow to materialize—as measured by a smaller value of $\alpha$. Moreover, there is an intuitive trade-off embedded in the selection of the bandwidth parameter. We observe that a longer bandwidth is preferable if the explosion rate of the intensity burst is small, and vice versa.

Overall, the test statistic catches the vast majority of medium and large intensity bursts, whereas smaller ones with slow rates of divergence may go unnoticed. This is not a cause of too much concern, since we are primarily interested in the behavior of large and rapid surges in trading activity.

Table 1: Rejection rate for the Poisson process.

| | Indicator kernel | | | | | | | | | Exponential kernel | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\ell_n = 60$ | | | $\ell_n = 300$ | | | $\ell_n = 600$ | | | $\ell_n = 60$ | | | $\ell_n = 300$ | | | $\ell_n = 600$ | | |
| | $z_{0.950}$ | $z_{0.990}$ | $z_{0.995}$ | $z_{0.950}$ | $z_{0.990}$ | $z_{0.995}$ | $z_{0.950}$ | $z_{0.990}$ | $z_{0.995}$ | $z_{0.950}$ | $z_{0.990}$ | $z_{0.995}$ | $z_{0.950}$ | $z_{0.990}$ | $z_{0.995}$ | $z_{0.950}$ | $z_{0.990}$ | $z_{0.995}$ |
| *Panel A: size (c = 0.000, no burst)* | | | | | | | | | | | | | | | | | | |
| | 6.0 | 1.2 | 0.6 | 4.8 | 1.1 | 0.7 | 5.1 | 0.9 | 0.4 | 5.5 | 1.5 | 0.9 | 5.1 | 0.8 | 0.8 | 5.8 | 1.1 | 0.4 |
| *Panel B: power (c = 0.025, small burst)* | | | | | | | | | | | | | | | | | | |
| $\alpha = 0.25$ | 59.2 | 32.0 | 24.3 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 83.2 | 58.6 | 48.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| $\alpha = 0.50$ | 100.0 | 100.0 | 99.9 | 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 |
| $\alpha = 0.75$ | 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 | 100.0 | 100.0 |
| *Panel C: power (c = 0.050, medium burst)* | | | | | | | | | | | | | | | | | | |
| $\alpha = 0.25$ | 83.1 | 54.9 | 41.9 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 97.9 | 83.9 | 75.3 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| $\alpha = 0.50$ | 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 | 100.0 | 100.0 |
| $\alpha = 0.75$ | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| *Panel D: power (c = 0.100, large burst)* | | | | | | | | | | | | | | | | | | |
| $\alpha = 0.25$ | 96.1 | 72.6 | 57.9 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 97.1 | 91.7 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| $\alpha = 0.50$ | 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 | 100.0 | 100.0 |
| $\alpha = 0.75$ | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |

*Note.* We simulate a point process as the sum of a Poisson process ("normal") and an intensity burst ("burst"), $N_t = N_t^{\mathrm{normal}} + N_t^{\mathrm{burst}}$. $\alpha$ measures the explosion rate of the latter. We test the hypothesis of no intensity burst. $c$ is the proportion of the total transaction count induced by burst process, so $c = 0.000$ corresponds to $\mathcal{H}_0$. The test is calculated at $t = \tau_{\mathrm{ib}}$, which is the intensity burst time under $\mathcal{H}_a$. $\ell_n$ is the bandwidth (in seconds) for the intensity estimator. The bandwidth for the observed local asymptotic variance is $K_n = 10\ell_n$. The table reports rejection rates of the intensity burst test statistic evaluated against the $(1 - \varsigma)$th percentile of the standard normal distribution, $z_{(1-\varsigma)}$, for $\varsigma = (0.050, 0.010, 0.005)$.

Table 2: Rejection rate for the Heston process.

| | Indicator kernel | | | | | | | | | Exponential kernel | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\ell_n = 60$ | | | $\ell_n = 300$ | | | $\ell_n = 600$ | | | $\ell_n = 60$ | | | $\ell_n = 300$ | | | $\ell_n = 600$ | | |
| | $z_{0.950}$ | $z_{0.990}$ | $z_{0.995}$ | $z_{0.950}$ | $z_{0.990}$ | $z_{0.995}$ | $z_{0.950}$ | $z_{0.990}$ | $z_{0.995}$ | $z_{0.950}$ | $z_{0.990}$ | $z_{0.995}$ | $z_{0.950}$ | $z_{0.990}$ | $z_{0.995}$ | $z_{0.950}$ | $z_{0.990}$ | $z_{0.995}$ |
| *Panel A: size (c = 0.000, no burst)* | | | | | | | | | | | | | | | | | | |
| | 7.2 | 3.4 | 2.4 | 5.7 | 2.6 | 1.1 | 4.9 | 1.5 | 1.0 | 8.2 | 3.8 | 3.0 | 7.1 | 2.5 | 1.7 | 6.0 | 1.9 | 1.2 |
| *Panel B: power (c = 0.025, small burst)* | | | | | | | | | | | | | | | | | | |
| $\alpha = 0.25$ | 18.1 | 8.5 | 6.4 | 61.7 | 39.0 | 30.2 | 37.9 | 15.0 | 11.1 | 25.4 | 13.6 | 10.6 | 73.3 | 47.5 | 40.3 | 60.6 | 34.1 | 27.2 |
| $\alpha = 0.50$ | 69.0 | 43.9 | 36.7 | 66.4 | 43.8 | 34.9 | 40.2 | 17.8 | 12.2 | 89.1 | 75.3 | 68.6 | 86.0 | 62.9 | 55.0 | 67.9 | 42.3 | 33.8 |
| $\alpha = 0.75$ | 99.3 | 98.1 | 97.0 | 78.7 | 55.7 | 48.4 | 51.0 | 26.3 | 18.9 | 100.0 | 100.0 | 100.0 | 97.4 | 86.9 | 83.7 | 83.3 | 64.2 | 55.3 |
| *Panel C: power (c = 0.050, medium burst)* | | | | | | | | | | | | | | | | | | |
| $\alpha = 0.25$ | 36.4 | 16.5 | 11.5 | 97.2 | 88.7 | 83.5 | 84.0 | 62.5 | 52.3 | 55.2 | 30.0 | 23.5 | 99.6 | 96.8 | 93.0 | 97.5 | 90.7 | 84.8 |
| $\alpha = 0.50$ | 97.5 | 90.8 | 87.2 | 98.3 | 92.7 | 88.2 | 85.8 | 67.2 | 58.7 | 99.8 | 99.4 | 99.1 | 99.9 | 99.6 | 98.5 | 98.9 | 94.9 | 92.1 |
| $\alpha = 0.75$ | 100.0 | 100.0 | 100.0 | 99.3 | 98.0 | 96.0 | 93.6 | 81.3 | 75.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 99.6 | 99.2 |
| *Panel D: power (c = 0.100, large burst)* | | | | | | | | | | | | | | | | | | |
| $\alpha = 0.25$ | 69.0 | 35.1 | 25.8 | 100.0 | 100.0 | 100.0 | 99.7 | 99.0 | 98.1 | 89.0 | 65.3 | 53.2 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| $\alpha = 0.50$ | 100.0 | 99.9 | 99.8 | 100.0 | 100.0 | 100.0 | 99.7 | 99.3 | 98.5 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| $\alpha = 0.75$ | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 99.7 | 99.6 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |

*Note.* We simulate a point process as the sum of a Heston process ("normal") and an intensity burst ("burst"), $N_t = N_t^{\mathrm{normal}} + N_t^{\mathrm{burst}}$. $\alpha$ measures the explosion rate of the latter. We test the hypothesis of no intensity burst. $c$ is the proportion of the total transaction count induced by burst process, so $c = 0.000$ corresponds to $\mathcal{H}_0$. The test is calculated at $t = \tau_{\mathrm{ib}}$, which is the intensity burst time under $\mathcal{H}_a$. $\ell_n$ is the bandwidth (in seconds) for the intensity estimator. The bandwidth for the observed local asymptotic variance is $K_n = 10\ell_n$. The table reports rejection rates of the intensity burst test statistic evaluated against the $(1 - \varsigma)$th percentile of the standard normal distribution, $z_{(1-\varsigma)}$, for $\varsigma = (0.050, 0.010, 0.005)$.

Table 3: Rejection rate for the Hawkes process.

| | Indicator kernel | | | | | | | | | Exponential kernel | | | | | | | | |
| | $\ell_n = 60$ | | | $\ell_n = 300$ | | | $\ell_n = 600$ | | | $\ell_n = 60$ | | | $\ell_n = 300$ | | | $\ell_n = 600$ | | |
| | $z_{0.950}$ | $z_{0.990}$ | $z_{0.995}$ | $z_{0.950}$ | $z_{0.990}$ | $z_{0.995}$ | $z_{0.950}$ | $z_{0.990}$ | $z_{0.995}$ | $z_{0.950}$ | $z_{0.990}$ | $z_{0.995}$ | $z_{0.950}$ | $z_{0.990}$ | $z_{0.995}$ | $z_{0.950}$ | $z_{0.990}$ | $z_{0.995}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Panel A: size ($c = 0.000$, no burst)* | | | | | | | | | | | | | | | | | | |
| | 4.6 | 1.2 | 1.0 | 5.2 | 1.6 | 1.0 | 5.6 | 1.1 | 0.8 | 5.0 | 1.5 | 0.9 | 5.5 | 1.6 | 0.7 | 5.5 | 1.2 | 0.8 |
| *Panel B: power ($c = 0.025$, small burst)* | | | | | | | | | | | | | | | | | | |
| $\alpha = 0.25$ | 49.1 | 25.2 | 18.2 | 100.0 | 100.0 | 99.7 | 99.3 | 94.5 | 93.2 | 74.5 | 47.7 | 38.0 | 100.0 | 100.0 | 100.0 | 100.0 | 99.7 | 99.6 |
| $\alpha = 0.50$ | 99.8 | 99.4 | 98.3 | 100.0 | 100.0 | 100.0 | 99.5 | 96.9 | 94.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 99.9 | 99.7 |
| $\alpha = 0.75$ | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 99.7 | 99.2 | 98.1 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| *Panel C: power ($c = 0.050$, medium burst)* | | | | | | | | | | | | | | | | | | |
| $\alpha = 0.25$ | 77.1 | 47.7 | 36.8 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 95.5 | 78.1 | 65.3 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| $\alpha = 0.50$ | 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 | 100.0 | 100.0 |
| $\alpha = 0.75$ | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| *Panel D: power ($c = 0.100$, large burst)* | | | | | | | | | | | | | | | | | | |
| $\alpha = 0.25$ | 94.7 | 67.3 | 54.4 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 95.9 | 89.5 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| $\alpha = 0.50$ | 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 | 100.0 | 100.0 |
| $\alpha = 0.75$ | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |

*Note.* We simulate a point process as the sum of a Hawkes process ("normal") and an intensity burst ("burst"), $N_t = N_t^{\text{normal}} + N_t^{\text{burst}}$. $\alpha$ measures the explosion rate of the latter. We test the hypothesis of no intensity burst. $c$ is the proportion of the total transaction count induced by burst process, so $c = 0.000$ corresponds to $\mathcal{H}_0$. The test is calculated at $t = \tau_{\text{ib}}$, which is the intensity burst time under $\mathcal{H}_a$. $\ell_n$ is the bandwidth (in seconds) for the intensity estimator. The bandwidth for the observed local asymptotic variance is $K_n = 10\ell_n$. The table reports rejection rates of the intensity burst test statistic evaluated against the $(1 - \varsigma)$th percentile of the standard normal distribution, $z_{(1-\varsigma)}$, for $\varsigma = (0.050, 0.010, 0.005)$.

Figure 3: Daily price and log-return in EUR/USD.
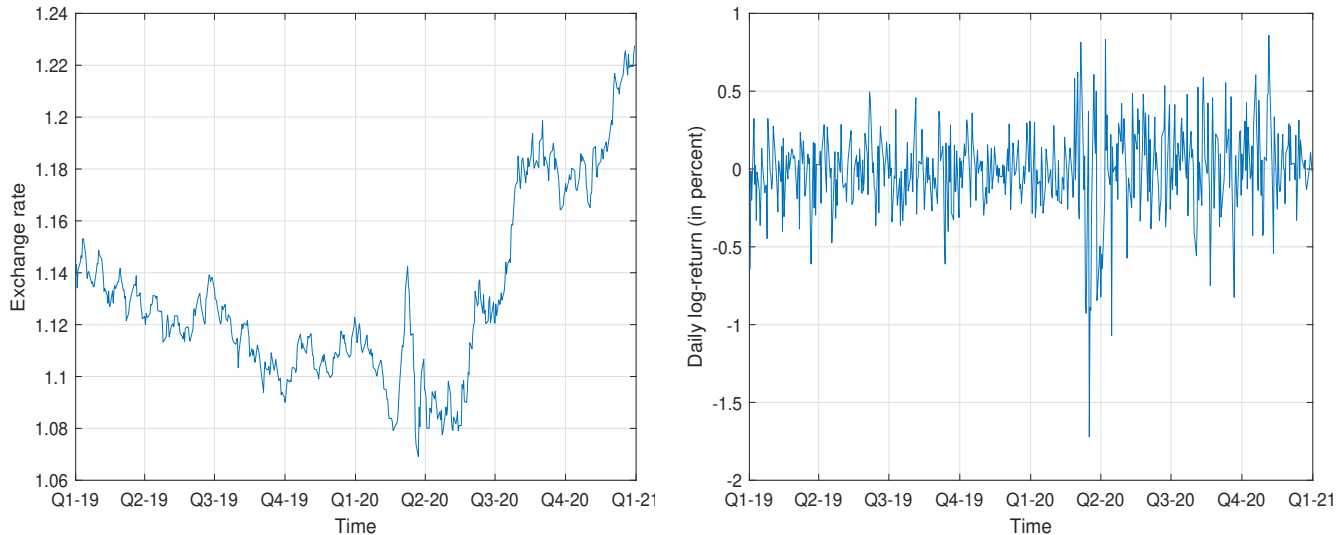


*Note.* In Panel A, we plot the EUR/USD spot exchange rate at noon CET for the sample period January 1, 2019 to December 31, 2020. In Panel B, we show the associated daily log-return in percent.

# 5 Empirical application

We examine high-frequency data obtained via Electronic Broking Services (EBS) on the Chicago Mercantile Exchange (CME). EBS is an electronic platform operating at the wholesale level and is the main interdealer venue in the foreign exchange market. Our dataset consists of all spot transactions and every tick-by-tick update to the first ten levels in the limit order book of the EUR/USD. The sample period January 1, 2019 to December 31, 2020. We exclude days with limited trading (e.g. weekends and public holidays) and retain $T = 518$ days. We also restrict attention to the most active trading hours from 8:00am to 4:00pm Central European Time (CET), covering the European and American session. In Panel A of Figure 3, we plot the daily midquote of the spot exchange rate at noon, while Panel B shows the associated log-return.

