ARTICLE IN PRESS

Journal of Econometrics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Testing for jumps in a discretely observed price process with endogenous sampling times☆

Qiyuan Li [a], Yifan Li [b], Ingmar Nolte [c], Sandra Nolte [c], Shifan Yu [d,*]

[a] Faculty of Business and Economics, University of Hong Kong, Hong Kong Special Administrative Region
[b] Alliance Manchester Business School, University of Manchester, United Kingdom
[c] Lancaster University Management School, United Kingdom
[d] Oxford-Man Institute of Quantitative Finance, University of Oxford, United Kingdom

ARTICLE INFO

JEL classification:
C12
C14
C22
C58

Keywords:
High-frequency data
Jump test
Market microstructure noise
Stochastic sampling scheme
First exit time

ABSTRACT

This paper introduces a novel nonparametric high-frequency jump test for discretely observed Itô semimartingales. Based on observations sampled recursively at first exit times from a symmetric double barrier, our method distinguishes between threshold exceedances caused by the Brownian component and jumps, which enables the construction of a feasible, noise-robust statistical test. Simulation results demonstrate superior finite-sample performance of our test compared to existing methods. An empirical analysis of NYSE-traded stocks provides clear statistical evidence for jumps, with the results highly robust to spurious detections.

1. Introduction

There exists a consensus in the financial literature that modeling asset price dynamics requires the specification of different components. In addition to the stochastic volatility component, which accounts for the persistence of volatility, "jumps" in asset prices serve as an explanation for abnormally large variations that play an important role for the tail behavior of return distributions. Jumps are believed to contain predictive information, so correctly identifying them often leads to improved price or volatility forecasts and better portfolio outcomes (see, e.g., Yan, 2011; Jiang and Yao, 2013; Cremers et al., 2015, for empirical applications using daily or monthly financial data, and Andersen et al., 2007a; Corsi et al., 2010; Nolte and Xu, 2015; Bollerslev et al., 2015, 2020; Pelger, 2020, for those using high-frequency intraday data). The increased availability of high-frequency financial data has further motivated the development of methodologies designed to test the model specification based on a discretely observed semimartingale.

Over the past two decades, a number of nonparametric jump tests have been developed. Starting from the seminal work of Barndorff-Nielsen and Shephard (2004), most of these tests are constructed on jump-robust measures of returns or their variations, see, e.g., Huang and Tauchen (2005), Barndorff-Nielsen and Shephard (2006), Andersen et al. (2007b), Jiang and Oomen (2008), Lee and Mykland (2008), Aït-Sahalia and Jacod (2009b), Corsi et al. (2010), Podolskij and Ziggel (2010), Andersen et al. (2012), Lee and Mykland (2012), and Aït-Sahalia et al. (2012), among others. Some recent works focus on modified versions of these tests when

☆ This article is part of a Special issue entitled: 'Themed Issue: High Frequency Econometrics' published in Journal of Econometrics.

* Corresponding author.
E-mail addresses: qiyuanli@hku.hk (Q. Li), yifan.li@manchester.ac.uk (Y. Li), i.nolte@lancaster.ac.uk (I. Nolte), s.nolte@lancaster.ac.uk (S. Nolte), shifan.yu@eng.ox.ac.uk (S. Yu).

https://doi.org/10.1016/j.jeconom.2025.106132
Received 7 October 2022; Received in revised form 8 September 2025; Accepted 1 November 2025

Please cite this article as: Q. Li, Y. Li, I. Nolte et al., Testing for jumps in a discretely observed price process with endogenous sampling times. Journal of Econometrics (2025) 106132, https://doi.org/10.1016/j.jeconom.2025.106132.

conventional assumptions are violated, see, e.g., Laurent and Shi (2020) and Kolokolov and Renò (2024), and tests for co-jumps in a collection of assets, see, e.g., Bibinger and Winkelmann (2015) and Caporin et al. (2017).

Despite the theoretical developments in the literature, these jump tests can sometimes deliver inconsistent results in practice. Unlike the noiseless theoretical framework, the presence of market microstructure noise in real-world high-frequency data requires practitioners to either sample sparsely or pre-average the tick-level data. Although the literature establishes asymptotic consistency of calendar-grid tests under both the null and alternative hypotheses, their finite-sample performance deteriorates markedly at commonly used sampling frequencies in practice (Dumitru and Urga, 2012; Maneesoonthorn et al., 2020). Under sparse, exogenous and typically equidistant sampling, many tick-level returns are bundled into one sampling interval, which dilutes the relative contribution of jumps and thus reduces finite-sample power of those tests. This practical tension motivates a key question: Can we sample in a way that retains more jump information than equidistant calendar-time sampling?

In this paper, we introduce an innovative nonparametric method to test for jumps in a discretely observed semimartingale based on endogenous sampling. Different from the conventional equidistant calendar-time sampling, our methodology adopts a stochastic and endogenous approach that recursively samples tick-by-tick observations at first exit times from a symmetric double barrier, inspired by Engle and Russell (1998), Andersen et al. (2008), Fukasawa and Rosenbaum (2012), Vetter and Zwingmann (2017), and Hong et al. (2023), among others. This endogenous sampling scheme is tailored to be sensitive to jumps. Jumps of size larger than the barrier width will terminate the sampling interval immediately and produce large "overshoots", i.e., threshold exceedances. By stopping the sampling interval when the barrier is breached, the sampled return captures the jump in a clean way, without mixing it with post-jump diffusion increments, which amplifies the jump signal. To distinguish between threshold exceedances caused by discrete Brownian steps and those by jumps, we censor the returns between consecutive sampling times with a specific threshold,[1] and construct a standardized test statistic to measure the potential distortion caused by disproportionately large overshoots in the sample moment of returns. While our approach is in the spirit of the standard truncation technique of Mancini (2009), the objects are first-exit ladder increments rather than Brownian increments over equidistant calendar-time intervals, so the relevant limit theorems differ and require new asymptotic results under the first-exit framework. Furthermore, we develop a two-step noise reduction method based on the pre-averaging approach of Jacod et al. (2009) and the wild bootstrap to mitigate the impact of weakly dependent market microstructure noise in the pre-sampling tick-level observations, which helps to bridge the gap between real-world features and the theoretical framework.

Simulation results reveal that our new high-frequency jump test exhibits reliable finite-sample size and power performance across various aggregation levels, and its performance is robust to measurement errors simulated with a realistically calibrated specification. A comparison with commonly used tests constructed from equidistantly sampled observations and some noise-robust versions based on ultra-high-frequency data is conducted thereafter. We find that (i) most calendar-time-sampled tests exhibit less consistent performance across different sampling frequencies and are poorly sized in the presence of noise, which is in line with the Monte Carlo results of Dumitru and Urga (2012) and Maneesoonthorn et al. (2020), (ii) while noise-robust tests achieve reliable sizes in the presence of noise, their power performance is still inferior to our test across a wide range of simulation settings, and (iii) the truncation-based jump filtering and detection techniques commonly applied in the recent literature suffer from spurious detections and become unreliable when noise is substantial at high frequencies, which echoes the findings in Aït-Sahalia et al. (2025). In an empirical application, our test is applied to transaction data of 10 selected stocks listed on the New York Stock Exchange (NYSE). Clear statistical evidence of jumps is found for all selected stocks, with jumps occurring on approximately 10% to 15% of trading days. Furthermore, the test rejections are highly robust to the correction of spurious detections based on the method of Bajgrowicz et al. (2016).

The remainder of this paper is structured as follows: Section 2 lays out the basic setup and key assumptions. Section 3 discusses the test statistic and its asymptotic theory, along with the noise reduction technique. Section 4 assesses the finite-sample performance of our new test with Monte Carlo simulations. After discussing the empirical application for selected NYSE stocks in Section 5, we conclude in Section 6. All proofs and additional simulation and empirical results are relegated to the Online Appendix.

2. Setting and assumptions

On a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, let the one-dimensional process $X = (X_t)_{t\geq 0}$ denote the efficient logarithmic price of a financial asset. We assume that X follows a possibly discontinuous Itô semimartingale of the following form:

$$\begin{aligned} X &= X' + X'', \\ X'_t &= X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s, \\ X''_t &= \int_0^t \int_{\mathbb{R}} \delta(s,x) \mathbb{1}_{\{|\delta(s,x)|\leq 1\}} (\underline{p} - \underline{q})(ds, dx) + \int_0^t \int_{\mathbb{R}} \delta(s,x) \mathbb{1}_{\{|\delta(s,x)|>1\}} \underline{p}(ds, dx), \end{aligned} \tag{1}$$

where t stands for time, W is a standard Brownian motion, $\underline{p}(dt, dx)$ is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}$ with a compensator $\underline{q}(dt, dx) = dt \otimes \lambda(dx)$, and λ is a σ-finite measure on $\mathbb{R}_+$. We assume that X satisfies the following regularity conditions:

[1] Related works about the boundary crossing problems for random walks, especially those with Gaussian steps, include Rogozin (1964), Lorden (1970), Lotov (1996), and Khaniyev and Kucuk (2004).

Assumption 1. The following properties hold for the processes in Eq. (1):

(i) The process μ is optional and locally bounded;
(ii) The process σ is càdlàg (i.e., right-continuous with left limits), adapted, and strictly positive;
(iii) There exists a sequence $(\tau_m)_{m\geq 1}$ of stopping times increasing to ∞, and a sequence $(K_m)_{m\geq 1}$ of finite constants, such that it holds for each $m \geq 1$ that $\mathbb{E}[|\sigma_{t\wedge\tau_m} - \sigma_{s\wedge\tau_m}|^2] \leq K_m|t-s|$ for all $s,t \in [0,T]$ with some finite T;
(iv) The function $\delta(\omega,t,x)$ on $\Omega\times\mathbb{R}_+\times\mathbb{R}$ is predictable;
(v) There is a localizing sequence $(\tau_n)_{n\geq 1}$ of stopping times increasing to ∞, and a sequence $(f_n)_{n\geq 1}$ of deterministic nonnegative functions on $\mathbb{R}$, which satisfies $|\delta(\omega,t,x)| \wedge 1 \leq f_n(x)$ for all (ω,t,x) with $t \leq \tau_n(\omega)$, and $\int_{\mathbb{R}} |f_n|^r \lambda(dx) < \infty$ for some $r \in [0,1)$.

Remark 1. Assumption 1 entails some very mild technical conditions that the processes in Eq. (1) should meet. Conditions (i), (ii) and (iv) are standard in the literature. Condition (iii) states that the spot volatility process is locally 1/2-Hölder continuous under the L_2-norm. The smoothness condition is satisfied whenever σ is an Itô semimartingale, or a long-memory process driven by a fractional Brownian motion (Li and Liu, 2021). The parameter r in Condition (v) sets a bound on the degree of jump activity, which can be interpreted as a generalized version of the Blumenthal–Getoor index for a Lévy process (Aït-Sahalia and Jacod, 2009a; Jing et al., 2012). With some $r \in [0,1)$, we consider jumps of both finite and infinite activities, but restrict them to be of finite variation, i.e., they are absolutely summable, such that in Eq. (1) we dispense with the integral with $\underline{p} - \underline{q}$; see Jacod et al. (2019) for more details.

The quadratic variation (QV) of X over a finite interval $[0,t]$ is defined as

$$\langle X,X\rangle_t = \int_0^t \sigma_s^2 ds + \sum_{0\leq s\leq t} (\Delta X_s)^2, \quad \text{with } \Delta X_t = X_t - X_{t-}, \tag{2}$$

where the integrated variance (IV), $\int_0^t \sigma_s^2 ds$, summarizes the variation from X'.

Testing for jumps is a procedure to answer the fundamental question of whether the realized sample path $X(\omega)$ is continuous or not over a finite time interval, e.g., $(0,1)$.[2] Technically speaking, we decompose the sample space Ω into two complementary subsets:

$$\begin{aligned} \Omega' &= \{\omega : X_t(\omega) \text{ is continuous on } (0,1)\}, \\ \Omega'' &= \{\omega : X_t(\omega) \text{ is discontinuous on } (0,1)\}, \end{aligned} \tag{3}$$

where Ω' (resp. Ω'') represents the null hypothesis (alternative hypothesis) for a jump test, which assesses the plausibility of these two hypotheses based on discrete observations of $X(\omega)$.

2.1. Observation scheme

We now describe how observations take place.[3] At stage n, we assume that the successive observations of $X(\omega)$ occur at times $0 = t_{n,0} < t_{n,1} < \dots$ for a sequence $(t_{n,i})$ of discrete times over a fixed interval (such as a trading day), which is normalized to the unit interval $[0,1]$. We set

$$N_t^n = \sum_{i\geq 1} \mathbb{1}_{\{t_{n,i}\leq t\}} \qquad \text{and} \qquad \Delta_{n,i} = t_{n,i} - t_{n,i-1}, \tag{4}$$

where $N \equiv N_1^n$ stands for the number of observations on $(0,1]$, and $\Delta_{n,i}$ is the i-th inter-observation lag at stage n. It is easily seen from the empirical tick-level data that the observation times are far from evenly spaced and usually dependent on $X(\omega)$ itself. Our assumption for the observation scheme over $[0,1]$ is outlined as follows:

Assumption 2. Let Δ_n be a positive sequence of real numbers satisfying $\Delta_n \to 0$ as $n \to \infty$. We define an intensity process of observations $\lambda = (\lambda_t)_{0\leq t\leq 1}$ with $\lambda_t = K\sigma_t^2$ for some $K > 0$. There exists a localizing sequence $(\tau_m)_{m\geq 1}$ of stopping times and positive constants $K_{m,p}$ and κ such that:

(i) With $(\mathcal{F}_t^n)_{t\geq 0}$ the smallest filtration containing $(\mathcal{F}_t)_{t\geq 0}$ and with respect to which all observation times $t_{n,i}$ are stopping times, for each $i = 1,2,\dots$, the variable $\Delta_{n,i}$ is, conditionally on $\mathcal{F}_{i-1}^n \equiv \mathcal{F}_{t_{n,i-1}}^n$, independent of $\mathcal{F}_\infty = \bigvee_{t\geq 0} \mathcal{F}_t$.
(ii) With the restriction $\{t_{i-1} < \tau_m\}$, we have for all $p \geq 2$,

$$\begin{aligned} \mathbb{E}[|\Delta_{n,i}\lambda_{t_{n,i-1}} - \Delta_n\|\mathcal{F}_{i-1}^n] &\leq K_{m,1}\Delta_n^{2+\kappa}, \\ \mathbb{E}[|\Delta_{n,i}\lambda_{t_{n,i-1}}|^p|\mathcal{F}_{i-1}^n] &\leq K_{m,p}\Delta_n^p. \end{aligned} \tag{5}$$

[2] We restrict the alternative hypothesis to contain at least one jump on $(0,1)$ as it is not feasible for a test to identify jumps occurring right at both endpoints of the interval.

[3] We would like to distinguish the terms "observation scheme" and "sampling scheme" in this paper. We allow both tick-level and sampled observations to form discrete-time processes, and the term "sampling" refers to a subsampling or subset selection procedure for the discrete observations at the highest frequency.

A useful consequence of this Assumption is the following convergence in probability:

$$\Delta_n N_t^n \xrightarrow{\mathbb{P}} \tau(t) = \int_0^t \sigma_s^2 ds. \tag{6}$$

Remark 2. Assumption 2 is inspired by Assumption (O) of Jacod et al. (2017) and Assumption (O-ρ, ρ') of Jacod et al. (2019). The process λ controls for the "spot" observation arrival rates, and the unobserved Δ_n can be interpreted as an "average mesh size" between successive observations in an alternative time scale that evolves with volatility (see Remark 3). Our choice of the intensity $\lambda = K\sigma^2$ implies higher observation frequencies of $X(\omega)$ during periods of high local volatility, which captures the diurnal patterns of transaction activities and intraday volatility. This is motivated by the empirical evidence of the E-mini S&P 500 futures contract in Andersen et al. (2018), which illustrates a notable similarity in the intraday U-shaped patterns of one-minute transaction counts and return variation, where the pronounced spikes, typically align with market openings or announcements, roughly coincide. Note that λ is defined up to scale, which allows $K = 1$ to be set without loss of generality (by scaling Δ_n correspondingly), as further discussed in Jacod et al. (2017).