In Panel A of Figure 4, we show the total number of transactions per day in the retained sample. There are typically thousands of trades. To account for the expected increases in trading activity that are associated with the normal periodicity of this market or regular releases of pre-scheduled news announcements, we start by extracting a nonparametric estimate of the time-of-the-day mean transaction count. We calculate the latter as the average number of transactions in each one-second bucket, where the average is taken across days in the sample. The sequence of these estimates is then normalized to sum to one. The seasonality in the intraday transaction count is reported in Panel B of Figure 4.

We compute the spot intensity estimator in (3.3) each second from 8:00am to 4:00pm CET with a bandwidth of $\ell_n = 300$ and the exponential kernel. Each estimate is then divided by the appropriate

Figure 4: Trading activity in EUR/USD.
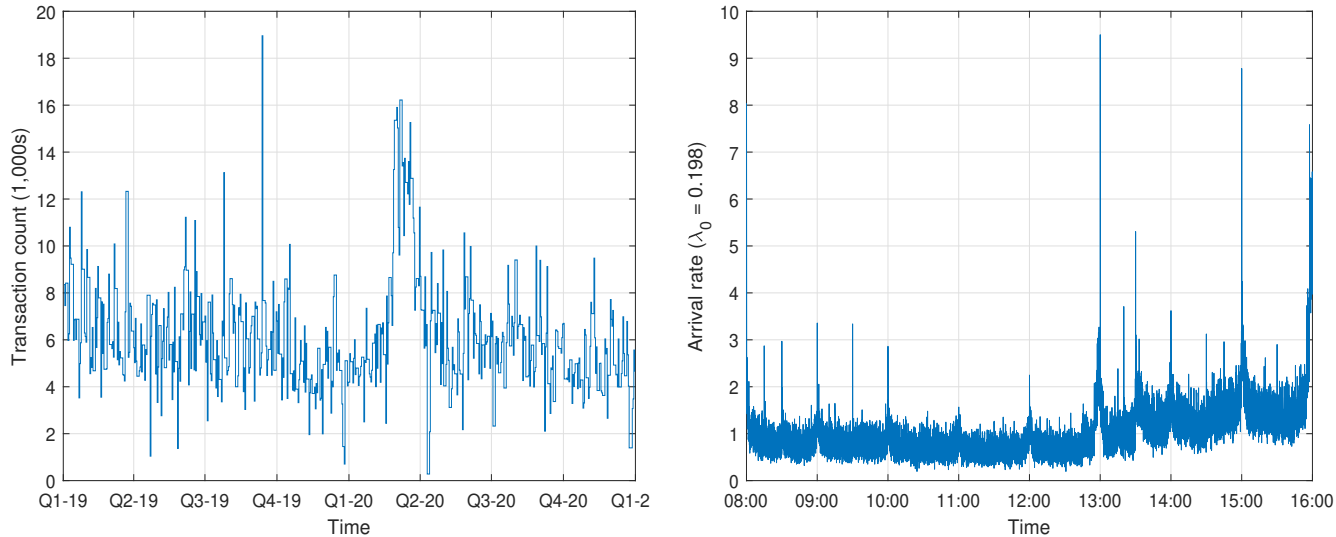
*Note.* In Panel A, we report the daily number of transactions (in 1,000s) on the EBS platform in the EUR/USD spot exchange rate. In Panel B, we show a nonparametric estimate of the time-of-the-day mean transaction intensity between 8:00am and 4:00pm CET, which is normalized to sum to one. $\lambda_0$ is an estimate of the unconditional arrival rate over the whole sample.

seasonality factor to remove the inherent diurnal variation. Next, to avoid the computational load of running the test statistic at such a refined grid, every day we select the 20 largest local maxima of the diurnal-corrected spot intensity estimator, allowing for at most one maximum in a 5-minute neighbourhood—equal to the bandwidth of the intensity estimator—to avoid double-counting. In Panel A of Figure 5, we show how this rule operates on July 2, 2019. The spot intensity is shown from 8:00am to 4:00pm. It is scaled by the average intensity on that day and log-transformed, as further motivated below.

In the subsequent analysis, we limit our attention to the points indicated with a square, which are the most probable intensity burst times.[11] We implement the test here with $K_n = 10\ell_n$ for the observed asymptotic local variance estimator, as in the simulation section. In doing so, only the crossed square is identified as an intensity burst. Overall, we monitor a total of 10,360 test statistics across the 518 days. The median value of the sequence is 1.24, which is more than a standard deviation away from the unconditional mean of zero of the asymptotic standard normal distribution. There are 913, 331, 119, and 36 observations with a value higher than 4, 5, 6, and 7, amounting to 8.81%, 3.20%, 1.15%, and 0.35%. Hence, intensity bursts—even more extreme ones—happen on a frequent basis.

In the remainder of the empirical investigation, we confine the analysis to a subsample of intensity bursts, where the test statistic is greater than 5. This value is large enough to be highly significant

---

[11]By Lemma 1, the unobserved intensity diverges at an intensity burst time. Thus, if an intensity burst occurs near a given point, the local intensity estimator should be at its highest here.

Figure 5: Candidate times for an intensity burst.
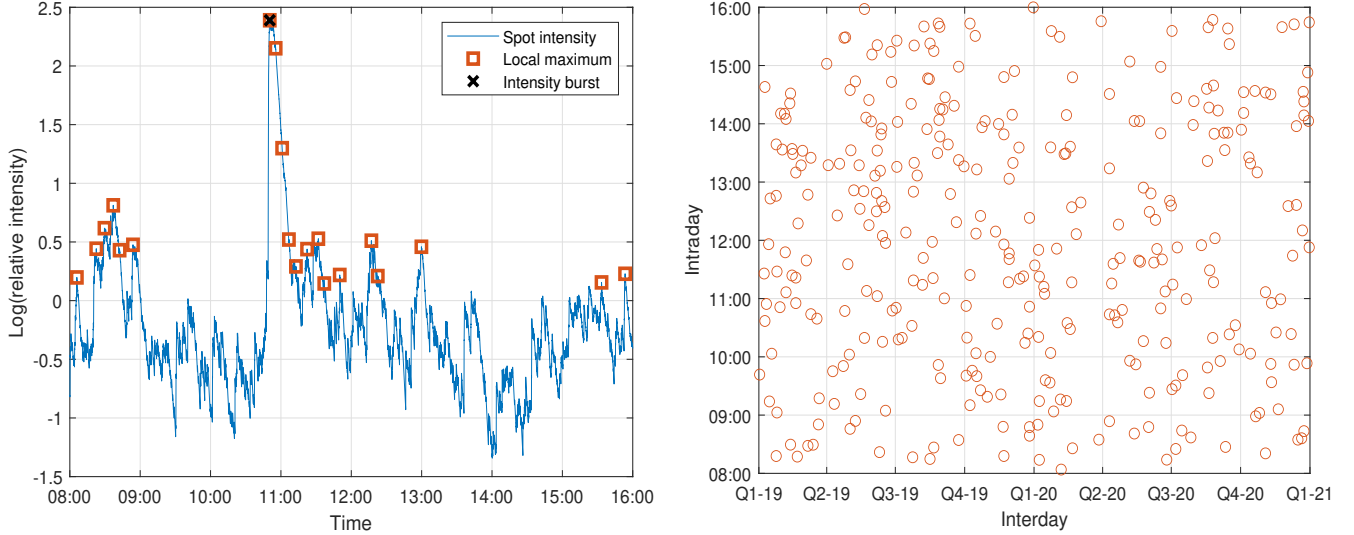
*Note.* In Panel A of this figure, we show the candidate times for an intensity burst on July 2, 2019 highlighted with a square. These are defined as the 20 largest local maxima of the estimated spot intensity within a 5-minute window (identical to the bandwidth of the estimator). The crossed square is identified as an intensity burst. In Panel B, we show how the intensity burst test statistics with a value greater than 5 are scattered at the intraday level over the whole sample.

and small enough to retain a sufficient number of observations.[12] The reduced sample of 331 events are highlighted in Panel B of Figure 5. There is no systematic placement in the points, which are more or less uniformly distributed at the intraday level.[13]

## 5.1 A burst or jump?

We first explore whether the above events correspond to genuine intensity bursts or are, in fact, more likely to be associated with false alarms caused by discrete changes in the intensity process,

---

[12]We face a theoretical concern in looking at the test statistic at many local maxima of the intensity estimator, since this can cause over-rejection under the null hypothesis of no intensity burst. We therefore do not evaluate it against a customary quantile from the standard normal distribution, but we select a much higher value following the multiple comparison literature to control the family-wise error rate at the $\varsigma = 0.01$ level of significance. On the one hand, we conduct 10,380 tests in total, so a Bonferroni correction recommends a critical value of $\Phi^{-1}(1-0.01/10,380) = 4.7610$. On the other hand, according to extreme value theory, the normalized maximum test statistic converges in law to a Gumbel distribution, i.e. $\frac{\max_{1 \le i \le m} \phi_{t_i}^{\mathrm{ib}} - a_m}{b_m} \xrightarrow{\mathfrak{D}} \xi$ as $m \to \infty$, where $F_\xi(x) = \mathbb{P}(\xi \le x) = \exp(-\exp(-x))$, $a_m = \sqrt{2\log(m)} - \frac{\log(\pi) + \log(\log(m))}{2\sqrt{2\log(m)}}$, and $b_m = (2\log(m))^{-1/2}$. Hence, if $\phi_{t_i}^{\mathrm{ib}}$ is computed at a one-second grid, as consistent with our implementation, and the daily maximum test statistic is extracted from this sequence, the critical value for the nonnormalized maximum test statistic is 5.1846. We base our inference on the average of these approaches by adopting a critical value of 5.

[13]We calculated the fraction of the total daily transaction count that occurs over a 10-minute interval around each such event. The sample average is 3.60% (median of 2.91%) with a 1st and 3rd quartile of 1.93% and 8.95%. This crudely aligns with the our choices of $c$ made in the simulation section.

Figure 6: Histogram of ratio statistic and explosion rate.
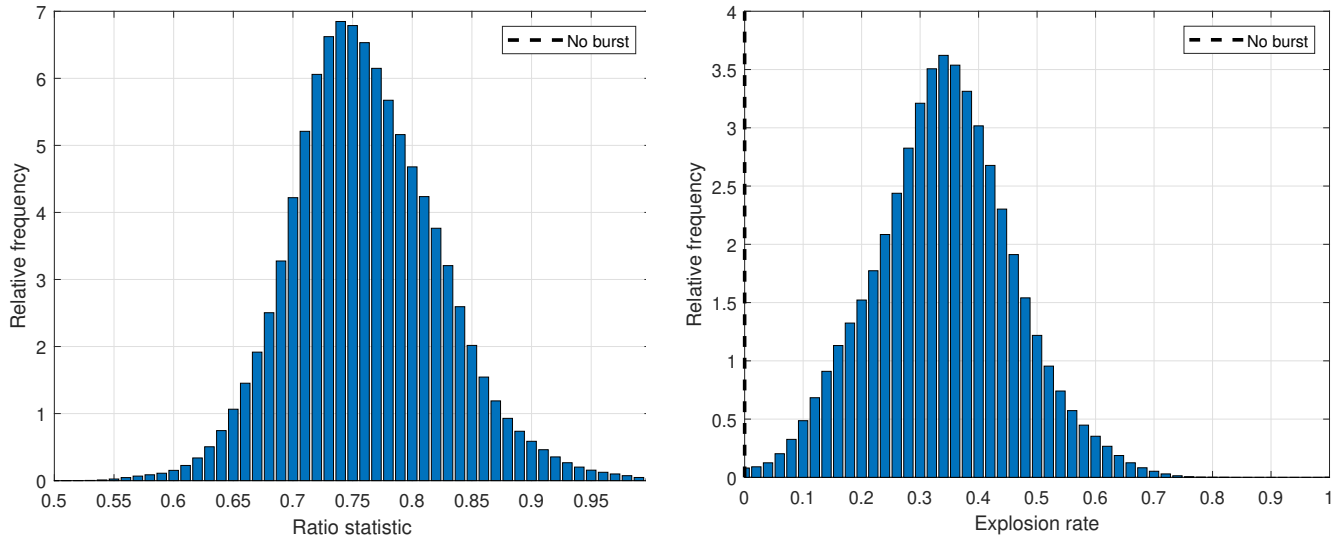


*Note.* In Panel A, we plot a histogram of the ratio statistic from Lemma 2, $\widetilde{\lambda}_\theta(k\delta_n)/\widetilde{\lambda}_\theta(\delta_n)$, with $k = 2$. It is centered around one in the absence of a burst, and strictly smaller than one otherwise. In Panel B, we report the distribution of the estimated explosion rate, $\hat{\alpha}$, based on the negative slope coefficient of an OLS regression through the origin of $\log\left(\widetilde{\lambda}_\theta(k\delta_n)\right)$ on $\log(k)$ for $k = 1, 2, 3, 5,$ and $10$.

as discussed in Section 3.3.

We calculate the two-sided intensity estimator from (3.29) with $k = 2$ and construct the ratio statistic from Lemma 2, where the baseline in the denominator corresponds to an implicit $k = 1$. It should be centered around one if the event coincided with a jump in intensity, whereas it should be strictly smaller than one—or $k^{-\alpha}$—for an intensity burst. In Panel A of Figure 6, we plot a histogram of the quotient. As evident, the empirical density function is concentrated around 0.75 with little probability mass near one. In total, only three of the ratio statistics are greater than or equal to one, where it is 1.05 on average. This strongly indicates that our test statistic is not driven by a jump component.

Conditional on an intensity burst, we can convert the ratio statistic to an estimate of the explosion rate, $\alpha$, as described in (3.35) and the subsequent comment. We regress $\log\left(\widetilde{\lambda}_\theta(k\delta_n)\right)$ through the origin against $\log(k)$ for $k = 1, 2, 3, 5,$ and $10$. The distribution of the negative slope estimates are reported in Panel B of Figure 6. As consistent with the previous analysis, they are strictly above the no burst lower bound of zero. The overall impression is that an intensity burst is often modest with an explosion rate in the middle of the permissible range.