Remark 3. With the convergence result in Eq. (6), Assumption 2 implies a time-changed regular observation scheme under infill asymptotics: As $n \to \infty$, the observation time $t_{n,i}$ converges to $\check{t}_{n,i} = \inf\{t \in [0,1] : \tau(t) = i\Delta_n\}$. This limiting observation scheme corresponds to Example 2.2 in Jacod et al. (2017). In contrast to the calendar time t, the "intrinsic time" $\tau(t)$ evolves endogenously with respect to the variation from X'. The time change induces a certain level of endogeneity, and extends the commonly assumed equidistant observation scheme in high-frequency financial econometrics literature (Li et al., 2014; Dimitriadis and Halbleib, 2022). With the irregular calendar-time mesh sizes $\Delta_{n,i}$ regulated by Condition (ii), the deviation of $(t_{n,i})$ from $(\check{t}_{n,i})$ vanishes as $n \to \infty$. Importantly, this discrepancy does not affect the limit theorems developed in the next section, a conclusion supported by strong approximation results in the spirit of Chernozhukov et al. (2013, 2019). Further details can be found in Remark 6 and Online Appendix A.1.

2.2. Price duration sampling

Sparse sampling is widely adopted in both the financial econometrics literature and by practitioners to mitigate the impact of market microstructure noise, with some popular choices like 1-minute and 5-minute sampling in calendar time (Aït-Sahalia et al., 2005; Liu et al., 2015). However, such sparse sampling aggregates a substantial amount of tick-level returns exogenously, which dilutes the relative size of jumps and inevitably reduces the power of jump tests. This phenomenon is evident in the Monte Carlo results of Dumitru and Urga (2012) and Maneesoonthorn et al. (2020): Nearly all traditional tests constructed from calendar-time-sampled returns exhibit rapid power loss as sampling becomes sparser.[4]

In response to this issue, a path-dependent sampling scheme seems a natural solution. We consider a stochastic and endogenous sampling scheme for all observations of $X(\omega)$ on $[0,1]$: Let $(X_i)_{0\le i\le N}$ collect all observations under Assumption 2. With a selected barrier width $c > 0$, the price duration sampling (PDS) is defined as the following sampling algorithm:

1. Set $\Pi_0^{(c)} = 0$.
2. For $j = 1, 2, \dots$, sample X_i for all $i = \Pi_j^{(c)}$ that are decided recursively by

$$\Pi_j^{(c)} = \inf\left\{ \Pi_{j-1}^{(c)} < i \le N : \left| X_i - X_{\Pi_{j-1}^{(c)}} \right| \ge c \right\}. \tag{7}$$

We therefore obtain a subsequence $X^{(c)} = (X_{\Pi_j^{(c)}})_{0\le j\le N^{(c)}}$, where $N^{(c)} = \max_{j\ge 1}\{\Pi_j^{(c)} \le N\}$ counts the total number of sampled observations. Moreover, we define the PDS returns as the increments of $X^{(c)}$, i.e., $r_j^{(c)} = X_{\Pi_j^{(c)}} - X_{\Pi_{j-1}^{(c)}}$ for all $j \in \{1, 2, \dots, N^{(c)}\}$.

Remark 4. The above sampling algorithm is a discrete-time version of PDS in Hong et al. (2023). The idea of sampling financial observations based on hitting or exit times was initially proposed by Engle and Russell (1998), and has been further developed since then, see, e.g., Gerhard and Hautsch (2002), Andersen et al. (2008), Tse and Yang (2012), Fukasawa and Rosenbaum (2012), Potiron and Mykland (2017), Vetter and Zwingmann (2017), and Hong et al. (2023). While previous studies have primarily focused on volatility estimation based on this alternative sampling scheme, our contribution stands out as the first to demonstrate that this scheme can be exploited to construct more effective high-frequency jump tests.

From a technical standpoint, our design belongs to the broader class of stochastic and endogenous sampling in high frequency, see, e.g., Fukasawa (2010), Li et al. (2014), and Koike (2017). We work with a symmetric first-exit scheme on the intrinsic-time clock, a tractable special case under which asymptotic properties for realized variance (RV) estimators follow directly from Fukasawa (2010). However, further theoretical developments are needed for a jump test based on this sampling scheme; see Remark 6 for further details.

[4] Some noise-robust tests constructed from filtered data, such as those proposed by Lee and Mykland (2012) and Aït-Sahalia et al. (2012), can utilize all available observations without sampling. As alternative methods that exploit data more sufficiently than classical approaches, we compare their finite-sample performance with our method through simulations in Section 4.

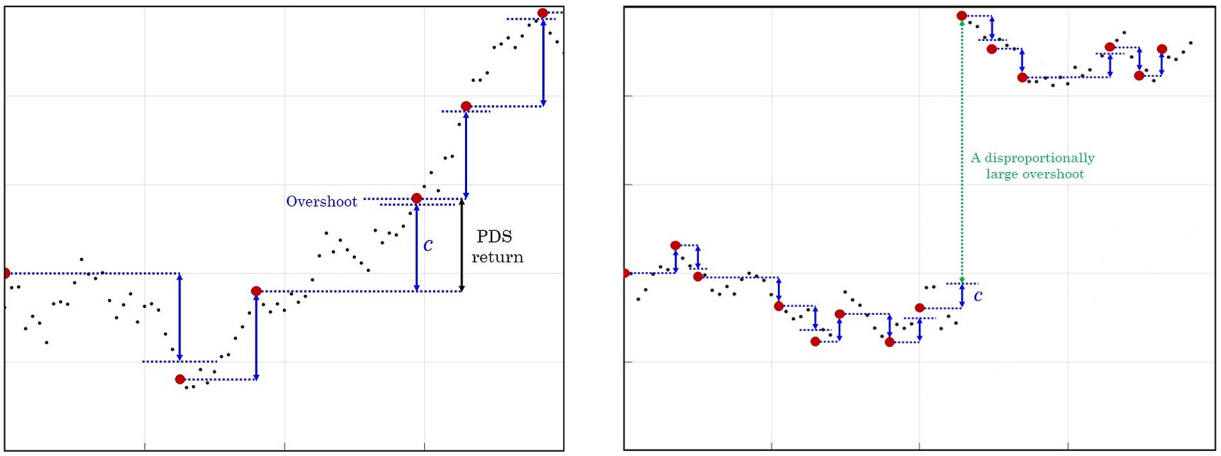

Fig. 1. Examples of PDS when $X(\omega)$ is continuous and discontinuous, respectively. Jumps will almost surely lead to the sampling of the next available observation, and induce a large overshoot.

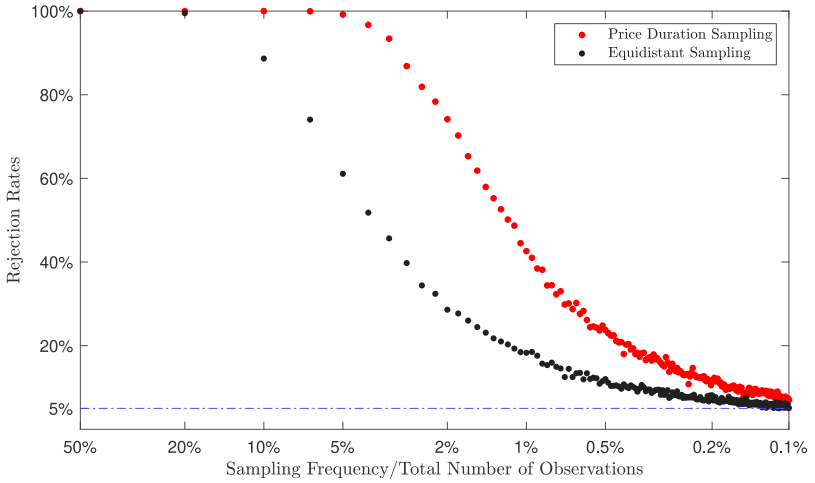

Fig. 2. Rejection rates under two different sampling schemes. We simulate 2000 random walk paths with 10^6 standard normal steps (null). A jump of fixed size 10 is randomly inserted in each path. Under two different sampling schemes, the absolute returns containing jumps are compared with the 95% quantiles of absolute sampled returns under the null.

This endogenous sampling scheme is designed to be highly sensitive to the presence of jumps. Fig. 1 shows examples where $X(\omega)$ is continuous and discontinuous, respectively. When $X(\omega)$ is continuous, each sampled return under PDS ("PDS return", i.e., first ladder height with respect to c) consists of the barrier width c plus a small exceedance, i.e., the extra movement needed before the next discrete observation time is stamped. By contrast, if $X(\omega)$ is discontinuous, a jump with magnitude $> c$ triggers the stopping rule immediately and produces an "overshoot" that is visibly larger than the continuous returns.