## 5.2  State of the market near an intensity burst

We next take a closer look at how various state variables evolve in the vicinity of an intensity burst. In Panel A of Figure 7, we show the evolution of trading activity. We compute a proxy for the burst

Figure 7: Trading activity and return volatility near intensity burst.
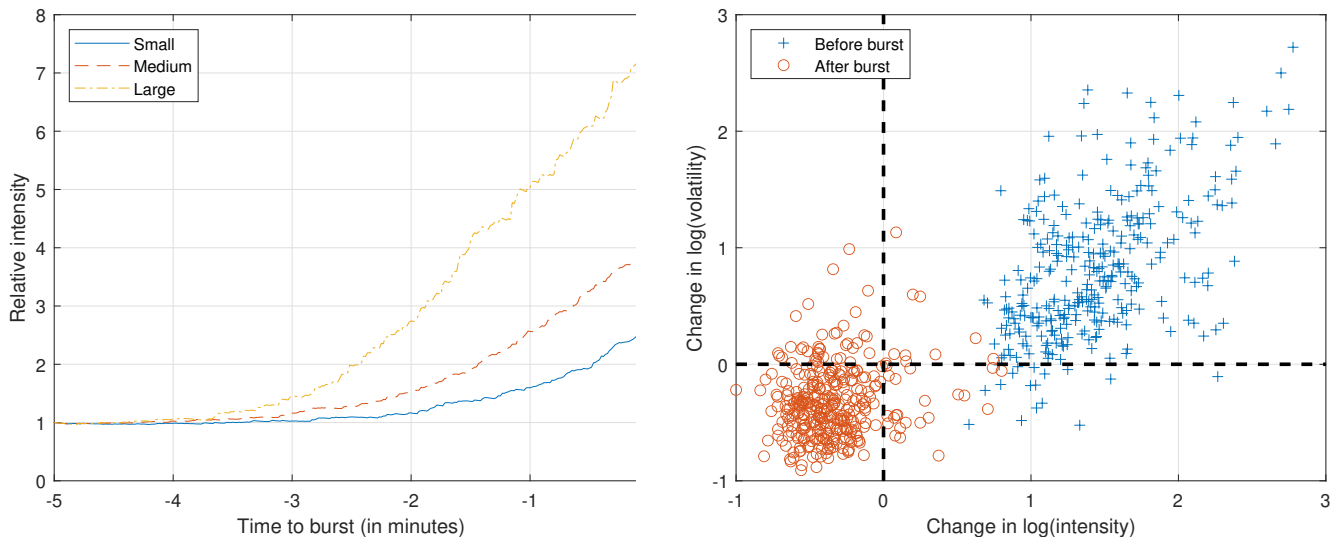
*Note.* In Panel A, we plot the evolution of the relative intensity estimator each second in a 5-minute window prior an intensity burst time. The sample has been split in four based on the magnitude of this measure and we then report the median intensity in the first (small), second–third (medium), and fourth (large) quartile. In Panel B, we plot the change in log-spot intensity against the change in log-spot volatility of the EUR/USD exchange rate process before and after the burst.

size as the relative change in the intensity estimator at each second in a 5-minute interval prior to the intensity burst time. We split the sample in four based on the magnitude of this measure and report the median intensity in the first (small), second–third (medium), and fourth (large) quartile. As expected, there is a steady rise in trading activity prior to a detected event. The curvature indicates a good fit with an intensity burst of the form (2.4), while the level is consistent with the simulation of this model illuminated in Figure 2.

In Panel B of the figure, we plot the change in log-spot intensity against the change in log-spot volatility of the EUR/USD exchange rate process. The latter is calculated with the price jump- and microstructure noise-robust estimator from Christensen, Oomen, and Renò (2022). To construct the increment, we measure the spot processes at the 5-minute mark before and after the intensity burst time, together with an estimate at the peak. As readily seen, there is a strong positive association between the change in intensity and change in volatility. This is consistent with the mixture of distribution hypothesis. The correlation is stronger before an intensity burst, around 0.55, which is on par with Stoltenberg, Mykland, and Zhang (2022). However, there is a significant drop in the level of the correlation, around 0.20, in the aftermath of an intensity burst.

How does liquidity change around an intensity burst? To understand this, we construct the realized illiquidity of Lacave, Ranaldo, and de Magistris (2023), which is a high-frequency version of the Amihud (2002) measure. In Panel A of Figure 8, we chart of the average evolution of this measure on a second-by-second basis. At each time point, we extract one-minute log-returns over the preceding 5-minute interval and calculate the statistic as the realized power variation (i.e. the

Figure 8: Realized illiquidity and transaction volume near intensity burst.
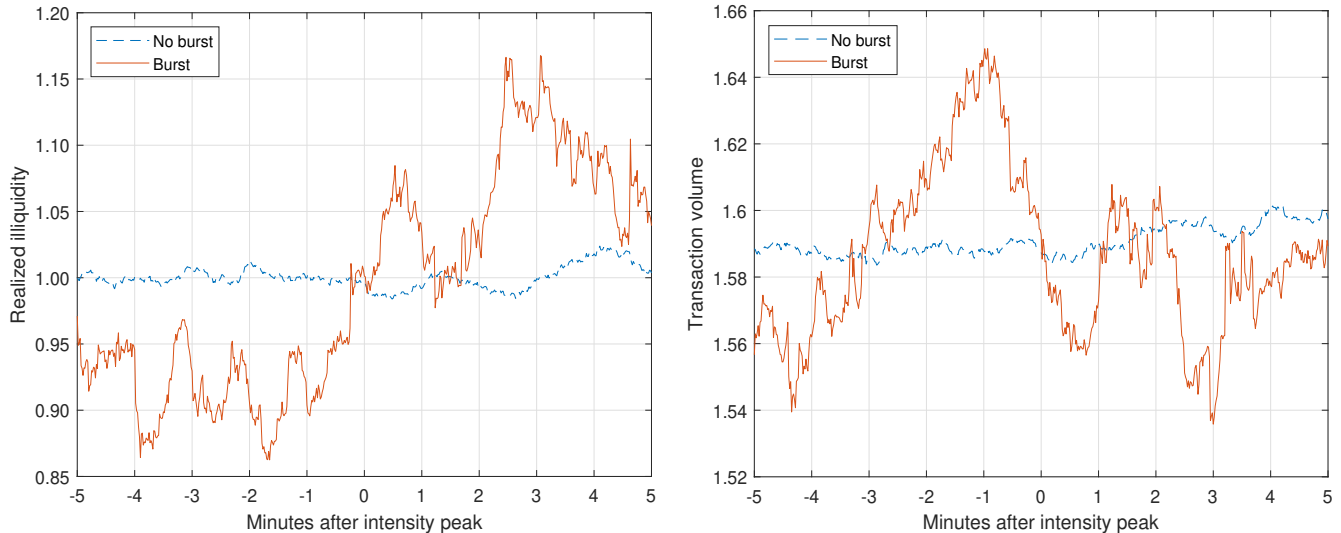


*Note.* In Panel A, we show the average realized illiquidity of Lacave, Ranaldo, and de Magistris (2023) over a 10-minute interval centered at the candidate times associated with the local maximum intensity estimates from Panel A in Figure 5. The sample has been divided based on whether the intensity burst statistic at these points is greater than 5 [burst] or not [no burst]. In Panel B, the average transaction volume (in million €) is reported. The latter is proportional to the denominator of the realized illiquidity measure.

sum of absolute log-returns) divided by the transaction volume over that window. We normalize it to one on average for ease of interpretation. We restrict attention to the candidate times from Panel A of Figure 5 and split the sample dependent on whether the intensity burst statistic is greater than 5 (burst) or not (no burst). While there is no material change in liquidity for an insignificant peak in the intensity estimator, during an intensity burst we observe a substantial rise in the illiquidity index both before and after the intensity burst time. In Panel B of the figure, we plot the cumulative transaction volume in million € (i.e. the denominator of realized illiquidity). There is no discernible shift in transaction volume, which hovers about in a very narrow interval confined to a few percent around its average level. This suggests that the upsurge in the realized illiquidity is mainly induced by the increase in volatility (i.e. the numerator of the statistic).

## 5.3   What is the relationship with a drift burst?

In Figure 9, we follow the recipe of May 14, 2019 displayed in Figure 1 and report the price and transaction data in the EUR/USD from 10:00am to 10:30am on July 10, 2019. As in the previous example, this day contains an intensity burst, which is started by sell-side activity around 10:15am. However, there is only a minuscule downtick in the exchange rate, which does not get caught as a drift burst. So here, the intensity burst is an isolated event.

Why is the market sometimes able to withstand the abnormal trading activity observed during an intensity burst with no discernible movement in the price, while at other times it goes through

Figure 9: Intensity burst in the EUR/USD without a drift burst.
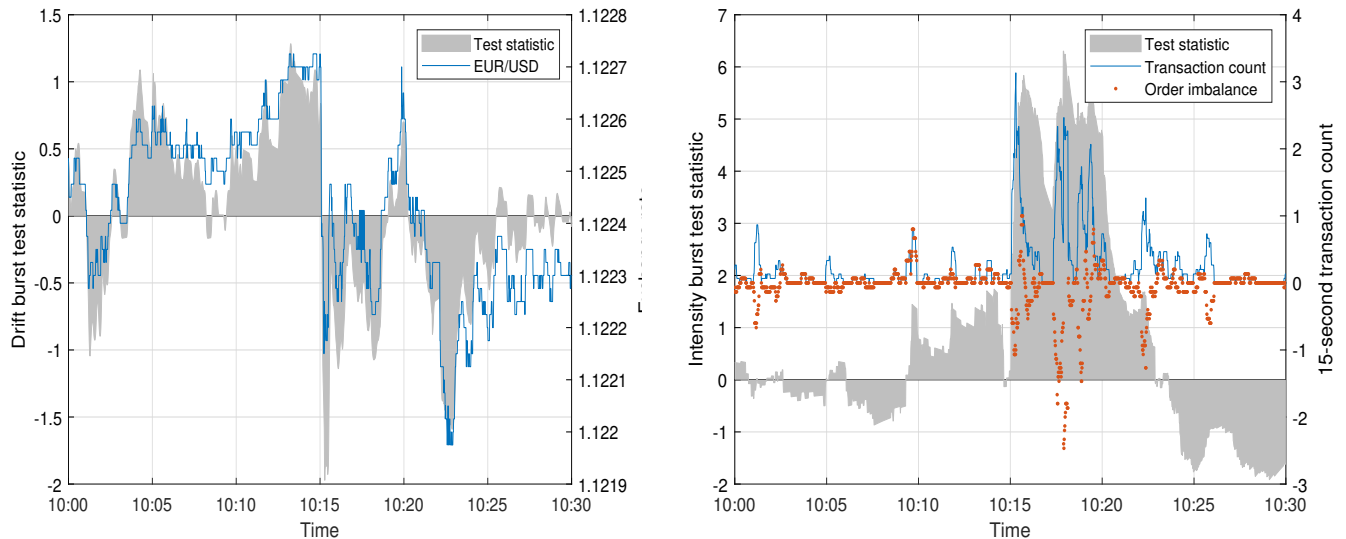


*Note.* This figure shows the EUR/USD spot exchange rate on July 10, 2019, where we detect an intensity burst but no a drift burst. In the left panel, we plot the exchange rate from 10:00am to 10:30am CET along with the drift burst test statistic proposed in Christensen, Oomen, and Renò (2022). In the right panel, we plot a nonparametric estimator of the time-varying trading intensity (as proxied by the number of transactions in a 15-second window) along with the intensity burst test statistic and a measure of order imbalance.

a severe correction? It is possible that a piece of news (or public signal) does not alter the average opinion of investors, even if their interpretation of the information differ. This is related to the differences-of-opinion literature (e.g., Bollerslev, Li, and Xue, 2018; Kandel and Pearson, 1995). Roughly speaking, in these models price change is related to the average opinion (or 1st moment), while trading volume is related to the dispersion of opinion (or 2nd moment). So there can be a lot of trading without a big impact on price, because mostly disagreeing investors trade with each other. However, absolute price change and trading volume should still be positively correlated (Harris and Raviv, 1993).[14]

To explore this question in more detail, we extract the drift burst test statistic at a one-second frequency in a 10-minute window centered at the intensity burst time.[15] The attractiveness of the drift burst test statistic is that it is detects abnormal price changes on a volatility-adjusted basis, since the test statistic is calculated as the average log-price increment divided by spot volatility over an interval.

We employ a separate linear regression with the minimum (most negative) and maximum (most positive) value of the drift burst test statistic as dependent variable. As covariates, we include the

---

[14]A small trade can cause a relatively big change in price if the market impact is large, such as in the presence of asymmetric information (e.g. Glosten and Milgrom, 1985; Kyle, 1985).

[15]We do not pretend that there is a cause and effect from an intensity burst to a drift burst. In practice, these events are the result of a highly complicated and interrelated process. The aim here is merely to determine correlation with suspected drivers of the drift burst test statistic during an intensity burst.

Figure 10: Properties of relative trading intensity.
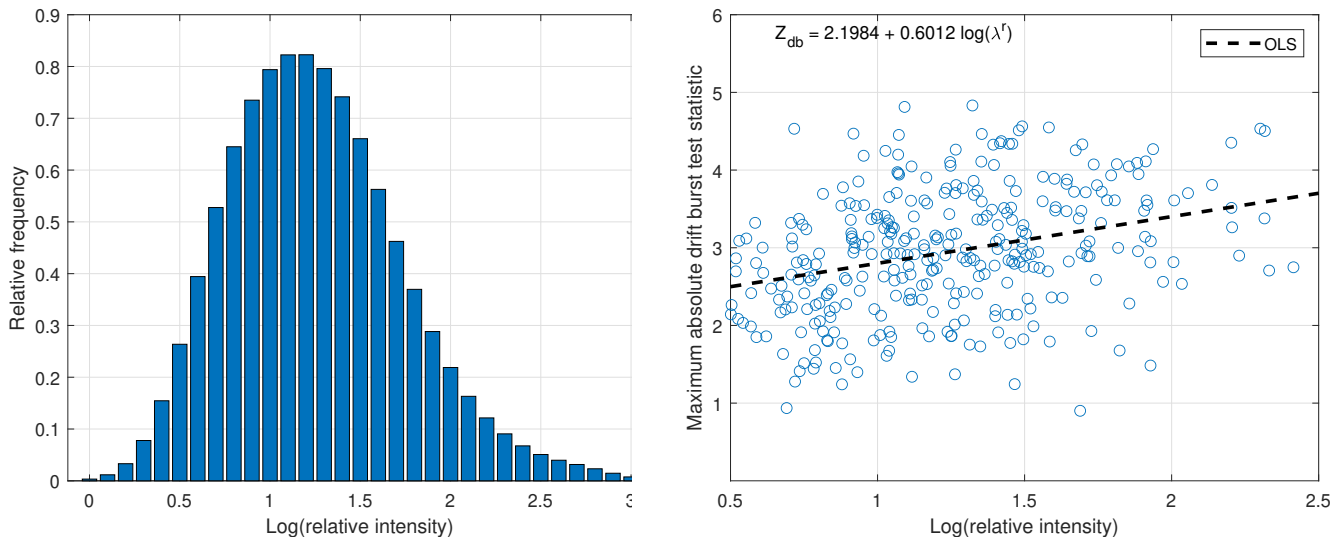


*Note.* In Panel A, we plot a histogram of the logarithm of relative trading intensity at the peak of an intensity burst. In Panel B, we show how the log-relative intensity is related to the maximum absolute value of the drift burst statistic in a 10-minute interval around the intensity burst time. A least squares fit is superimposed as a reference. The slope coefficient is significant with a *P*-value less than a percent. The sample is conditional on an intensity burst statistic being greater than 5. It contains a total of 331 events.

level of trading activity that should, everything else equal, inflate the magnitude of the drift burst test statistic. The rationale is, as stated above, that it requires trades (or quote updates) to move the price. We therefore add the intensity estimated at the peak of the intensity burst as the first predictor. However, because there is a great variety in the absolute level of trading over time, we scale the diurnal-corrected spot estimator by the average level of intensity for the entire day to construct a relative trading activity, which is more homogenous across events. As shown in Panel A of Figure 10, a log-transformation renders this variable close to normally distributed, apart from a slightly elevated right-tail. In Panel B of the figure, we plot the logarithmic measure against the maximum absolute value of the drift burst statistic over the calculation window, which indeed shows a positive relationship.

Our transaction data also contains an indicator variable of whether a deal was buyer- or seller-initiated. One-sided order flow is a good proxy for price pressure—that can be caused by either a change in the average opinion of investors or asymmetric information—so we include the order imbalance in the 5-minute interval before the intensity burst time as our second predictor. We calculate this as the number of buyer trades minus the number of seller trades. A positive value is therefore indicative of buy-side activity, and vice versa. We scale the difference by the spot intensity to again get a relative measure.

As our third predictor, we construct an intuitive concept of market depth derived from the information in the limit order book. We compute the elasticity of the price with respect to quantity following the procedure described in Næs and Skjeltorp (2006). This is referred to as the slope of

the order book.[16] It accounts for the resilience of the market by indicating whether it can absorb relatively large amounts of trading volume with minimal price impact. In particular, a high price elasticity suggests that the market is in a very fragile state, such that a sudden burst in trading activity can induce pronounced price volatility. The calculation is as follows:

$$\mathcal{S}_{\mathrm{ob}} = \frac{1}{L} \left( \frac{|P_1/M - 1|}{v_1} + \sum_{l=1}^{L} \frac{|P_{l+1}/P_l - 1|}{v_{l+1}/v_l - 1} \right),\tag{5.1}$$

where $L$ is the number of levels in the limit order book (10 here), $M$ is the midquote of the best bid and offer, $P_l$ and $v_l$ are the price and natural logarithm of accumulated volume at tick level $l$, for $l = 1, \ldots, L$. We calculate the slope separatively for the bid and ask side at a one-second frequency and include the moving average of the sequence over the 5-minute interval preceding the intensity burst in the regression.[17]

We estimate the model:

$$Z_{\mathrm{db}}^{\mp} = a + b_1 \log(\lambda^r) + b_2 \mathcal{O} + b_3 \mathcal{S}_{\mathrm{ob}}^{\mp} + b_4 \sqrt{RV} + b_5 \mathcal{P}_0 + \epsilon,\tag{5.2}$$

where $Z_{\mathrm{db}}^{-}$ ($Z_{\mathrm{db}}^{+}$) is the minimum (maximum) drift burst test statistic, $\lambda^r$ is the relative intensity estimate, $\mathcal{O}$ is the relative order imbalance, and $\mathcal{S}_{\mathrm{ob}}^{-}$ ($\mathcal{S}_{\mathrm{ob}}^{+}$) is the slope of the order book on the bid (ask) side.

There are a few controls in (5.2). $\sqrt{RV}$ is the square-root of 5-minute realized variance computed between three- to one-hour before the intensity burst, while $\mathcal{P}_0$ is the percent of daily 5-minute log-returns that are zero. The latter is a proxy of staleness in the aggregate trading environment, see Bandi, Kolokolov, Pirino, and Renò (2020). We first run the regression without these two additional variables and then add them as a robustness check.[18]

Table 4 presents the outcome of the regression estimated for the minimum (maximum) drift burst statistic in the left-hand (right-hand) side using the bid (ask) slope as price elasticity. We only comment on the results for the minimum drift burst test statistic, while summarizing the main differences to the other regression below.

The results are compelling. In particular, the sign of the slope coefficient estimates are in line with our economic intuition. The logarithm of relative intensity makes the minimum drift burst test statistic more negative, and the effect is significant. This is perhaps a bit surprising, since we are conditioning on an intensity burst. What also ought to matter is the ability of the market to handle the surge in trading activity, as proxied by order imbalance and price elasticity. The coefficient on order flow is positive and highly significant. This implies that sell-side activity (a negative value of

---

[16]Næs and Skjeltorp (2006) look at the elasticity of quantity with respect to price. We employ the reciprocal of their formula for the slope of the order book by reversing the role of price and volume.

[17]The one-sided slope calculation is a bit sensitive to the flip-flopping of large orders up and down in tick level, when the market bounces around at the inside spread. The slight averaging alleviates this problem.

[18]We also estimated a regression, where the realized illiquidity was included as a covariate. However, it was highly insignificant, so we removed it again. This is rather intuitive, since the measure was found to be driven mostly by volatility and the drift burst test statistic controls for this effect.

Table 4: Regression with drift burst test statistic.

| | | | | |
|---|---|---|---|---|
| Constant | $-1.145$ <br> $(-4.591)$ | $-1.318$ <br> $(-3.936)$ | $0.525$ <br> $(1.618)$ | $0.722$ <br> $(1.885)$ |
| $\log(\lambda_r)$ | $-0.430$ <br> $(-2.512)$ | $-0.456$ <br> $(-2.657)$ | $0.318$ <br> $(1.942)$ | $0.330$ <br> $(2.013)$ |
| $\mathcal{O}$ | $0.594$ <br> $(3.121)$ | $0.610$ <br> $(3.217)$ | $0.476$ <br> $(2.587)$ | $0.423$ <br> $(2.289)$ |
| $\mathcal{S}_{\text{ob}}$ | $-0.713$ <br> $(-2.495)$ | $-1.062$ <br> $(-3.273)$ | $2.281$ <br> $(4.036)$ | $2.601$ <br> $(4.459)$ |
| $\sqrt{RV}$ | | $0.689$ <br> $(0.431)$ | | $-1.069$ <br> $(-0.701)$ |
| $\mathcal{P}_0$ | | $6.467$ <br> $(2.214)$ | | $-5.260$ <br> $(-2.086)$ |
| $R^2$ | $0.082$ | $0.098$ | $0.109$ | $0.123$ |

*Note.* We estimate the linear regression: $Z_{\text{db}}^{\mp} = a + b_1 \log(\lambda^r) + b_2 \mathcal{O} + b_3 \mathcal{S}_{\text{ob}}^{\mp} + b_4 \sqrt{RV} + b_5 \mathcal{P}_0 + \varepsilon$, where $Z_{\text{db}}^{-}$ ($Z_{\text{db}}^{+}$) is the minimum (maximum) drift burst test statistic in a 10-minute window centered at the intensity burst time, $\lambda^r$ is the relative intensity, $\mathcal{O}$ is the relative order imbalance, $\mathcal{S}_{\text{ob}}^{-}$ ($\mathcal{S}_{\text{ob}}^{+}$) is the bid slope (ask slope) of the order book, $\sqrt{RV}$ is the 5-minute realized variance between three-to one-hour before the intensity burst, while $\mathcal{P}_0$ is the percent of daily 5-minute log-returns that are zero. The table reports parameter estimates with the $t$-statistic shown in parenthesis underneath. The sample size is 331. $R^2$ is the coefficient of determination.

this variable) moves the minimum drift burst statistic farther away from zero. In addition, a large price elasticity, as measured by the bid slope of the order book, reinforces this effect and pushes the drift burst statistic even further into negative territory.[19]

The maximum regression shows identical results, apart from a few expected sign changes in appropriate places, i.e. the intercept, and the coefficient in front of log-relative intensity, the order book slope, square-root realized variance, and the fraction of zeros in the log-return series. The main highlight is that the $t$-statistic on the log-relative intensity and order imbalance are a bit more compressed, although both are still at the margin of being significant. This can possibly be reconciled by a costly market presence, as developed in Huang and Wang (2009). In their framework, if market participation has a cost, the need of financial intermediaries to rebalance their equilibrium portfolios in response to news creates an endogenous order imbalance, which always leads to excessive selling. This effect exacerbates the demand for cash during a sell-off, while rallies remove a smaller portion of volume resting on the ask side. Hence, trading activity and order imbalances put less strain on the system during an upsurge. On the other hand, since EUR/USD is an exchange rate, it is not evident how to define a sell-off.

The additional controls do not alter the coefficients estimates of the above predictors much. The realized volatility reduces the magnitude of the drift burst test statistic, even if the effect is insignificant. Moreover, zero returns further attenuates it. The explanation is probably that in a slow market with limited trading, a large burst in relative trading activity may still be economically

---

[19]We also incorporated an interaction effect between order imbalance and price elasticity. However, it was insignificant both in the restricted and full version of the model. We therefore omitted it again.

small and incapable of shifting the price much.

In summary, our analysis shows it is asymmetric trading combined with a vulnerable order book that causes the absolute drift burst test statistic to inflate during an intensity burst, thus increasing the likelihood of observing a drift burst.

# 6   Conclusion

In this paper, we propose a model with a locally 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. This is the definition of an intensity burst. We show the model is capable of generating rather extreme clusters of observations over small time intervals that are far more concentrated compared to what a standard doubly stochastic Poisson or self-exciting Hawkes processes with locally bounded intensity can produce.

We develop an inference strategy for detecting an intensity burst. The theory relies on a heavy traffic assumption, which permits to consistently estimate and draw inference about the properties of a point process over a finite time interval. It resembles the standard in-fill condition for asymptotic theory of realized variance. In absence of an intensity burst (null hypothesis), the asymptotic distribution of our test statistic is standard normal, but the rate of convergence depends on a nuisance parameter, the speed with which the process accumulates points under the heavy traffic condition. Hence, the feasible implementation of our test statistic is based on an automatic inference procedure, where 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, so the power of the testing procedure converges to one.

A simulation study demonstrates that the test statistic has good finite sample properties for a variety of point processes. It controls size and has high power under the alternative. With our numerical experiments, we also show that with self-exciting exponential Hawkes (1971) process, a more general model than the theoretical framework allows, the test statistic does not produce false positives. Furthermore, we demonstrate robustness to diurnal variation in the intensity process based on a deflated intensity estimator.

At last, we implement the test statistic on an extensive set of EUR/USD foreign exchange rate high-frequency data extracted from the electronic broking services (EBS) platform, where it captures abnormal surges in trading activity. We detect a nontrivial amount of intensity bursts in these data and describe their basic properties. We examine how intensity is related to other fundamental indicators, such as volatility and liquidity. We also show that the more excessive the trading activity observed during an intensity burst is, the higher is the chance to observe a drift burst. This effect is stronger with an unbalanced order flow and when the price elasticity of the limit order book is high. We leave a further investigation of these findings to future research.

# 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 the processes $a_t$, $\nu_t$, and $\delta_t(x)$ as $(t, x)$ vary within $[0, T] \times \mathbb{R}$. Moreover, we can assume that $|\delta_t(x)| \leq \bar{\Gamma}(x)$, where $\bar{\Gamma}$ is a bounding function such that $\int_{\mathbb{R}} \bar{\Gamma}(x)^2 F(\mathrm{d}x) < \infty$ and, for all $\kappa \in (0, 1)$, $\int_{\{x : \bar{\Gamma}(x) \leq \kappa\}} \bar{\Gamma}(x) F(\mathrm{d}x) < \infty$.

We make extensive use of a universal positive constant, $C$, whose value changes from equation-to-equation and line-to-line without notification. As a shorthand notation, we write the intensity burst time as $\tau_{\mathrm{ib}} \equiv \tau$.

To establish stable convergence of the local intensity estimator in (3.3), we need an auxiliary result, which is a reproduction of Alvarez, Panloup, Pontier, and Savy (2012, Lemma 8).

**Lemma 3.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a probability space. For each $n \geq 1$, suppose $\zeta_2^n, \zeta_3^n, \ldots, \zeta_{k_n}^n$ are martingale increments with respect to the sub-$\sigma$-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 $\sigma$-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 $\eta$ 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 $\eta$.*

## A.1   Proof of Lemma 1

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

$$\widehat{\lambda}_t - \lambda_{t-} = o_p(1), \tag{A.1}$$

which shows the first statement.

Next, under $\mathcal{H}_1$, 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 $n\mu_t$ and $n\beta_t$.

Therefore, at time $\tau$,

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

It follows from (A.1) 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 $n \geq N_\epsilon : \mathbb{P}(\delta_n^\alpha \widehat{\beta}_\tau > \Delta_\epsilon) < \epsilon$.

By Markov's inequality:

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

Next, conditioning on $\sigma$ and employing the law of iterated expectations:

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

$$\leq \frac{C}{\delta_n}\int_{\tau-\delta_n}^\tau |\tau - u|^{-\alpha} du = \frac{C}{1-\alpha}\delta_n^{-\alpha},$$

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

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

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

## 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 3.

To this end, we define the sequence $\rho_n$ 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 $\delta_n$ can be expressed as $\delta_n = \ell_n \Delta_n$, where $\Delta_n = n^{-1}$ and $\ell_n$ 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 $\Delta_n$, 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

35

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

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

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:

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

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

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 $A_t = \int_0^t a_s^*\mathrm{d}s + \int_0^t\int_{\mathbb{R}}\delta_s(x)\left(\mathfrak{p}(\mathrm{d}s,\mathrm{d}x) - \mathfrak{q}(\mathrm{d}s,\mathrm{d}x)\right)$, where $a_s^* = a_s + \int_{|\delta_s(x)|>1}\delta_s(x)F(\mathrm{d}x)$, and $M_t = \int_0^t\nu_s\mathrm{d}W_s$.

36

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

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

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

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

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

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

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

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

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

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

As a result,

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

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

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

and, as $\delta_n \to 0$,

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

37

From the definition of $\rho_n$:

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

Therefore,

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

Next,

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

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

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

which further means that

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

On the other hand, multiplying and dividing by $\delta_n$,

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

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,

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

38

which implies that

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

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

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

By the Cauchy-Schwarz and Chebyshev's inequalities,

$$
\mathbb{E}\left[\left(\zeta_i^n(3) + \zeta_i^n(5)\right)^2 \mathbf{1}_{\left\{(\zeta_i^n(3)+\zeta_i^n(5))^2\ge\epsilon\right\}} \mid \mathcal{F}_{(i-1)\Delta_n}\right]
$$

$$
\le \sqrt{\mathbb{E}\left[\left(\zeta_i^n(3) + \zeta_i^n(5)\right)^4 \mid \mathcal{F}_{(i-1)\Delta_n}\right] \mathbb{E}\left[\mathbf{1}_{\left\{(\zeta_i^n(3)+\zeta_i^n(5))^2\ge\epsilon\right\}} \mid \mathcal{F}_{(i-1)\Delta_n}\right]}
$$

$$
= \sqrt{\mathbb{E}\left[\left(\zeta_i^n(3) + \zeta_i^n(5)\right)^4 \mid \mathcal{F}_{(i-1)\Delta_n}\right] \mathbb{P}\left(\left(\zeta_i^n(3) + \zeta_i^n(5)\right)^2 \ge \epsilon \mid \mathcal{F}_{(i-1)\Delta_n}\right)}
$$

$$
\le \frac{1}{\epsilon}\mathbb{E}\left[\left(\zeta_i^n(3) + \zeta_i^n(5)\right)^4 \mid \mathcal{F}_{(i-1)\Delta_n}\right]
$$

$$
\le \frac{8}{\epsilon}\left(\mathbb{E}\left[\left(\zeta_i^n(3)\right)^4 \mid \mathcal{F}_{(i-1)\Delta_n}\right] + \mathbb{E}\left[\left(\zeta_i^n(5)\right)^4 \mid \mathcal{F}_{(i-1)\Delta_n}\right]\right),
$$

where the last line follows from the inequality $|x+y|^p \le 2^{p-1}(|x|^p + |y|^p)$, for all real $x$ and $y$ and any $p\ge 1$.