To motivate the use of PDS over exogenous sampling schemes in testing for jumps, we highlight the intuition that, over a given sampling interval, the price return containing a jump is more easily identified when the jump size is large relative to the aggregated continuous price increments in that interval. This relative magnitude can be naturally interpreted as a "signal-to-noise" ratio for jump tests. With a simple motivating Monte Carlo example, we demonstrate that the PDS consistently generates a higher signal-to-noise ratio than equidistant sampling in finite samples under the same sampling frequency. In each replication, we simulate a Gaussian random walk with a fixed number of i.i.d. increments (corresponding to the limiting observation scheme in Remark 3). Under the alternative, we insert one fixed-size jump at a uniformly chosen time index. For each simulated path we then obtain sampled returns with both PDS and equidistant sampling across a broad range of sampling frequencies. A (PDS or equidistantly) sampled return is labeled a jump whenever its absolute value exceeds the 95% quantile of the corresponding null distribution, which by construction fixes the empirical size at 5% for every sampling frequency. In our simulation, both the PDS barrier width and the equidistant interval length are calibrated such that both sampling schemes produce the same expected number of observations. Therefore, any discrepancy in rejection rates under the alternative can be attributed to the effective signal-to-noise ratio achieved by the respective sampling method.

Fig. 2 illustrates the rejection rates under both sampling schemes — under the alternative that the tested interval contains a jump — across a continuum of expected sampling frequencies. For both sampling methods, the rejection rates decline toward the nominal 5% size as sampling becomes increasingly sparse. This is because each sampled return aggregates the jump with an increasing number of Gaussian increments, thereby deteriorating the signal-to-noise ratio. However, the speed of this convergence differs substantially between the two schemes. Intuitively, as the PDS samples the price observations whenever a jump-induced overshoot occurs, the price increments after the jump are excluded from that sampling interval, thus inflating the signal-to-noise ratio. By contrast, equidistant sampling aggregates returns exogenously at fixed intervals, where the jump size is diluted much more rapidly as the sampling interval lengthens. This advantageous property of PDS contributes to a diminished probability of committing a Type II error, and thereby serves as the main motivation for the new statistical test proposed in the next section.

We now formally introduce our asymptotic setting under PDS: We let the barrier width c shrink proportionally to $\sqrt{\Delta_n}$ under infill asymptotics, i.e.,

$$c \equiv c_n = m\sqrt{\Delta_n}, \quad \text{for some constant } m > 0. \tag{8}$$

When $X(\omega)$ is continuous, each absolute PDS return $|r_i^{(c)}|$ is a sum of the barrier width c and a small threshold exceedance caused by the discreteness of observations, such that the ratio $|r_i^{(c)}|/c$ is bounded in probability. In contrast, jumps of a higher asymptotic order than $\sqrt{\Delta_n}$ will almost surely trigger the stopping rule in Eq. (7), and generate disproportionally large PDS returns for which $|r_i^{(c)}|/c$ diverges in the limit.

To distinguish between the "small" overshoots induced by continuous price increments and the "big" overshoots caused by genuine jumps, we censor the (absolute) PDS returns with a threshold $\varphi_\epsilon(c)$ that shrinks to zero at the same rate $\sqrt{\Delta_n}$ as the barrier width c, i.e., for all $i \in \{1, 2, \dots, N^{(c)}\}$,

$$|\overline{r}_i^{(c)}| = |r_i^{(c)}| \wedge \varphi_\epsilon(c), \quad \text{where } \varphi_\epsilon(c) = c\,(1+\epsilon) \text{ for some constant } \epsilon > 0. \tag{9}$$

Remark 5. The idea of censored returns is inspired by the standard truncation techniques of Mancini (2009). However, the key difference is that the sampled returns we work with are first-exit ladder increments, rather than Brownian increments over equidistant calendar-time intervals, for which Mancini's threshold is calibrated with local volatility estimates and justified by the Lévy modulus of continuity. Therefore, the relevant limit theorems are different and require additional theoretical development in the first-exit framework. Furthermore, the fixed choice of ϵ in Eq. (9) is unconventional in the literature. Unlike the standard calendar-time threshold of order slightly higher than $\sqrt{\Delta_n}$, our chosen threshold $\varphi_\epsilon(c) \asymp \sqrt{\Delta_n}$ affects increments from both X' and X'' under infill asymptotics. This circumvents the "perfect correlation" issue of censored and uncensored returns under the null (Podolskij and Ziggel, 2010), and allows the construction of feasible test statistics.[5]

3. Main results

In this section, we introduce and analyze our new test statistic, which is constructed from the PDS returns between sampled observations collected by $X^{(c)}$. Then we augment the test with an effective noise reduction method to mitigate the impact of market microstructure noise.

3.1. Test statistic

To prepare for the construction of our test statistic, we first introduce the notation for the moments of PDS returns from a standard Gaussian random walk $(Z_i)_{i=0,1,\dots}$ with a barrier width m, which is denoted as $Z_1^{(m)}$:

(i) Absolute moment of $Z_1^{(m)}$: $\mu_\gamma(m) = \mathbb{E}[|Z_1^{(m)}|^\gamma]$,
(ii) Absolute moment of censored $Z_1^{(m)}$: $\overline{\mu}_{\gamma,\epsilon}(m) = \mathbb{E}[|\overline{Z}_1^{(m)}|^\gamma] = \mathbb{E}[(|Z_1^{(m)}| \wedge \varphi_\epsilon(m))^\gamma]$,
(iii) Absolute cross moment of censored and uncensored $Z_1^{(m)}$: $\overline{\rho}_{\gamma,\epsilon}(m) = \mathbb{E}[|Z_1^{(m)}|^\gamma |\overline{Z}_1^{(m)}|^\gamma]$,

and two first-order differentiable and invertible functions:

$$h_2(m) = \frac{\mu_2(m)}{m^2} \quad \text{and} \quad \overline{h}_{2,\epsilon}(m) = \frac{\overline{\mu}_{2,\epsilon}(m)}{m^2}, \tag{10}$$

with the first-order derivatives $h_2'(m)$ and $\overline{h}'_{2,\epsilon}(m)$, and the inverse functions $h_2^{-1}(x)$ and $\overline{h}^{-1}_{2,\epsilon}(x)$.[6]

[5] In this paper, we can also adopt the truncation technique and discard all absolute PDS returns that are larger than $\varphi_\epsilon(c)$. However, the censoring approach does not change the total number of PDS returns and is therefore more convenient for both our theoretical derivation and empirical implementation.

[6] The invertibility and differentiability of $h_2(m)$ and $\overline{h}_{2,\epsilon}(m)$ are proved in Online Appendix A.2. In practice, we compute the required moments and functionals of $Z_1^{(m)}$ by simulation of standard Gaussian random walks, and obtain numerically the inverses and first-order derivatives of $h_2(m)$ and $\overline{h}_{2,\epsilon}(m)$ with local polynomial interpolation and local linear regression, respectively.

We will now proceed to define the testing procedures. For all observations $(X_i)_{0\le i\le N}$ under Assumption 2, we obtain the sampled observations in $X^{(c)}$ with the barrier width c that satisfies Eq. (8). To assess the distortion resulting from "large" overshoots, we compare two estimates from (normalized) sample moments of uncensored and censored PDS returns, respectively, i.e.,

$$M_c = h_2^{-1}\left(S_2\right) \qquad \text{and} \qquad \overline{M}_{c,\epsilon} = \overline{h}_{2,\epsilon}^{-1}\left(\overline{S}_{2,\epsilon}\right). \tag{11}$$

where

$$S_2 = \frac{1}{N^{(c)}}\sum_{i=1}^{N^{(c)}}\left(\frac{|r_i^{(c)}|}{c}\right)^2 \quad \text{and} \quad \overline{S}_{2,\epsilon} = \frac{1}{N^{(c)}}\sum_{i=1}^{N^{(c)}}\left(\frac{|\overline{r}_i^{(c)}|}{c}\right)^2. \tag{12}$$

Theorem 1 (*Consistency*). *Under Assumptions 1 and 2, it holds as $n\to\infty$ that*

$$\begin{aligned}(\overline{M}_{c,\epsilon}, M_c)' &\xrightarrow{\mathbb{P}} (m,m)', \qquad \text{if } \omega\in\Omega',\\ (\overline{M}_{c,\epsilon}, M_c)' &\xrightarrow{\mathbb{P}} (m,m^*)', \qquad \text{if } \omega\in\Omega'',\end{aligned} \tag{13}$$

where $m^* = h_2^{-1}(\kappa\cdot h_2(m))$ with κ the ratio between QV and IV over $[0,1]$.