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

$$
\sum_{i\in\mathcal{D}_{t-}^n} \mathbb{E}\left[\left(\zeta_i^n(3)\right)^4 \mid \mathcal{F}_{(i-1)\Delta_n}\right] = \sum_{i\in\mathcal{D}_{t-}^n} \frac{\rho_n^2}{\ell_n^4}\mathbb{E}\left[\Delta_i\widetilde{N}^n - \mu_{(i-1)\Delta_n}^4 \mid \mathcal{F}_{(i-1)\Delta_n}\right]
$$

$$
= \sum_{i\in\mathcal{D}_{t-}^n} \frac{\rho_n^2}{\ell_n^4}\mu_{(i-1)\Delta_n}(1+3\mu_{(i-1)\Delta_n})
$$

$$
\le C\frac{\rho_n^2}{\ell_n^3} \to 0,
$$

for both choices of $\rho_n$.

On the other hand, using the boundedness of $\nu$,

$$
\sum_{i\in\mathcal{D}_{t-}^n} \mathbb{E}\left[\left(\zeta_i^n(4)\right)^4 \mid \mathcal{F}_{(i-1)\Delta_n}\right] = \sum_{i\in\mathcal{D}_{t-}^n} \frac{\rho_n^2}{\ell_n^4}(tn-\ell_n-i)^4\, \mathbb{E}\left[\left(M_{i\Delta_n} - M_{(i-1)\Delta_n}\right)^4 \mid \mathcal{F}_{(i-1)\Delta_n}\right]
$$

$$
= 3\frac{\rho_n^2}{\ell_n^4}\sum_{i\in\mathcal{D}_{t-}^n}(tn-\ell_n-i)^4\, \mathbb{E}\left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n}\nu_s^2 ds\right)^2 \mid \mathcal{F}_{(i-1)\Delta_n}\right]
$$

39

$$\leq C \frac{\rho_n^2 \Delta_n^2}{\ell_n^4} \sum_{i \in \mathcal{D}_{t-}^n} (tn - \ell_n - i)^4 = C \frac{\rho_n^2 \Delta_n^2}{\ell_n^4} \frac{6\ell_n^5 + 15\ell_n^4 + 10\ell_n^3 - \ell_n}{30} \to 0,$$

again for both choices of $\rho_n$. Hence, Lindeberg's condition holds.

By Lemma 3, we therefore conclude that

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

where $\mathcal{Z} \sim N(0,1)$ 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

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

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

$$\mathbb{E}\left[ |\Lambda_1^n(t)|^2 \right] = \frac{\rho_n}{\ell_n^2} \sum_{i \in \mathcal{D}_{t-}^n} \mathbb{E}\left[ \left( \Delta_i N^n - \Delta_i \widetilde{N}^n - \mathbb{E}\left[ \Delta_i N^n - \Delta_i \widetilde{N}^n \mid \mathcal{F}_{(i-1)\Delta_n} \right] \right)^2 \right]$$

$$= \frac{\rho_n}{\ell_n^2} \sum_{i \in \mathcal{D}_{t-}^n} \mathbb{E}\left[ n \int_{(i-1)\Delta_n}^{i\Delta_n} \mu_s - \mu_{(i-1)\Delta_n} \mathrm{d}s \right]$$

$$\leq \frac{\rho_n n}{\ell_n^2} \sum_{i \in \mathcal{D}_{t-}^n} \int_{(i-1)\Delta_n}^{i\Delta_n} \mathbb{E}\left[ |\mu_s - \mu_{(i-1)\Delta_n}| \right] \mathrm{d}s,$$

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

$$\mathbb{E}\left[ |\mu_s - \mu_{(i-1)\Delta_n}| \right] \leq C \sqrt{\Delta_n}. \tag{A.2}$$

Thus, for any definition of $\rho_n$,

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

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 (A.2),

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

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

Now, $\Lambda_4^n(t)$ can be expressed as follows:

$$\Lambda_4^n(t) = \sum_{i \in \mathcal{D}_{t-}^n} (tn - \ell_n - i)\big(\zeta_i^n(4,1) + \zeta_i^n(4,2)\big),$$

where

$$\zeta_i^n(4,1) = \frac{\sqrt{\rho_n}}{\ell_n} \int_{(i-1)\Delta_n}^{i\Delta_n} a_s^* \mathrm{d}s,$$

and

$$\zeta_i^n(4,2) = \frac{\sqrt{\rho_n}}{\ell_n} \int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\mathbb{R}} \delta_s(x) \left(\mathfrak{p}(\mathrm{d}s, \mathrm{d}x) - \mathfrak{q}(\mathrm{d}s, \mathrm{d}x)\right).$$

Yet again, employing standard estimates for semimartingales from Jacod and Protter (2012) means that there exists a constant $C > 0$ such that

$$\mathbb{E}\left[\big| \int_{(i-1)\Delta_n}^{i\Delta_n} a_s^* \mathrm{d}s \big|\right] \leq C\Delta_n,$$

so that

$$\begin{aligned}
\mathbb{E}\left[\left| \sum_{i \in \mathcal{D}_{t-}^n} (tn - \ell_n - i)\zeta_i^n(4,1) \right|\right] &\leq C\frac{\sqrt{\rho_n}}{\ell_n} \sum_{i \in \mathcal{D}_{t-}^n} (tn - \ell_n - i)\Delta_n \\
&= C\frac{\Delta_n \sqrt{\rho_n}}{\ell_n} \sum_{j=1}^{\ell_n} j \\
&= C\Delta_n \sqrt{\rho_n}(\ell_n + 1) \\
&\sim C \begin{cases} \delta_n^{3/2}\sqrt{n} & \text{if } \delta_n\sqrt{n} \to 0, \\ \delta_n^{1/2} & \text{if } \delta_n\sqrt{n} \to \theta \text{ or } \delta_n\sqrt{n} \to \infty. \end{cases}
\end{aligned}$$

Thus,

$$\mathbb{E}\left[\left| \sum_{i \in \mathcal{D}_{t-}^n} (tn - \ell_n - i)\zeta_i^n(4,1) \right|\right] \to 0.$$

This implies that the first sum in $\Lambda_4^n(t)$ is asymptotically negligible.

The second sum in $\Lambda_4^n(t)$ is denoted by $R^n(t) = \sum_{i \in \mathcal{D}_{t-}^n} (tn - \ell_n - i)\zeta_i^n(4,2)$ and observe that $\zeta_i^n(4,2)$ is a pure-jump martingale with $\mathbb{E}\left[\zeta_i^n(4,2)\right] = 0$. We take a $\kappa \in (0,1)$ and decompose

41

$\zeta_i^n(4, 2)$ into a "small jump" and "big jump" component defined through the bounding function $\bar{\Gamma}$, $\zeta_i^n(4, 2) = \chi_i^{n,\prime}(\kappa) + \chi_i^{n,\prime\prime}(\kappa)$, where

$$\chi_i^{n,\prime}(\kappa) = \frac{\sqrt{\rho_n}}{\ell_n} \int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\left\{x : \bar{\Gamma}(x) \leq \kappa\right\}} \delta_s(x) \big(\mathfrak{p}(\mathrm{d}s, \mathrm{d}x) - \mathfrak{q}(\mathrm{d}s, \mathrm{d}x)\big),$$

and

$$\chi_i^{n,\prime\prime}(\kappa) = \frac{\sqrt{\rho_n}}{\ell_n} \int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\left\{x : \bar{\Gamma}(x) > \kappa\right\}} \delta_s(x) \big(\mathfrak{p}(\mathrm{d}s, \mathrm{d}x) - \mathfrak{q}(\mathrm{d}s, \mathrm{d}x)\big).$$

Next, define the subset $\Omega_n(\psi, \kappa) \subseteq \Omega$ such that the Poisson process $\mathfrak{p}\big([0, t] \times \{x : \bar{\Gamma}(x) > \kappa\}\big)$ has no jumps in the interval $[t - \ell_n^\psi \Delta_n, t]$ for all $0 < \psi < 1$ and $0 < \kappa < 1$. As $n \to \infty$, we note that $\Omega_n(\psi, \kappa) \to \Omega$, because $\ell_n^\psi \Delta_n \to 0$.

Now, for all $c > 0$, the law of total probability and Markov's inequality imply that

$$\mathbb{P}\big(|R^n(t)| > c\big) = \mathbb{P}\Big(|R^n(t)| > c \mid \Omega_n^\complement(\psi, \kappa)\Big) + \mathbb{P}\big(|R^n(t)| > c \mid \Omega_n(\psi, \kappa)\big)$$

$$\leq \mathbb{P}\Big(|R^n(t)| > c \mid \Omega_n^\complement(\psi, \kappa)\Big) + \frac{\mathbb{E}\left[\big(R^n(t)\big)^2 \mid \Omega_n(\psi, \kappa)\right]}{c^2},$$

where $\Omega_n^\complement(\psi, \kappa)$ is the complement of $\Omega_n(\psi, \kappa)$.

As there are no common jumps in $\chi_i^{n,\prime}(\kappa)$ and $\chi_i^{n,\prime\prime}(\kappa)$:

$$\mathbb{E}\left[\big(R^n(t)\big)^2 \mid \Omega_n(\psi, \kappa)\right] = \sum_{i \in \mathcal{D}_{t-}^n} (tn - \ell_n - i)^2 \left(\mathbb{E}\left[\big(\chi_i^{n,\prime}(\kappa)\big)^2 \mid \Omega_n(\psi, \kappa)\right] + \mathbb{E}\left[\big(\chi_i^{n,\prime\prime}(\kappa)\big)^2 \mid \Omega_n(\psi, \kappa)\right]\right).$$

The small jump term can be controlled as

$$\mathbb{E}\left[\big(\chi_i^{n,\prime}(\kappa)\big)^2\right] \leq C \frac{\rho_n}{\ell_n^2} \Delta_n \int_{\left\{x : \bar{\Gamma}(x) \leq \kappa\right\}} \bar{\Gamma}(x)^2 F(\mathrm{d}x).$$

Thus,

$$\sum_{i \in \mathcal{D}_{t-}^n} (tn - \ell_n - i)^2 \mathbb{E}\left[\big(\chi_i^{n,\prime}(\kappa)\big)^2\right] \leq C \frac{\rho_n}{\ell_n^2} \Delta_n \int_{\left\{x : \bar{\Gamma}(x) \leq \kappa\right\}} \bar{\Gamma}(x)^2 F(\mathrm{d}x) \sum_{i \in \mathcal{D}_{t-}^n} (tn - \ell_n - i)^2$$

$$= C \left[\frac{\rho_n \Delta_n \ell_n (\ell_n + 1)(2\ell_n + 1)}{6\ell_n^2}\right] \int_{\left\{x : \bar{\Gamma}(x) \leq \kappa\right\}} \bar{\Gamma}(x)^2 F(\mathrm{d}x),$$

where

$$\left[\frac{\rho_n \Delta_n \ell_n (\ell_n + 1)(2\ell_n + 1)}{6\ell_n^2}\right] \sim \rho_n \delta_n = \begin{cases} \delta_n^2 n, & \text{if } \delta_n \sqrt{n} \to 0, \\ 1, & \text{if } \delta_n \sqrt{n} \to \theta \text{ or } \delta_n \sqrt{n} \to \infty. \end{cases}$$

Furthermore, because there are no big jumps (larger than $\kappa$) in the interval $[t - \ell_n^\psi \Delta_n, t]$ on the event $\Omega_n(\psi, \kappa)$, we can also conclude that:

$$\sum_{i \in \mathcal{D}_{t-}^n} (tn - \ell_n - i)^2 \mathbb{E}\left[\big(\chi_i^{n,\prime\prime}(\kappa)\big)^2 \mid \Omega_n(\psi, \kappa)\right] = \sum_{i \in \mathcal{D}_{t-}^n, i\Delta_n \leq t - \ell_n^\psi \Delta_n} (tn - \ell_n - i)^2 \mathbb{E}\left[\big(\chi_i^{n,\prime\prime}(\kappa)\big)^2\right]$$

$$\leq C \frac{\rho_n}{\ell_n^2} \Delta_n \sum_{i \in \mathcal{D}_{t-}^n, i\Delta_n \leq t - \ell_n^\psi \Delta_n} (tn - \ell_n - i)^2,$$

where the sum on the right-hand side has the following form:

$$\sum_{i \in \mathcal{D}_{t-}^n, i\Delta_n \leq t - \ell_n^\psi \Delta_n} (tn - \ell_n - i)^2 = \sum_{j=1}^{\ell_n^\psi} j^2 = \frac{\ell_n^\psi (\ell_n^\psi + 1)(2\ell_n^\psi + 1)}{6} \sim \ell_n^{3\psi}. \tag{A.3}$$

This produces

$$\sum_{i \in \mathcal{D}_{t-}^n} (tn - \ell_n - i)^2 \mathbb{E}\left[ \left( \chi_i^{n,\prime\prime}(\kappa) \right)^2 \mid \Omega_n(\psi, \kappa) \right] \sim \rho_n \Delta_n \ell_n^{3\psi - 2} \to 0, \tag{A.4}$$

and it thus follows that

$$\limsup_{n\to\infty} \mathbb{P}\big( |R^n(t)| > c \big) \leq \limsup_{n\to\infty} \mathbb{P}\big( |R^n(t)| > c \mid \bar{\Omega}_n(\psi, \kappa) \big) + \limsup_{n\to\infty} A_n, \tag{A.5}$$

where

$$A_n \sim \rho_n \delta_n \int_{\{x : \bar{\Gamma}(x) \leq \kappa\}} \bar{\Gamma}(x)^2 F(\mathrm{d}x) + \rho_n \Delta_n \ell_n^{3\psi - 2}. \tag{A.6}$$

As $\kappa \to 0$, $\int_{\{x : \bar{\Gamma}(x) \leq \kappa\}} \bar{\Gamma}(x)^2 F(\mathrm{d}x) \to 0$, so that as $\kappa \to 0$ and $n \to \infty$, $A_n \to 0$. Moreover, as $n \to \infty$, $\Omega_n^{\complement}(\psi, \kappa) \to \emptyset$, so $\limsup_{n\to\infty} \mathbb{P}\big( |R^n(t)| > c \mid \Omega_n^{\complement}(\psi, \kappa) \big) = 0$. To conclude, $\limsup_{n\to\infty} \mathbb{P}\big( |R^n(t)| > c \big) = 0$ and, hence, $\lim_{n\to\infty} \mathbb{P}\big( |R^n(t)| > c \big) = 0$, as was to be demonstrated. ∎

## A.3 Proof of Theorem 2

As in proof of Theorem 1, we start with a decomposition of $\widehat{\lambda}_t - \mu_{t-}$:

$$\sqrt{\ell_n}(\widehat{\lambda}_t - \mu_{t-}) = \Lambda_1^n(t) + \Lambda_2^n(t) + \Lambda_3^n(t) + (\mathrm{III}),$$

where $\Lambda_1^n(t) - \Lambda_3^n(t)$ are defined as above, and

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

The treatment of the first triplet does not change from before, so it follows that $\Lambda_1^n(t)$ and $\Lambda_2^n(t)$ are asymptotically negligible, whereas $\Lambda_3^n(t)$ converges stably in law:

$$\Lambda_3^n(t) \xrightarrow{\mathfrak{D}_s} \sqrt{\mu_{t-}} \mathcal{Z}. \tag{A.7}$$

The difference is that we do not split (III) into a drift and martingale part, $\Lambda_4^n(t)$ and $\Lambda_5^n(t)$, as there is no such structural assumption to work with now. So, our sole task is to prove (III) is asymptotically negligible under Assumption 3.

We set $\zeta_i^n = \frac{1}{\sqrt{\ell_n}}(tn - \ell_n - i)(\mu_{i\Delta_n} - \mu_{(i-1)\Delta_n})$, such that $(\mathrm{III}) = \sum_{i \in \mathcal{D}_{t-}^n} \zeta_i^n$. It is then straight-

forward to show that:

$$\sum_{i\in\mathcal{D}_{t-}^n} \mathbb{E}\big[|\zeta_i^n|\big] \leq C\Delta_n^{H^\mu}\frac{1}{\sqrt{\ell_n}}\sum_{i\in\mathcal{D}_{t-}^n}(tn-\ell_n-i)$$

$$= C\Delta_n^{H^\mu}\frac{1}{\sqrt{\ell_n}}\sum_{j=1}^{\ell_n} j$$

$$= C\Delta_n^{H^\mu}\sqrt{\ell_n}(\ell_n+1)$$

$$\sim \ell_n^{3/2}\Delta_n^{H^\mu},$$

which converges to zero so long as we select $\ell_n = O\big(n^{\frac{2H^\mu}{3}-\epsilon}\big)$, for some $0 < \epsilon < \frac{2H^\mu}{3}$. This implies that (III) is asymptotically negligible. The upper bound on $\epsilon$ ensures that $\ell_n \to \infty$. ∎

## A.4    Proof of Theorem 3

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

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

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

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

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

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

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

Then,

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

44

where

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

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

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

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

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

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

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

it follows that

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

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

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

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

So $\mathbb{E}\left[\mathcal{E}_1'^n(t)\right] = 0$, 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:

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

Next,

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

45

Alluding to the boundedness of $\mu$, we get the estimates:

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

Based on this, we deduce that:

$$\mathbb{E}\left[\left(\Lambda_3^n(t_j)\right)^4\right] \le \frac{\rho_n^2}{\ell_n^4}\left(\sum_{i=1}^{2\ell_n} C + 4\sum_{i=1}^{2\ell_n-1}\sum_{s=i+1}^{2\ell_n} C\right) = C\frac{\rho_n^2(2\ell_n + 4\ell_n(2\ell_n - 1))}{\ell_n^4}. \tag{A.8}$$

Ergo,

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

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

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

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

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

we have:

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

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

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

46

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

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

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

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

$$\sum_{i \in \mathcal{D}_{t_j}^n} (t_j n - 2\ell_n - i)^2 = \frac{8\ell_n^3 + 6\ell_n^2 + \ell_n}{3}. \tag{A.9}$$

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 $\nu$ is càdlàg and $K_n\ell_n\Delta_n \to 0$, $\overline{M} = o_p(1)$. The sum was computed in (A.9), so combining the terms shows that

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

47

where

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

As such, $\mathcal{E}_2'''^n(t) = o_p(1)$.

Following the above steps, we also deduce that

$$\mathcal{E}_2''''^n(t) \xrightarrow{p} 0.$$

The last term can be written as

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

where the sum on the right-hand side is given by (A.9). As a consequence,

$$\mathcal{E}_2'''''^n(t) \xrightarrow{p} \begin{cases} 0, & \text{if } \ell_n\sqrt{\Delta_n} \rightarrow 0, \\ \dfrac{8}{3}\nu_t^2, & \text{if } \ell_n\sqrt{\Delta_n} \rightarrow \theta > 0 \text{ or } \ell_n\sqrt{\Delta_n} \rightarrow \infty, \end{cases}$$

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.5   Proof of Theorem 4

We largely copy from the proof of Theorem 3. 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:

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

where $t_j = t - j\Delta_n$.

This can further be split into

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

where

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

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

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

48

The only difference here compared to the proof of Theorem 3 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:

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

To show $\mathbb{E}\left[ \left( \mathcal{E}_1'^n(t) \right)^2 \right]$ converges to zero, we look at each part in the above equation separately. First, by Theorem 3—employing the estimate in (A.8)—we deduce that:

$$\frac{1}{K_n^2} \sum_{j=0}^{K_n-1} \mathbb{E}\left[ \left( \epsilon(t_j) \right)^2 \right] \leq C \frac{\ell_n}{K_n^2} \sum_{j=0}^{K_n-1} \mathbb{E}\left[ \left( \Lambda_3^n(t_j) \right)^4 \right] = O\left( K_n^{-1}\ell_n^{-2}\rho_n^2 \right).$$

Next, to deal with the cross-product term we express $\epsilon(t_j)$ as

$$\epsilon(t_j) = \sum_{i \in \mathcal{D}_{t_j}^n} \chi_i + \mathcal{R}_{t_j}(t),$$

where $\chi_i = \left( \left( \zeta_i(3) \right)^2 - \frac{\rho_n}{\ell_n}\mu_{(i-1)\Delta_n} \right)$ is an $\mathcal{F}_{i\Delta_n}$-martingale difference sequence and $\mathcal{R}_{t_j}(t)$ is a reminder term of the form:

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

We note that for $s > 2\ell_n j$, $\mathcal{D}_{t_j}^n \cap \mathcal{D}_{t_s}^n = \emptyset$, which implies that $\mathbb{E}\left[ \epsilon(t_j)\epsilon(t_s) \right] = 0$. In contrast for $s \leq 2\ell_n j$, $\mathcal{D}_{t_j}^n \cap \mathcal{D}_{t_s}^n \neq \emptyset$, but nevertheless

$$\mathbb{E}\left[ \mathcal{R}_{t_j}(t)\mathcal{R}_{t_s}(t) \right] = \mathbb{E}\left[ \epsilon(t_j)\mathcal{R}_{t_s}(t) \right] = \mathbb{E}\left[ \epsilon(t_s)\mathcal{R}_{t_j}(t) \right] = 0,$$

since $\zeta_i(3)$ is also a $\mathcal{F}_{i\Delta_n}$-martingale difference.

49

Thus,

$$\frac{2}{K_n^2} \sum_{j=0}^{K_n-2} \sum_{s=j+1}^{K_n-1} \mathbb{E}\big[\epsilon(t_j)\epsilon(t_s)\big] = \frac{2}{K_n^2} \sum_{j=0}^{K_n-2} \sum_{s=j+1}^{j+2\ell_n} \mathbb{E}\Big[\sum_{i\in\mathcal{D}_{t_j}^n} \chi_i \sum_{i\in\mathcal{D}_{t_s}^n} \chi_i\Big].$$

For each summand on the right-hand side of the above equation,

$$\mathbb{E}\Big[\sum_{i\in\mathcal{D}_{t_j}^n} \chi_i \sum_{i\in\mathcal{D}_{t_s}^n} \chi_i\Big] = \sum_{i\in\mathcal{D}_{t_j}^n \cap \mathcal{D}_{t_s}^n} \mathbb{E}\big[\chi_i^2\big].$$

By the boundedness of $\mu$, $\mathbb{E}\Big[\big(\zeta_i^n(3)\big)^4\Big] \leq C\frac{\rho_n^2}{\ell_n^4}$ and this bound also applies to $\mathbb{E}\big[\chi_i^2\big]$. As the intersection $\mathcal{D}_{t_j}^n \cap \mathcal{D}_{t_s}^n$ contains at most $2\ell_n$ terms, it therefore shows that:

$$\sum_{i\in\mathcal{D}_{t_j}^n \cap \mathcal{D}_{t_s}^n} \mathbb{E}\big[\chi_i^2\big] \leq C\ell_n^{-3}\rho_n^2.$$

Consequently,

$$\frac{2}{K_n^2} \sum_{j=0}^{K_n-2} \sum_{s=j+1}^{j+2\ell_n} \mathbb{E}\Big[\sum_{i\in\mathcal{D}_{t_j}^n} \chi_i \sum_{i\in\mathcal{D}_{t_s}^n} \chi_i\Big] \leq CK_n^{-1}\ell_n^{-2}\rho_n^2.$$

Combining the orders of the terms that make up $\mathbb{E}\Big[\big(\mathcal{E}_1'^n(t)\big)^2\Big]$, we deduce that

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

Now, by the assumptions made in the theorem, it follows that $\mathbb{E}\Big[\big(\mathcal{E}_1'^n(t)\big)^2\Big] = o(1)$. 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 $\mu$ is càdlàg, as $\Delta_n \to 0$,

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

Thus, alluding to the definition of $\rho_n$, 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.6 Proof of Theorem 5

The statement of the theorem under $\mathcal{H}_0$ follows immediately by combining Theorem 1 and Theorem 4 with Slutsky's theorem.

We next look at $\mathcal{H}_1$. We use the version of the test statistic based on the non-overlapping observed asymptotic local variance in (3.20), $\widetilde{\text{avar}}(\nabla\widehat{\lambda}_t)$. In that instance,

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

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

We next derive a lower (upper) bound on the stochastic order of the numerator (denominator). We show that the numerator diverges *at least* as fast as the numerator. 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

$$\begin{aligned}
\mathbb{E}\big[|D_{\tau,j}|\big] &\leq \mathbb{E}\left[\frac{N^{n,\beta}(\tau-(1+2j)\delta_n, \tau-2j\delta_n) + N^{n,\beta}(\tau-(2+2j)\delta_n, \tau-(1+2j)\delta_n)}{n\delta_n}\right]\\
&= \mathbb{E}\left[\frac{N^{n,\beta}(\tau-(2+2j)\delta_n, \tau-2j\delta_n)}{n\delta_n}\right]\\
&= \mathbb{E}\left[\frac{1}{\delta_n}\int_{\tau-(2+2j)\delta_n}^{\tau-2j\delta_n}\sigma_u|\tau-u|^{-\alpha}\mathrm{d}u\right]\\
&\leq C\delta_n^{-\alpha}\big((j+1)^{1-\alpha} - j^{1-\alpha}\big),
\end{aligned} \tag{A.10}$$

by the proof of Lemma 1.

By the mean-value theorem there exists a $j^* \in (j, j+1)$, such that

$$(j+1)^{1-\alpha} - j^{1-\alpha} = (j^*)^{-\alpha} \leq j^{-\alpha}.$$

Therefore,

$$\mathbb{E}\big[|D_{\tau,j}|\big] \leq C\delta_n^{-\alpha}j^{-\alpha}.$$

Thus, for $j$ fixed:

$$D_{\tau,j} = O_p(\delta_n^{-\alpha}).$$

as $\delta_n \to 0$.

Regarding the denominator,

$$\mathbb{E}\left[\frac{1}{K_n}\sum_{j=0}^{K_n-1}D_{\tau,j}^2\right] = \frac{1}{K_n}\sum_{j=0}^{K_n-1}\mathbb{E}\left[D_{\tau,j}^2\right],$$

where

$$\mathbb{E}\left[D_{\tau,j}^2\right] = \mathbb{E}\left[\left(\frac{N^{n,\beta}(\tau-(1+2j)\delta_n,\tau-2j\delta_n)-N^{n,\beta}(\tau-(2+2j)\delta_n,\tau-(1+2j)\delta_n)}{n\delta_n}\right)^2\right]$$

$$\leq \frac{1}{n^2\delta_n^2}\mathbb{E}\left[N^{n,\beta}(\tau-(2+2j)\delta_n,\tau-2j\delta_n)^2\right]$$

$$= \frac{1}{n^2\delta_n^2}\left(\mathbb{E}\left[N^{n,\beta}(\tau-(2+2j)\delta_n,\tau-2j\delta_n)\right]+\mathbb{E}\left[N^{n,\beta}(\tau-(2+2j)\delta_n,\tau-2j\delta_n)\right]^2\right),$$

due to the properties of the Poisson distribution.

It follows from (A.10) that

$$\mathbb{E}\left[N^{n,\beta}(\tau-(2+2j)\delta_n,\tau-2j\delta_n)\right] \leq Cn\delta_n^{1-\alpha}j^{-\alpha},$$

Hence, by the rate conditions imposed in Theorem 5, $n\delta_n \to \infty$ and $\delta^{-\alpha} \to \infty$, so the above expectation diverges. This implies that the square of the expectation is the leading term, so it suffices to look at the second summand in the above decomposition:

$$\frac{1}{K_n}\sum_{j=0}^{K_n-1}\frac{1}{n^2\delta_n^2}\mathbb{E}\left[N^{n,\beta}(\tau-(2+2j)\delta_n,\tau-2j\delta_n)\right]^2 \leq C\delta_n^{-2\alpha}\frac{1}{K_n}\sum_{j=0}^{K_n-1}j^{-\alpha} = O\left(\left[\delta_nK_n\right]^{-2\alpha}\right)$$

since $\sum_{j=0}^{K_n-1}j^{-2\alpha} = O(K_n^{1-2\alpha})$. Thus, the denominator is *at most* $O_p\left(\left[\delta_nK_n\right]^{-\alpha}\right)$.