Both $\overline{M}_{c,\epsilon}$ and M_c are jointly asymptotically normal under the null with a known variance–covariance matrix, which naturally leads to a well-defined ratio test.

Theorem 2 (*Asymptotic normality*). *Under the same conditions as in Theorem 1, the estimators $\overline{M}_{c,\epsilon}$ and M_c are jointly normally distributed when $\omega\in\Omega'$:*

$$\sqrt{N}\begin{pmatrix}\overline{M}_{c,\epsilon}-m\\ M_c-m\end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\phi_{11}(m) & \bullet\\ \phi_{21}(m) & \phi_{22}(m)\end{pmatrix}\right), \tag{14}$$

where

$$\phi_{11}(m) = \frac{\mu_2(m)(\overline{\mu}_{4,\epsilon}(m)-\overline{\mu}_{2,\epsilon}^2(m))}{m^4(\overline{h}'_{2,\epsilon}(m))^2}, \tag{15}$$

$$\phi_{22}(m) = \frac{\mu_2(m)(\mu_4(m)-\mu_2^2(m))}{m^4(h'_2(m))^2}, \tag{16}$$

$$\phi_{21}(m) = \frac{\mu_2(m)(\overline{\rho}_{2,\epsilon}(m)-\mu_2(m)\overline{\mu}_{2,\epsilon}(m))}{m^4(h'_2(m)\overline{h}'_{2,\epsilon}(m))^2}. \tag{17}$$

Remark 6. Based on a strong approximation argument of Chernozhukov et al. (2013, 2019), we couple the complicated observation scheme under Assumption 2 with the much simpler limiting observation scheme $(\check{\iota}_{n,i})$ in Remark 3, and the PDS returns $(r_i^{(c)})_{1\le i\le N^{(c)}}$ are strongly approximated by the sampled returns from a corresponding homogeneous Gaussian random walk in intrinsic time; see Online Appendix A.1.3 for details. The CLT in Theorem 2 is obtained from a joint convergence of the PDS-based RV, its censored version, and the sum of squared sampling thresholds. While the marginal stable CLT of the PDS-based RV can also be derived from Fukasawa (2010), the joint convergence requires additional non-trivial theoretical results, which are new in the literature; see Online Appendix A.4 for details.

Moreover, under the alternative, $\overline{M}_{c,\epsilon}$ remains robust to jumps because $\varphi_\epsilon(c)$ shrinks at the same rate as $\sqrt{\Delta_n}$, so censoring caps jump-induced overshoots. By contrast, M_c is inflated by these overshoots and converges to a different level. The resulting separation implies that the standardized test statistic diverges, as shown in the following Corollary 1.

Theorems 1 and 2 directly imply the following result, which indicates that our test is correctly sized under the null and consistent under the alternative:

Corollary 1. *Under the same conditions, the standardized ratio test statistic $T_{c,\epsilon}$ satisfies*

$$T_{c,\epsilon} = \frac{\overline{M}_{c,\epsilon}/M_c - 1}{\sqrt{\widehat{V}_\epsilon(\overline{M}_{c,\epsilon})}} \begin{cases}\xrightarrow{\mathcal{L}} \mathcal{N}(0,1) & \text{if } \omega\in\Omega',\\ \xrightarrow{\mathbb{P}} \infty & \text{if } \omega\in\Omega'',\end{cases} \tag{18}$$

where the denominator is the estimated standard deviation of $\overline{M}_{c,\epsilon}/M_c$ with

$$\widehat{V}_\epsilon(m) = \frac{1}{m^2 N}(\phi_{11}(m)+\phi_{22}(m)-2\phi_{21}(m)). \tag{19}$$

When $X''(\omega)\equiv 0$ on the interval $(0,1)$, the test statistic $T_{c,\epsilon}$ converges in distribution to a standard normal random variable, which is implied by Theorem 2. When $X''(\omega)\neq 0$ for some $t\in(0,1)$, the numerator of $T_{c,\epsilon}$ converges to a finite non-zero level determined by κ, whereas its denominator shrinks to zero as $n\to\infty$. Consequently, the standardized test statistic diverges in the limit, thereby implying the consistency of the test under the alternative hypothesis.

3.2. Noise mitigation

As discussed in Remark 6, our asymptotic results derived in Section 3.1 are based on the strong approximation result that couples the observation scheme $(t_{n,i})$ under Assumption 2 and the limiting scheme $(\check{t}_{n,i})$ in Remark 3. However, this rationale becomes untenable when the observations are contaminated by measurement errors, such as market microstructure noise. In this section, we propose an empirically plausible approach to mitigate the impact of the noise. With a two-step noise reduction method, we transform the noise-contaminated observations into a sequence of pseudo-observations, which behaves locally like a Gaussian random walk in the limit. Since each sampled return is only determined by finitely many tick-level returns within a local horizon, our test statistic relying solely on the sample moments of normalized PDS returns remains valid.

To this end, we assume an additive noise term with a weak dependence structure, before which we recall the definition of α-mixing (Fan and Yao, 2003): The α-mixing coefficient of a stationary sequence $(X_i)_{i\in\mathbb{Z}}$ of variables indexed by $i \in \mathbb{Z}$ is defined as

$$\alpha(h) = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{F}_i,\ B \in \mathcal{F}^{i+h}\}, \tag{20}$$

where the pre- and post-σ-fields are defined as $\mathcal{F}_j = \sigma(\{X_i : i \le j\})$ and $\mathcal{F}^j = \sigma(\{X_i : i \ge j\})$. The process (X_i) is said to be α-mixing if $\alpha(h) \to 0$ as $h \to \infty$.

Assumption 3. Let $\varepsilon = (\varepsilon_i)_{0\le i\le N}$ be a stationary sequence with $\mathbb{E}[\varepsilon_i] = 0$ and $\mathbb{E}[|\varepsilon_i|^{2+\delta}] < \infty$ for some $\delta > 0$, where ε_i are identically distributed with variance σ_ε^2 and autocovariance function $\Gamma_h = \mathbb{E}[\varepsilon_i\varepsilon_{i+h}]$. The process ε is α-mixing with $\sum_{h=1}^{\infty}\alpha(h)^{\delta/(2+\delta)} < \infty$, and exogenous to X. The sequence $Y = (Y_i)_{0\le i\le N}$ collects all observations contaminated by noise $Y_i = X_i + \varepsilon_i$, with the log-returns $r_i = Y_i - Y_{i-1}$ for all $1 \le i \le N$.

Remark 7. The autocovariance function Γ_h satisfies $\Gamma_0 = \sigma_\varepsilon^2$ and $\Gamma_{-h} = \Gamma_h$. For Γ_h, the standard absolute summability condition, i.e., $\sum_{h\in\mathbb{Z}}|\Gamma_h| < \infty$, is well-known to be sufficient for ergodicity and necessary for α-mixing under stationarity (Ibragimov and Linnik, 1971). Furthermore, the assumed conditions on the $(2+\delta)$th moment and the α-mixing coefficient $\alpha(h)$ are sufficient for a CLT for the centered, stationary and α-mixing ε (Theorem 1.7, Ibragimov, 1962; Theorem 8.3.7, Durrett, 2019).

Remark 8. The additive noise term ε_i summarizes a diverse array of market frictions. An i.i.d. additive noise with non-zero variance, firstly introduced by Zhou (1996), is commonly assumed in earlier literature of high-frequency volatility estimation, see, e.g., Aït-Sahalia et al. (2005) and Zhang et al. (2005). However, some previous studies including Hansen and Lunde (2006), Ubukata and Oya (2009), and Aït-Sahalia et al. (2011) find empirical evidence of self-dependent noise in financial markets. Recent work by Jacod et al. (2017) summarizes the common statistical properties of market microstructure noise and develops estimators for its autocovariances and autocorrelations, which further confirms this point. Assumption 3 allows for a weak dependence structure of the noise. This standard Itô semimartingale plus locally dependent noise framework has been employed by a number of recent studies, see, e.g., Jacod et al. (2017, 2019), Varneskov (2017), Christensen et al. (2022), and Li and Linton (2022).