Now, we look at the numerator:

$$D_{\tau,0} = \frac{N^{n,\beta}(\tau-\delta_n,\tau)-N^{n,\beta}(\tau-2\delta_n,\tau-\delta_n)}{n\delta_n}.$$

Let

$$\mathfrak{B}_{1,\tau}^n = \frac{N^{n,\beta}(\tau-\delta_n,\tau)}{n\delta_n}-\frac{1}{\delta_n}\int_{\tau-\delta_n}^{\tau}\frac{\sigma_s}{|\tau-s|^\alpha}\mathrm{d}s.$$

As in the proof of Lemma 1, we can condition on $\mathcal{F}^\sigma$, the $\sigma$-algebra generated by $\sigma$, and then employ the law of iterated expectations to obtain:

$$\mathbb{E}\left[\mathfrak{B}_{1,\tau}^n\right] = \mathbb{E}\left[\mathbb{E}\left[\mathfrak{B}_{1,\tau}^n\mid\mathcal{F}^\sigma\right]\right] = 0,$$

and, by the law of total variance,

$$\mathrm{var}\left(\mathfrak{B}_{1,\tau}^n\right) = \mathbb{E}\left[\mathrm{var}\left(\mathfrak{B}_{1,\tau}^n\mid\mathcal{F}^\sigma\right)\right]+\underbrace{\mathrm{var}\left(\mathbb{E}\left[\mathfrak{B}_{1,\tau}^n\mid\mathcal{F}^\sigma\right]\right)}_{=0},$$

where

$$\mathrm{var}\left(\mathfrak{B}_{1,\tau}^n\mid\mathcal{F}^\sigma\right) = \frac{1}{n^2\delta_n^2}n\int_{\tau-\delta_n}^{\tau}\frac{\sigma_s}{|\tau-s|^\alpha}\mathrm{d}s$$

$$\leq \frac{C}{n\delta_n^2} \int_{\tau-\delta_n}^{\tau} \frac{1}{|\tau-s|^\alpha} ds = \frac{C}{n\delta_n^2} \delta_n^{1-\alpha} = C\frac{1}{n\delta_n^{(1+\alpha)}} = O\left(n^{-1}\delta_n^{-(1+\alpha)}\right).$$

Thus,

$$\mathrm{var}\left(\delta_n^{\alpha/2}\mathfrak{B}_{1,\tau}^n\right) \leq \frac{C}{n\delta_n} \to 0,$$

which has the implication that

$$\delta_n^{\alpha/2}\left(\frac{N^{n,\beta}(\tau-\delta_n,\tau)}{n\delta_n} - \frac{1}{\delta_n}\int_{\tau-\delta_n}^{\tau} \frac{\sigma_s}{|\tau-s|^\alpha} ds\right) \xrightarrow{p} 0.$$

Analogously,

$$\delta_n^{\alpha/2}\left(\frac{N^{n,\beta}(\tau-2\delta_n,\tau-\delta_n)}{n\delta_n} - \frac{1}{\delta_n}\int_{\tau-2\delta_n}^{\tau-\delta_n} \frac{\sigma_s}{|\tau-s|^\alpha} ds\right) \xrightarrow{p} 0,$$

such that

$$\delta_n^{\alpha/2}\left(D_{\tau,0} - \left[\frac{1}{\delta_n}\int_{\tau-\delta_n}^{\tau} \frac{\sigma_s}{|\tau-s|^\alpha} ds - \frac{1}{\delta_n}\int_{\tau-2\delta_n}^{\tau-\delta_n} \frac{\sigma_s}{|\tau-s|^\alpha} ds\right]\right) \xrightarrow{p} 0.$$

Next, we define $\underline{\sigma} = \inf_{s\in[\tau-2\delta_n]}\sigma_s$ and $\overline{\sigma} = \sup_{s\in[\tau-2\delta_n]}\sigma_s$. Then, by construction

$$\frac{1}{\delta_n}\int_{\tau-\delta_n}^{\tau} \frac{\sigma_s}{|\tau-s|^\alpha} ds - \frac{1}{\delta_n}\int_{\tau-2\delta_n}^{\tau-\delta_n} \frac{\sigma_s}{|\tau-s|^\alpha} ds \geq \frac{1}{\delta_n}\left(\int_{\tau-\delta_n}^{\tau} \frac{\underline{\sigma}}{|\tau-s|^\alpha} ds - \int_{\tau-2\delta_n}^{\tau-\delta_n} \frac{\overline{\sigma}}{|\tau-s|^\alpha} ds\right)$$

$$= \frac{1}{\delta_n}\left(\frac{\underline{\sigma}}{1-\alpha}\delta_n^{1-\alpha} - \frac{\overline{\sigma}}{1-\alpha}(2^{1-\alpha}-1)\delta_n^{1-\alpha}\right)$$

$$= \frac{\underline{\sigma}-\overline{\sigma}(2^{1-\alpha}-1)}{1-\alpha}\delta_n^{-\alpha}.$$

Since $\overline{\sigma}-\underline{\sigma} \geq |\sigma_\tau - \sigma_{\tau-2\delta_n}|$, it follows that for every $\epsilon > 0 : \{\omega \in \Omega : \overline{\sigma} - \underline{\sigma} > \epsilon\} \subseteq \{\omega \in \Omega : |\sigma_\tau - \sigma_{\tau-2\delta_n}| > \epsilon\}$, and so

$$\mathbb{P}\left(|\overline{\sigma}-\underline{\sigma}| > \epsilon\right) \leq \mathbb{P}\left(|\sigma_\tau - \sigma_{\tau-\delta_n}| > \varepsilon\right) \longrightarrow 0,$$

as $\delta_n \to 0$, where the right-hand side follows by stochastic continuity of $\sigma$. In other words, for a large enough $n$, $\underline{\sigma}$ is arbitrarily close to $\overline{\sigma}$ in probability. Consequently,

$$\mathbb{P}\left(\frac{\underline{\sigma}-\overline{\sigma}(2^{1-\alpha}-1)}{1-\alpha} > 0\right) \longrightarrow 1,$$

as $\delta_n \to 0$. Hence,

$$\frac{1}{\delta_n}\int_{\tau-\delta_n}^{\tau} \frac{\sigma_s}{|\tau-s|^\alpha} ds - \frac{1}{\delta_n}\int_{\tau-2\delta_n}^{\tau-\delta_n} \frac{\sigma_s}{|\tau-s|^\alpha} ds \xrightarrow{p} \infty. \tag{A.11}$$

so the explosion rate of the numerator is *at least* $\delta^{-\alpha}$. This is faster than the explosion rate of the denominator (at most $(\delta_n K_n)^{-\alpha}$), so the test statistic diverges under the alternative. ∎

## A.7 Proof of Lemma 2

*Part (i) — Conditional on $\omega \in \Omega_1^B$:*

The observed process can be broken into the parts

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

where $N_t^{n,\mu}$ and $N_t^{n,\beta}$ are inhomogeneous Poisson processes with rates $n\mu_t$ and $n\beta_t$. This yields the following decomposition for $\widetilde{\lambda}_\theta(k\delta_n)$:

$$\widetilde{\lambda}_\theta(k\delta_n) = \widetilde{\mu}_\theta(k\delta_n) + \widetilde{\beta}_\theta(k\delta_n),$$

where $\widetilde{\mu}_\theta(k\delta_n) \xrightarrow{p} \mu_{\theta-}$ and

$$\widetilde{\beta}_\theta(k\delta_n) = \frac{N_{\theta+k\delta_n}^{n,\beta} - N_{\theta-k\delta_n}^{n,\beta}}{2nk\delta_n}.$$

Since $\delta_n \to 0$ and $\alpha > 0$, the convergence $\widetilde{\mu}_\theta(k\delta_n) \xrightarrow{p} \mu_{\theta-}$ further implies that $\delta_n^{\alpha/2}\widetilde{\mu}_\theta(k\delta_n) \xrightarrow{p} 0$. It immediately follows that

$$\delta_n^{\alpha/2}\left(\widetilde{\lambda}_\theta(k\delta_n) - \sigma_{\theta-}(2k\delta_n)^{-\alpha}\right) - \delta_n^{\alpha/2}\left(\widetilde{\beta}_\theta(k\delta_n) - \sigma_{\theta-}(2k\delta_n)^{-\alpha}\right) \xrightarrow{p} 0.$$

Note that $\sigma_\theta(2k\delta_n)^{-\alpha} = \frac{1}{2k\delta_n}\int_{\theta-k\delta_n}^{\theta+k\delta_n}\frac{\sigma_\theta}{|\theta-s|^\alpha}ds$. This implies that

$$\delta_n^{\alpha/2}\left(\widetilde{\beta}_\theta(k\delta_n) - \sigma_\theta(2k\delta_n)^{-\alpha}\right) = \delta_n^{\alpha/2}\left(\mathfrak{B}_{1,\theta}^n + \mathfrak{B}_{2,\theta}^n\right), \qquad (A.12)$$

where

$$\mathfrak{B}_{1,\theta}^n = \widetilde{\beta}_\theta(k\delta_n) - \frac{1}{2k\delta_n}\int_{\theta-k\delta_n}^{\theta+k\delta_n}\frac{\sigma_s}{|\theta-s|^\alpha}ds,$$

$$\mathfrak{B}_{2,\theta}^n = \frac{1}{2k\delta_n}\int_{\theta-k\delta_n}^{\theta+k\delta_n}\frac{\sigma_s}{|\theta-s|^\alpha}ds - \frac{1}{2k\delta_n}\int_{\theta-k\delta_n}^{\theta+k\delta_n}\frac{\sigma_\theta}{|\theta-s|^\alpha}ds.$$

Following the footsteps in the proof of Theorem 5 by conditioning on $\mathcal{F}^\sigma$ and employing the laws of iterated expectation and total variance, we find that

$$\mathbb{E}\left[\mathfrak{B}_{1,\theta}^n\right] = 0 \quad \text{and} \quad \text{var}\left(\mathfrak{B}_{1,\theta}^n\right) = \mathbb{E}\left[\text{var}\left(\mathfrak{B}_{1,\theta}^n \mid \mathcal{F}^\sigma\right)\right],$$

where $\text{var}\left(\mathfrak{B}_{1,\theta}^n \mid \mathcal{F}^\sigma\right) = O_p\left(n^{-1}\delta_n^{-(1+\alpha)}\right)$, so that $\delta_n^{\alpha/2}\mathfrak{B}_{1,\theta}^n \xrightarrow{p} 0$.

$\mathfrak{B}_{2,\theta}^n$ can be bounded by

$$\begin{aligned}
|\mathfrak{B}_{2,\theta}^n| &= \left|\frac{1}{2k\delta_n}\int_{\theta-k\delta_n}^{\theta+k\delta_n}\frac{\sigma_s - \sigma_\theta}{|\theta-s|^\alpha}ds\right| \\
&\leq \frac{1}{2k\delta_n}\int_{\theta-k\delta_n}^{\theta+k\delta_n}\frac{|\sigma_s - \sigma_\theta|}{|\theta-s|^\alpha}ds \\
&\leq \frac{(2k\delta_n)^{H^\sigma}}{2k\delta_n}\int_{\theta-k\delta_n}^{\theta+k\delta_n}\frac{L}{|\theta-s|^\alpha}ds = (2k\delta_n)^{H^\sigma-\alpha}L.
\end{aligned}$$

Hence,

$$\delta_n^{\alpha/2}|\mathfrak{B}_{2,\theta}^n| \leq C\delta_n^{H^\sigma-\alpha/2}L \xrightarrow{p} 0,$$

because $L$ is a bounded random variable and $H^\sigma - \alpha/2 > 0$. Thus, $\delta_n^{\alpha/2} |\mathfrak{B}_{1,\theta}^n|$ and $\delta_n^{\alpha/2} |\mathfrak{B}_{2,\theta}^n|$ converge to zero in probability.

Combining this fact with (A.12),

$$\delta_n^{\alpha/2} \left( \widetilde{\lambda}_\theta(k\delta_n) - \sigma_\theta(2k\delta_n)^{-\alpha} \right) \xrightarrow{p} 0.$$

for any $k \geq 1$. This means that:

$$\frac{\widetilde{\lambda}_\theta(k\delta_n)}{\widetilde{\lambda}_\theta(\delta_n)} = \frac{\delta_n^{\alpha/2} \left( \widetilde{\lambda}_\theta(k\delta_n) - \sigma_\theta(2k\delta_n)^{-\alpha} \right) + \delta_n^{\alpha/2} \left( \sigma_\theta(2k\delta_n)^{-\alpha} \right)}{\delta_n^{\alpha/2} \left( \widetilde{\lambda}_\theta(k\delta_n) - \sigma_\theta(2k\delta_n)^{-\alpha} \right) + \delta_n^{\alpha/2} \left( \sigma_\theta(2\delta_n)^{-\alpha} \right)} \xrightarrow{p} \frac{\delta_n^{\alpha/2} \left( \sigma_\theta(2k\delta_n)^{-\alpha} \right)}{\delta_n^{\alpha/2} \left( \sigma_\theta(2\delta_n)^{-\alpha} \right)} = k^{-\alpha},$$

which completes the proof of part (i).

*Part (ii) — Conditional on $\omega \in \Omega_1^J$:*

We start by observing that $\widetilde{\lambda}_\theta(k\delta_n)$ can be represented as the average of a forward- and backward-looking intensity estimator:

$$\widetilde{\lambda}_\theta(k\delta_n) = \frac{1}{2} \left( \widehat{\lambda}_\theta^{(-)}(k\delta_n) + \widehat{\lambda}_\theta^{(+)}(k\delta_n) \right).$$

The proof of Theorem 2 implies that $\widehat{\lambda}_\theta^{(-)}(k\delta_n) \xrightarrow{p} \mu_{\theta-}$. As the forward-looking estimator $\widehat{\lambda}_\theta^{(+)}(k\delta_n)$ is the mirror image of $\widehat{\lambda}_\theta^{(-)}(k\delta_n)$, we also deduce that $\widehat{\lambda}_\theta^{(+)}(k\delta_n) \xrightarrow{p} \mu_{\theta+}$, where $\mu_{\theta+} = \mu_{\theta-} + \Delta\mu_\theta$. In summary, by the continuous mapping theorem

$$\widetilde{\lambda}_\theta(k\delta_n) \xrightarrow{p} \frac{1}{2} \left( \mu_{\theta-} + \mu_{\theta-} + \Delta\mu_\theta \right) = \mu_{\theta-} + \frac{\Delta\mu_\theta}{2}.$$

As this convergence holds for every $k \geq 1$,

$$\frac{\widetilde{\lambda}_\theta(k\delta_n)}{\widetilde{\lambda}_\theta(\delta_n)} \xrightarrow{p} 1.$$

∎

## A.8 Proof of Theorem 6

The proof of the theorem consists of deriving the bivariate asymptotic distribution of the change-of-frequency intensity estimator computed with different $k$:

$$\sqrt{n\delta_n} \begin{pmatrix} \widetilde{\lambda}_\theta(k\delta_n) - \mu_{\theta-} - \dfrac{\Delta\mu_\theta}{2} \\ \widetilde{\lambda}_\theta(\delta_n) - \mu_{\theta-} - \dfrac{\Delta\mu_\theta}{2} \end{pmatrix} \xrightarrow{\mathfrak{D}_s} \Sigma^{1/2} \mathcal{Z}_2, \tag{A.13}$$

where $\mathcal{Z}_2 \sim N(0, I_2)$ and $\Sigma$ is the asymptotic covariance matrix. The statement of the theorem then follows by application of the delta method.