However, it is worth noting that Assumption 3 is in fact more stringent than needed, given that Proposition 1 only necessitates the convergence of the pre-averaged returns defined in Eq. (21) to an α-mixing and stationary Gaussian process. This convergence result requires an appropriate limit theorem to hold for a weighted-average of the tick-level returns $r_i = \Delta_i^N Y = \Delta_i^N X + \Delta_i^N \varepsilon$, which is satisfied when the assumed α-mixing and stationary ε is exogenous to X. However, the same result holds when (r_i) itself satisfies such conditions for an appropriate limit theorem, which permits certain dependence structure between X and ε. For brevity, we will stick with the exogenous noise assumption in the analysis henceforward, and examine its potential impact with a more general specification of ε via extensive simulations in Section 4.

With the additive noise under Assumption 3, the noisy observations clearly do not resemble a Gaussian random walk in the limit. There are two main problems:

(i) The noise term dominates the variance of tick-level returns (r_i) and does not shrink as $n \to \infty$;
(ii) The tick-level returns are no longer independent due to the self-dependence of ε.

We now introduce a two-step noise reduction method which facilitates the construction of a sequence of pseudo-observations with desirable properties in the limit:

Step 1: Pre-averaging

We implement the pre-averaging approach of Jacod et al. (2009): We choose a sequence of positive integers k_n satisfying $k_n\sqrt{\Delta_n} = \theta$ for some $\theta > 0$. We calculate log-returns on $(Y_i)_{0\le i\le N}$ that are pre-averaged in a local neighborhood of k_n observations:

$$r_i^* = \frac{1}{k_n}\sum_{j=k_n/2+1}^{k_n} Y_{i+j} - \frac{1}{k_n}\sum_{j=1}^{k_n/2} Y_{i+j} = \sum_{j=1}^{k_n-1} g\Big(\frac{j}{k_n}\Big) r_{i+j}, \tag{21}$$

where $g(s) = s \wedge (1-s)$, for all $i \in \{1, \dots, N'\}$ with $N' = N - 2k_n/2 + 2$.

Step 2: Random sign flip and permutation

We compute the "wild-bootstrapped" returns based on the pre-averaged returns $(r_i^*)_{1\le i\le N'}$ obtained from Step 1:

$$\tilde{r}_i = r^*_{\pi(i)}\delta_{\pi(i)}, \tag{22}$$

where $(\delta_i)_{1\le i\le N'}$ is a sequence of i.i.d. Rademacher random variables, i.e., $\mathbb{P}(\delta_i = -1) = \mathbb{P}(\delta_i = 1) = 1/2$, and $\pi : \{1,\dots,N'\} \mapsto \{1,\dots,N'\}$ is a uniform random permutation of the index set $\{1,\dots,N'\}$.

Under the null, we show that the sequence of "wild-bootstrapped" returns $(\tilde{r}_i)_{1\le i\le N'}$ behave locally like a sequence of i.i.d. Gaussian random variables in the limit:

Proposition 1. *Let ϵ and Y follow Assumption 3. Under the null hypothesis and as $n \to \infty$, the sequence $(\tilde{r}_i)_{1\le i\le N'}$ converges in distribution to a sequence of locally independent[7] and identically distributed Gaussian random variables with variances of order $\sqrt{\Delta_n}$.*

We first discuss why this two-step method can mitigate the impact of noise under the null hypothesis. In Step 1, the standard choice of pre-averaging window balances the orders of X increments and ϵ, such that the pre-averaged returns $(r_i^*)_{1\le i\le N'}$ converge to a centered, stationary and self-dependent Gaussian process as $n \to \infty$. The dependence structure of (r_i^*) arises from both the assumed self-dependent ϵ and overlapping pre-averaging windows. Therefore, we proceed to Step 2 to remove the local dependence, which is inspired by the wild bootstrap introduced by Wu (1986). The random sign flip eliminates serial correlations in (r_i^*). The uniform random permutation assigns equal probability to each of the $N'!$ possible permutations, which ensures that any two variables in $(\tilde{r}_i)_{1\le i\le N'}$ are independent when their indices are not sufficiently far apart from each other in $\{1,\dots,N'\}$ under infill asymptotics.

Proposition 1 inspires the construction of our test in the presence of noise as follows: We generate a sequence of pseudo-observations $(\tilde{Y}_i)_{0\le i\le N'}$ as partial sums of $(\tilde{r}_i)$, where $\tilde{Y}_0 = Y_0$ and $\tilde{Y}_i = \sum_{j=1}^{i}\tilde{r}_j$. Next, we choose a sequence of barrier widths $c = m\Delta_n^{1/4}$ and obtain the sampled observations $(\tilde{Y}_i^{(c)})$. Finally, we follow Section 3.1 to construct the standardized test statistic $\tilde{T}_{c,\epsilon}$ from $(\tilde{Y}_i^{(c)})$ in place of $(X_i^{(c)})$. Formal establishment of its asymptotic properties requires further assumptions about the noise, and is left for future research. We next discuss some plausible properties of $\tilde{T}_{c,\epsilon}$, which are verified through comprehensive simulations with a realistically calibrated noise specification in the next section.

The choice of $c = m\Delta_n^{1/4}$ ensures that the normalized increments $\tilde{r}_i/c$ are invariant to Δ_n, which is analogous to the case without noise. Assuming that $(\tilde{r}_i)_{1\le i\le N'}$ is a sequence of i.i.d. centered Gaussian random variables, $(\tilde{Y}_i)_{0\le i\le N'}$ forms a genuine Gaussian random walk, and thus the same CLT in Theorem 2 would hold for $\tilde{T}_{c,\epsilon}$ under the null. Our simulation results reveal that this CLT still holds for $\tilde{T}_{c,\epsilon}$ constructed from $(\tilde{Y}_i)$. This is because each sampled return is only determined by finitely many increments of $(\tilde{Y}_i)$ within a local horizon, which are indeed asymptotically i.i.d.. Importantly, the convergence rate of $\tilde{T}_{c,\epsilon}$ remains $\sqrt{N}$, which apparently contradicts the optimal $N^{1/4}$ rate of noise-robust IV estimators (Gloter and Jacod, 2001; Xiu, 2010; Reiß, 2011) that also appears in some noise-robust jump tests (Aït-Sahalia et al., 2012). This discrepancy arises because our test statistic does not rely on a noise-robust IV estimator, but rather on a consistent estimator of the scale-invariant barrier width m, which is identified through the variance of $\tilde{r}_i$. As $\tilde{r}_i$ has the same order as the pre-averaged noise, a consistent estimator of m has the same $\sqrt{N}$ rate as that of a noise variance estimator. This finding also reveals that a noise-robust IV estimator is not a pre-requisite for noise-robust jump tests.

4. Monte Carlo simulations

4.1. Simulation design

We simulate an empirically realistic discretized diffusion model for asset prices, which incorporates both time varying tick-variances and transaction activities. Firstly, we simulate a Heston model for the efficient price process X and obtain its tick-level observations, to which we add jumps with different sizes:

$$\begin{aligned} dX_t &= \left(\mu - \frac{\sigma_t^2}{2}\right)dt + \sigma_t dW_t + dX_t'', \quad t\in[0,1] \\ d\sigma_t^2 &= \alpha(\theta - \sigma_t^2)dt + \eta\sigma_t dB_t, \end{aligned} \tag{23}$$

where $W = (W_t)$ and $B = (B_t)$ are standard Brownian motions with $\mathrm{Corr}(W_t, B_t) = \rho$, and X'' is a compound Poisson process, i.e.,

$$X_t'' = \sum_{i=1}^{N_t} J_i, \tag{24}$$

where $N = (N_t)$ is a Poisson process with rate λ, and jump sizes J_i follow a double exponential distribution (Laplace distribution) with location parameter 0 and scale parameter b. To generate all tick-level observations, we discretize X equidistantly on $t = i/n$ for $n = 23{,}400$. Then we modify the observation times $0 \le t_{n,1} < t_{n,2} < \cdots \le 1$ following an inhomogeneous Poisson process with the rate

$$\alpha(t) = 1 - \frac{1}{2}\cos 2\pi t, \tag{25}$$

[7] A formal definition of local independence is given in Eq. (A.141) in Online Appendix A.5.

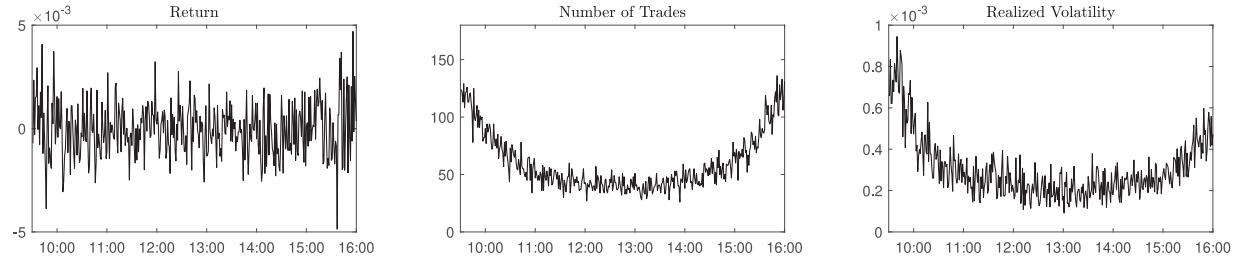

Fig. 3. Some market activity variables of simulated prices. The tick-level observations are simulated with the Heston model in Eq. (23), and we assign randomized observation times with an inverted U-shape rate function in Eq. (25) to all observations. The returns, numbers of transactions, and annualized RVs are computed at a granularity of one minute.

where $t \in [0,1]$. The inverted U-shaped rate function $\alpha(t)$ is employed to mimic the empirical feature of more transactions that occur in the early morning and late afternoon than in the middle of the trading day (Jacod et al., 2017). We draw 10,000 simulated price paths for each experiment.

For the additive noise,[8] we denote

$$\epsilon_i = 2\sqrt{\frac{\sigma^2_{t_{n,i}}}{n}}\left(\omega_i^A + \omega_i^B\sqrt{\frac{\nu-2}{\nu}}\right), \tag{26}$$

where ω_i^A are autocorrelated Gaussian random variables defined as

$$\omega_i^A = \phi_i + \sum_{j=1}^{\Lambda}\beta_j\phi_{i-j}, \quad \text{with } \phi_i \sim \text{i.i.d. } \mathcal{N}(0,1), \text{ and } \beta_j = \frac{d(1+d)\cdots(j-1+d)}{j!}, \tag{27}$$

for $d \in (-0.5, 0.5)$ and a large cutoff value Λ, which form a moving-average series that approximates a fractionally differenced process (Jacod et al., 2019), and ω_i^B are i.i.d. draws from a Student's t distribution with the degree of freedom ν.

The instantaneous standard deviation of the Gaussian-t mixture noise is about four times as much as that of diffusive increments, i.e., $\sqrt{\sigma^2_{t_{n,i}}/n}$, so that the diffusive increments are clearly dominated by the additive noise.[9] This specification of ϵ_i captures some important features of market microstructure noise in financial markets, e.g., temporal heteroskedasticity, slowly-decaying serial correlation, intraday seasonality, and dependence on the latent prices. The t-distributed noise ω_i^B is introduced to capture the large bouncebacks commonly observed in high-frequency transaction data (Aït-Sahalia et al., 2012). Besides the additive noise, we also consider the rounding errors on the price level, i.e., let the observed prices $e^{Y_i} = e^{X_i+\epsilon_i}$ be further rounded to cents. The observed logarithmic prices are given as

$$Y_i = \log\left(\left[\frac{e^{X_i+\epsilon_i}}{0.01}\right]\times 0.01\right), \tag{28}$$

where the function $[x]$ rounds a number x to the nearest integer.[10]

The annualized parameters for the Heston model are fixed at $(\mu,\alpha,\theta,\eta,\rho) = (0.05, 5, 0.16, 0.5, -0.5)$, where the volatility parameters satisfy the Feller's condition $2\alpha\theta \geq \eta^2$ which ensures the positivity of σ. The parameter choices follow both Aït-Sahalia and Jacod (2009b), Aït-Sahalia et al. (2012), which are calibrated according to the empirical estimates in Aït-Sahalia and Kimmel (2007). For the jump components, we let $\lambda = 1$, and $b = 0.2\sqrt{\theta}$ and $0.4\sqrt{\theta}$ corresponding to moderate and relatively large jump sizes. The moderate (resp. large) jumps contribute about 7% (resp. 25%) of the daily QV on average when noise is absent. For the additive noise term, we let $(d,\Lambda,\nu) = (0.3, 100, 2.5)$ following Aït-Sahalia et al. (2012) and Jacod et al. (2019).

Fig. 3 depicts the intraday variation of some market activity variables of a simulated path in the absence of noise, which include the return, number of trades, and annualized RV in each one-minute interval. Both transaction intensity and return variation exhibit a U-shaped pattern over the trading hours, which is in line with some prior empirical findings (Harris, 1986; Wood et al., 1985; Andersen and Bollerslev, 1997; Andersen et al., 2018, 2019, 2024). Fig. 4 compares the simulated tick-level latent prices and the rounded, noise-contaminated price observations over an intraday episode.

[8] The simulation design of additive noise mainly follows Aït-Sahalia et al. (2012). In addition, we consider its serial correlation using the method of Jacod et al. (2019).

[9] In the simulations, we follow Aït-Sahalia et al. (2012) to truncate the t-distributed ω_i^B at $\pm 50\sqrt{\nu/(\nu-2)}$ to avoid large returns in the absence of jumps, which could lead to very misleading results. Hence, the instantaneous standard deviation of the t-distributed noise $2\omega_i^B\sqrt{\sigma^2_{t_{n,i}}/n}\sqrt{(\nu-2)/\nu}$ is slightly lower than $2\sqrt{\sigma^2_{t_{n,i}}/n}$.

[10] We also consider alternative specifications for the additive heteroscedastic noise, see the results in Online Appendix B.3.

Table 1
Finite-sample size and power (%).

Nominal size: 5%	Panel A No jump				Panel B Moderate price jumps				Panel C Large price jumps			
$c/\sigma(r_i)$	$N^{(c)}$	ϵ			$N^{(c)}$	ϵ			$N^{(c)}$	ϵ		
		0.05	0.07	0.10		0.05	0.07	0.10		0.05	0.07	0.10
3	1786	5.26	5.31	5.48	1697	58.21	61.68	65.22	1564	76.29	78.47	80.37
4	1100	5.51	5.58	5.89	1043	61.24	64.54	67.46	959	77.95	80.18	81.99
5	744	5.39	5.56	5.77	705	63.01	66.29	69.55	647	79.10	81.00	82.83
6	536	4.99	5.20	5.61	508	63.77	67.13	70.30	466	80.16	82.01	83.92
7	405	5.28	5.56	5.71	383	65.19	68.47	71.07	351	80.59	82.23	84.01
8	316	5.20	5.61	5.93	299	65.86	68.90	72.07	274	80.76	82.51	84.36
9	254	5.28	5.46	6.01	240	66.33	68.88	71.47	220	81.36	82.78	84.42
10	208	5.07	5.29	5.49	197	66.66	69.33	72.16	181	81.20	83.18	84.85

This table reports the finite-sample size and size-adjusted power (%) of 10,000 simulations of the test statistic $T_{c,\epsilon}$ at 5% nominal level in the absence of market microstructure noise. Tick-level observations are sampled with different PDS barrier widths $c = K\sigma(r_i)$, i.e., K times the standard deviation of tick-by-tick returns, where K ranges from 3 to 10. Different censoring thresholds with $\epsilon \in \{0.05, 0.07, 0.1\}$ are considered. $N^{(c)}$ stands for the average sampling frequencies.

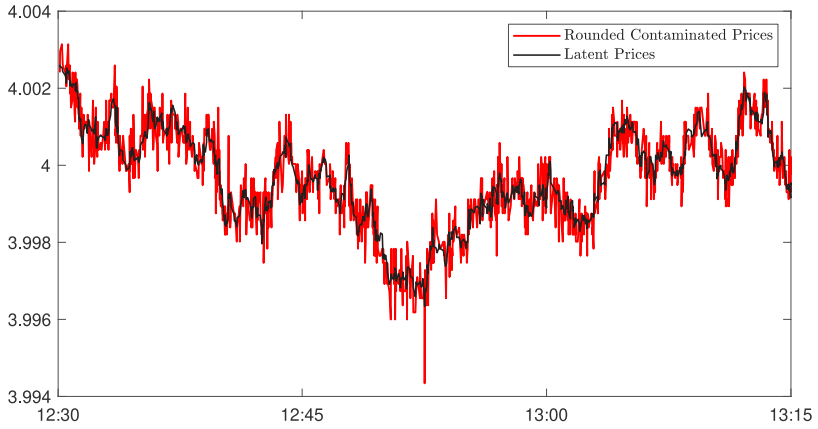

Fig. 4. Comparison of the simulated latent prices and the noise-contaminated prices with rounding errors.