To derive (A.13), by the Cramér-Wold device it suffices to show that every fixed linear combi-

nation

$$\mathbf{L}_n = c_1 \sqrt{n\delta_n} \left( \widetilde{\lambda}_\theta(k\delta_n) - \mu_{\theta-} - \frac{\Delta\mu_\theta}{2} \right) + c_2 \sqrt{n\delta_n} \left( \widetilde{\lambda}_\theta(\delta_n) - \mu_{\theta-} - \frac{\Delta\mu_\theta}{2} \right),$$

converges stably in law, i.e.

$$\mathbf{L}_n \xrightarrow{\mathfrak{D}_s} \sqrt{c_1^2 \Sigma_{1,1} + c_2^2 \Sigma_{2,2} + 2c_1 c_2 \Sigma_{1,2}} \, \mathcal{Z},$$

where $(c_1, c_2) \in \mathbb{R}^2$ and $\mathcal{Z} \sim N(0,1)$. To prove this, we split $\widetilde{\lambda}_\theta(k\delta_n)$ into the average of the forward- and backward-looking intensity estimator:

$$\sqrt{n\delta_n} \left( \widetilde{\lambda}_\theta(k\delta_n) - \mu_{\theta-} + \frac{\Delta\mu_\theta}{2} \right) = \frac{1}{2} \left[ \sqrt{n\delta_n} \left( \widetilde{\lambda}_\theta^{(-)}(k\delta_n) - \mu_{\theta-} \right) + \sqrt{n\delta_n} \left( \widetilde{\lambda}_\theta^{(+)}(k\delta_n) - \mu_{\theta+} \right) \right].$$

Next, both $\widetilde{\lambda}_\theta^{(-)}(k\delta_n)$ and $\widetilde{\lambda}_\theta^{(+)}(k\delta_n)$ are decomposed as in Theorem 2. For example,

$$\sqrt{n\delta_n} \left( \widetilde{\lambda}_\theta^{(-)}(k\delta_n) - \mu_{\theta-} \right) = \sum_{j=1}^{4} \Lambda_j^{n,(-)}(\theta; k),$$

where

$$\Lambda_1^{n,(-)}(\theta; k) = \frac{1}{k\sqrt{\ell_n}} \sum_{i=n\theta-k\ell_n+1}^{n\theta} \left( \Delta_i N^n - \Delta_i \widetilde{N}^n - \mathbb{E}\left[ \Delta_i N^n - \Delta_i \widetilde{N}^n \mid \mathcal{F}_{(i-1)\Delta_n} \right] \right),$$

$$\Lambda_2^{n,(-)}(\theta; k) = \frac{1}{k\sqrt{\ell_n}} \sum_{i=n\theta-k\ell_n+1}^{n\theta} \mathbb{E}\left[ \Delta_i N^n - \Delta_i \widetilde{N}^n \mid \mathcal{F}_{(i-1)\Delta_n} \right],$$

$$\Lambda_3^{n,(-)}(\theta; k) = \frac{1}{k\sqrt{\ell_n}} \sum_{i=n\theta-k\ell_n+1}^{n\theta} \left( \Delta_i \widetilde{N}^n - \mu_{(i-1)\Delta_n} \right),$$

$$\Lambda_4^{n,(-)}(\theta; k) = \frac{1}{k\sqrt{\ell_n}} \sum_{i=n\theta-k\ell_n+1}^{n\theta} (n\theta - \ell_n - i)(\mu_{i\Delta_n} - \mu_{(i-1)\Delta_n}).$$

Proceeding as above,

$$\mathbf{L}_n = L_n^{(-)} + L_n^{(+)} + R_n,$$

where

$$L_n^{(-)} = \frac{1}{2} \left( c_1 \Lambda_3^{n,(-)}(\theta; k) + c_2 \Lambda_3^{n,(-)}(\theta; 1) \right), \qquad L_n^{(+)} = \frac{1}{2} \left( c_1 \Lambda_3^{n,(+)}(\theta; k) + c_2 \Lambda_3^{n,(+)}(\theta; 1) \right),$$

are the leading terms and the remainder is

$$R_n = \sum_{j=1,j\neq 3}^{4} \left( c_1 \Lambda_1^{n,(+)}(\theta; k) + c_2 \Lambda_1^{n,(+)}(\theta; 1) \right) + \sum_{j=1,j\neq 3}^{4} \left( c_1 \Lambda_1^{n,(-)}(\theta; k) + c_2 \Lambda_1^{n,(-)}(\theta; 1) \right).$$

Notice that since $R_n$ is a sum of a finite number of asymptotically negligible terms, it is itself asymptotically negligible. Thus, it remains to inspect the behavior of $L_n^{(-)}$ and $L_n^{(+)}$. We do this in

detail for the former only by writing

$$L_n^{(-)} = \sum_{i=n\theta-k\ell_n+1}^{n\theta} \xi_i^n,$$

where

$$\xi_i^n = \frac{c_1}{2k\sqrt{\ell_n}}\big(\Delta_i\widetilde{N}^n - \mu_{(i-1)\Delta_n}\big) + \frac{c_2}{2\sqrt{\ell_n}}\big(\Delta_i\widetilde{N}^n - \mu_{(i-1)\Delta_n}\big)\mathbf{1}_{\{i\geq n\theta-\ell_n+1\}}.$$

It holds that $\mathbb{E}\left[\xi_i^n \mid \mathcal{F}_{(i-1)\Delta_n}\right] = 0$, and

$$\mathbb{E}\left[\left(\xi_i^2\right)^2 \mid \mathcal{F}_{(i-1)\Delta_n}\right] = \frac{\mu_{(i-1)\Delta_n}}{4\ell_n}\left[\frac{c_1^2}{k^2} + \left(c_2^2 + \frac{2c_1c_2}{k}\right)\mathbf{1}_{\{i\geq n\theta-\ell_n+1\}}\right].$$

Therefore,

$$\sum_{i=n\theta-k\ell_n+1}^{n\theta} \mathbb{E}\left[\left(\xi_i^2\right)^2 \mid \mathcal{F}_{(i-1)\Delta_n}\right] = c_1^2\frac{1}{4k}\sum_{i=n\theta-k\ell_n+1}^{n\theta}\frac{\mu_{(i-1)\Delta_n}}{k\ell_n} + \left(c_2^2 + \frac{2c_1c_2}{k}\right)\frac{1}{4}\sum_{i=n\theta-\ell_n+1}^{n\theta}\frac{\mu_{(i-1)\Delta_n}}{\ell_n}$$

$$\xrightarrow{p} c_1^2\frac{1}{4k}\mu_{\theta-} + c_2^2\frac{1}{4}\mu_{\theta-} + 2c_1c_2\frac{1}{4k}\mu_{\theta-}.$$

Furthermore,

$$\sum_{i=n\theta-k\ell_n+1}^{n\theta} \mathbb{E}\left[\left(\xi_i^n\right)^4 \mid \mathcal{F}_{(i-1)\Delta_n}\right] \leq C \sum_{i=n\theta-k\ell_n+1}^{n\theta}\frac{1}{\ell_n^2}\mathbb{E}\left[\left(\Delta_i\widetilde{N}^n - \mu_{(i-1)\Delta_n}\right)^4 \mid \mathcal{F}_{(i-1)\Delta_n}\right] \xrightarrow{p} 0,$$

where the convergence was already established in the proof of Theorem 1 (for the analogous term, which was denoted by $\zeta_i^n(3)$ back then). Thus, by Lemma 3

$$L_n^{(-)} \xrightarrow{\mathfrak{D}_s} \sqrt{c_1^2\frac{1}{4k}\mu_{\theta-} + c_2^2\frac{1}{4}\mu_{\theta-} + 2c_1c_2\frac{1}{4k}\mu_{\theta-}}\,\mathcal{Z}^{(-)},$$

where $\mathcal{Z}^{(-)} \sim N(0,1)$.

Analogously,

$$L_n^{(+)} \xrightarrow{\mathfrak{D}_s} \sqrt{c_1^2\frac{1}{4k}\mu_{\theta+} + c_2^2\frac{1}{4}\mu_{\theta+} + 2c_1c_2\frac{1}{4k}\mu_{\theta+}}\,\mathcal{Z}^{(+)},$$

where $\mathcal{Z}^{(+)} \sim N(0,1)$ independently from $\mathcal{Z}^{(-)}$. Since $\mu_{\theta+} = \mu_{\theta-} + \Delta\mu_\theta$:

$$\mathbf{L}_n \xrightarrow{\mathfrak{D}_s} \sqrt{c_1^2\frac{1}{2k}\left(\mu_{\theta-} + \frac{\Delta\mu_\theta}{2}\right) + c_2^2\frac{1}{2}\left(\mu_{\theta-} + \frac{\Delta\mu_\theta}{2}\right) + 2c_1c_2\frac{1}{2k}\left(\mu_{\theta-} + \frac{\Delta\mu_\theta}{2}\right)}\,\mathcal{Z}.$$

So the convergence in (A.13) holds with

$$\Sigma = \begin{bmatrix} \frac{1}{2k}\left(\mu_{\theta-} + \frac{\Delta\mu_\theta}{2}\right) & \frac{1}{2k}\left(\mu_{\theta-} + \frac{\Delta\mu_\theta}{2}\right) \\ \frac{1}{2k}\left(\mu_{\theta-} + \frac{\Delta\mu_\theta}{2}\right) & \frac{1}{2}\left(\mu_{\theta-} + \frac{\Delta\mu_\theta}{2}\right). \end{bmatrix}$$

$\blacksquare$

# References

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

Aït-Sahalia, Y., and J. Jacod, 2009, "Testing for jumps in a discretely observed process," *Annals of Statistics*, 37(1), 184–222.

Almgren, R. F., and N. Chriss, 2001, "Optimal execution of portfolio transactions," *Journal of Risk*, 3(2), 5–39.

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

Amihud, Y., 2002, "Illiquidity and stock returns: cross-section and time-series effects," *Journal of Financial Markets*, 5(1), 31–56.

Andersen, T. G., 1996, "Return volatility and trading volume: An information flow interpretation of stochastic volatility," *Journal of Finance*, 51(1), 169–204.

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.

———, 2012, "Jump-robust volatility estimation using nearest neighbour truncation," *Journal of Econometrics*, 169(1), 75–93.

Andersen, T. G., Y. Li, V. Todorov, and B. Zhou, 2023, "Volatility measurement with pockets of extreme return persistence," *Journal of Econometrics*, 237(2), 105048.

Bandi, F. M., A. Kolokolov, D. E. Pirino, and R. Renò, 2020, "Zeros," *Management Science*, 66(8), 3295–3798.

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.

Bollerslev, T., J. Li, and Y. Xue, 2018, "Volume, volatility, and public news announcements," *Review of Economic Studies*, 85(4), 2005–2041.

Bugni, F. A., J. Li, and Q. Li, 2023, "Permutation-based tests for discontinuities in event studies," *Quantitative Economics*, Forthcoming.

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.

Clark, P. K., 1973, "A subordinated stochastic process model with finite variance for speculative prices," *Econometrica*, 41(1), 135–155.

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

Comte, F., and E. Renault, 1998, "Long memory in continuous-time stochastic volatility models," *Mathematical Finance*, 8(4), 291–323.

Cox, D. R., 1955, "Some statistical methods connected with series of events," *Journal of the Royal Statistical Society: Series B*, 17(2), 129–157.

Cox, J. C., J. E. Ingersoll, and S. A. Ross, 1985, "A theory of the term structure of interest rates," *Econometrica*, 53(2), 385–407.

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.

Delbaen, F., and W. Schachermayer, 1994, "A general version of the fundamental theorem of asset pricing," *Mathematische Annalen*, 300(1), 463–520.

Duffie, D., 2005, "Credit risk modelling with affine processes," *Journal of Banking and Finance*, 29(11), 2751–2802.

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

Epps, T. W., and M. L. Epps, 1976, "The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distributions hypothesis," *Econometrica*, 44(2), 305–321.

Gatheral, J., T. Jaisson, and M. Rosenbaum, 2018, "Volatility is rough," *Quantitative Finance*, 18(6), 933–949.

Glosten, L. R., and P. R. Milgrom, 1985, "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders," *Journal of Financial Economics*, 14(1), 71–100.

Hansen, P. R., and A. Lunde, 2006, "Realized variance and market microstructure noise," *Journal of Business and Economic Statistics*, 24(2), 127–161.

Harris, M., and A. Raviv, 1993, "Differences of opinion make a horse race," *Review of Financial Studies*, 6(3), 473–506.

Hasbrouck, J., 1999, "The dynamics of discrete bid and ask quotes," *Journal of Finance*, 54(6), 2109–2142.

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.

Heston, S. L., 1993, "A closed-form solution for options with stochastic volatility with applications to bond and currency options," *Review of Financial Studies*, 6(2), 327–343.

Hong, H., and J. Wang, 2000, "Trading and returns under periodic market closure," *Journal of Finance*, 55(1), 297–354.

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.

Huang, J., and J. Wang, 2009, "Liquidity and market crashes," *Review of Financial Studies*, 22(7), 2607–2643.

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.

Kandel, E., and N. D. Pearson, 1995, "Differential interpretation of public signals and trade in speculative markets," *Journal of Political Economy*, 103(4), 831–872.

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.

Kyle, A. S., 1985, "Continuous auctions and insider trading," *Econometrica*, 53(6), 1315–1335.

Lacave, D., A. Ranaldo, and P. S. de Magistris, 2023, "Realized illiquidity," Working paper, Swiss Finance Institute.

Li, J., V. Todorov, and G. Tauchen, 2013, "Volatility occupation times," *Annals of Statistics*, 41(4), 1865–1891.

Li, J., and D. Xiu, 2016, "Generalized method of integrated moments for high-frequency data," *Econometrica*, 84(4), 1613–1633.

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

Næs, R., and J. A. Skjeltorp, 2006, "Order book characteristics and the volume-volatility relation: Empirical evidence from a limit order market," *Journal of Financial Markets*, 9(4), 408–432.

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.

Stoltenberg, E. A., P. A. Mykland, and L. Zhang, 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 M. Pitts, 1983, "The price variability-volume relationship on speculative markets," *Econometrica*, 51(2), 485–505.

Taylor, S. J., and X. Xu, 1997, "The incremental volatility information in one million foreign exchange quotations," *Journal of Empirical Finance*, 4(4), 317–340.

Thistlethwaite, D. L., and D. T. Campbell, 1960, "Regression-discontinuity analysis: An alternative to the ex post facto experiment," *Journal of Educational Psychology*, 51(6), 309–317.

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