4.2. Test performance in the absence of market microstructure noise

Table 1 reports the finite-sample size and size-adjusted power (at 5% nominal level) of the standardized test statistic $T_{c,\epsilon}$ when noise is absent. Tick-level observations are sampled with different PDS barrier widths $c = K\sigma(r_i)$, i.e., K times the standard deviation of tick-by-tick returns, where K ranges from 3 to 10. Different censoring thresholds with $\epsilon \in \{0.05, 0.07, 0.1\}$ are also considered. In Table 1, the rejection rates under the null (Panel A) are all closely aligned with the nominal level. For the finite-sample power under the alternative (Panels B and C), we find that the rejection rates are fairly robust across different sampling frequencies. Fig. 5 compares the finite-sample distributions of our test statistic with the limiting standard normal distribution. Under the null, the finite-sample distribution (solid line) closely resembles the standard normal (shaded area), while the distribution deviates significantly from $\mathcal{N}(0,1)$ when there exist jumps of either moderate or large sizes.

4.3. Test performance in the presence of market microstructure noise

Panel A in Table 2 summarizes the finite-sample size (at 5% nominal level) of the standardized test statistic $T_{c,\epsilon}$ constructed from the rounded noise-contaminated observations. We employ the two-step noise reduction method in Section 3.2 to construct the sequence of pseudo-observations with three different pre-averaging windows, i.e., $k_n = \lceil \theta\sqrt{N} \rceil$ with $\theta \in \{0.3, 0.4, 0.5\}$. The choices of θ follow the rule of thumb in Hautsch and Podolskij (2013). Similar to the results in the absence of noise, the rejection rates under the null are close to the nominal level across almost all choices of bandwidth c and censoring parameter ϵ. Panels B and C in Table 2 report the size-adjusted power under the alternative with moderate and large jumps, respectively. Compared with the simulation results in Table 1, the finite-sample power experiences a marginal reduction but remains above 40% for most of the parameter choices. Fig. 6 compares the finite-sample distributions of $T_{c,\epsilon}$ with $\mathcal{N}(0,1)$. It is observed that $T_{c,\epsilon}$ is almost a standard normal under the null, but it has a notably larger magnitude than $\mathcal{N}(0,1)$ under the alternative. Comparing this with Fig. 5, we observe that the right tails of the test statistic become smaller with the same jump specifications. This explains the slightly reduced power of our test in the presence of market microstructure noise.

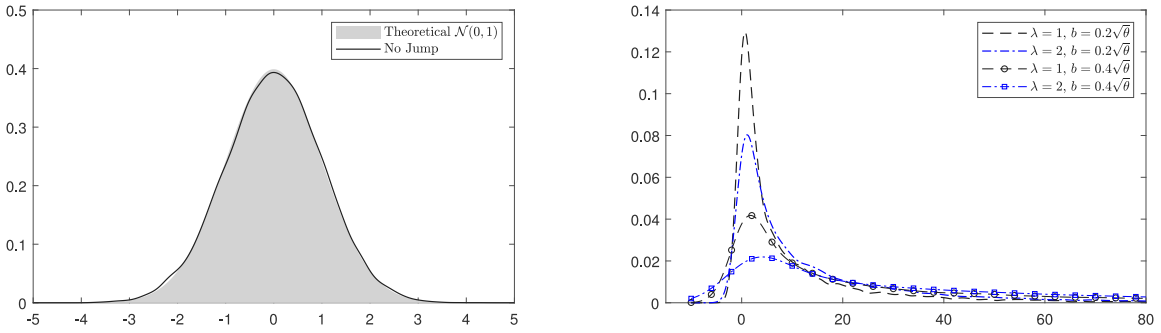

Fig. 5. Finite-sample distributions of the standardized test statistic $T_{c,\epsilon}$ in the absence of noise. We plot the finite-sample distribution under the null (solid line) and compare it with the simulated standard normal (shaded area). Jumps are simulated with a compounded Poisson process with the intensity λ, and their sizes follow a double exponential distribution with location parameter 0 and scale parameter b. We consider different parameter choices: (i) $\lambda = 1$, $b = 0.2\sqrt{\theta}$ (dash), (ii) $\lambda = 2$, $b = 0.2\sqrt{\theta}$ (dash-dot), (iii) $\lambda = 1$, $b = 0.4\sqrt{\theta}$ (dash-circle), and (iv) $\lambda = 2$, $b = 0.4\sqrt{\theta}$ (dash-square). In all cases, the PDS barrier width $c = 5\sigma(r_i)$, and the censoring parameter $\epsilon = 0.05$.

Table 2
Finite-sample size and power (%) in the presence of market microstructure noise.

Nominal size: 5%		$\theta = 0.3$				$\theta = 0.4$				$\theta = 0.5$			
	$c/\sigma(\tilde{r}_i)$	$N^{(c)}$	ϵ			$N^{(c)}$	ϵ			$N^{(c)}$	ϵ		
			0.05	0.07	0.10		0.05	0.07	0.10		0.05	0.07	0.10
	3	1784	4.90	5.15	5.20	1784	4.80	5.35	5.71	1783	5.06	5.07	5.70
	4	1099	4.84	4.95	5.42	1098	5.29	5.10	5.51	1098	5.14	5.08	5.79
	5	743	4.94	5.01	5.20	743	5.19	5.02	5.57	742	4.81	5.02	5.70
Panel A	6	536	4.74	4.89	5.57	536	4.78	5.11	5.58	535	4.96	5.11	5.47
No jump	7	404	4.99	5.11	5.29	404	4.86	5.05	5.76	404	4.86	5.17	5.46
	8	316	5.15	5.37	5.54	316	4.82	5.08	5.43	315	4.81	5.30	5.82
	9	253	5.04	5.41	5.13	254	4.84	5.10	5.63	253	4.96	5.28	5.73
	10	208	5.18	5.10	5.60	208	4.84	5.34	5.54	208	5.04	5.08	5.66
	3	1716	46.00	49.28	51.77	1717	44.98	46.84	49.27	1718	43.08	45.18	47.52
	4	1058	45.89	48.56	50.92	1059	43.31	46.76	48.56	1061	41.35	44.77	46.22
	5	717	44.95	47.45	50.42	719	42.99	45.50	47.53	720	40.81	43.12	44.93
Panel B	6	519	44.60	46.65	48.86	519	42.82	43.97	47.01	520	40.25	42.06	45.01
Moderate jumps	7	392	43.79	45.31	48.98	393	41.00	43.15	46.29	394	40.08	41.95	45.04
	8	307	42.45	45.21	48.64	308	40.97	42.47	46.81	308	39.98	41.03	43.83
	9	247	41.38	43.95	48.57	248	40.54	42.58	45.62	248	38.45	41.21	43.67
	10	203	41.08	44.57	47.51	204	40.04	41.30	45.45	204	38.12	40.03	43.86
	3	1594	68.85	70.38	72.44	1596	68.06	69.14	70.54	1599	66.37	68.54	69.68
	4	983	68.79	70.51	72.17	986	66.37	68.97	70.42	990	65.60	67.26	68.97
	5	668	67.26	69.92	71.80	671	66.37	68.22	69.52	673	65.11	66.50	68.19
Panel C	6	484	67.69	69.41	70.48	486	65.92	67.13	69.67	489	64.38	66.05	68.07
Large jumps	7	367	66.78	68.71	70.81	369	65.14	66.46	68.54	371	63.68	65.35	67.61
	8	288	65.84	68.11	70.24	290	64.22	66.40	68.42	292	62.93	64.89	66.90
	9	233	65.83	67.44	70.32	234	64.08	66.29	68.38	236	62.29	64.68	66.67
	10	192	65.00	66.96	69.70	193	63.44	65.54	68.23	194	61.91	64.12	66.69

This table reports the finite-sample size and size-adjusted power (%) of 10,000 simulations of the test statistic $T_{c,\epsilon}$ at 5% nominal level. All simulated prices are contaminated by the additive Gaussian-t mixture noise and rounding errors. We utilize the two-step noise reduction method in Section 3.2 to construct the sequence of pseudo-observations with three different pre-averaging windows, i.e., $k_n = \lceil \theta\sqrt{N} \rceil$ with $\theta \in \{0.3, 0.4, 0.5\}$. The observations are sampled with different PDS barrier widths $c = K\sigma(\tilde{r}_i)$, where K ranges from 3 to 10. Different censoring thresholds with $\epsilon \in \{0.05, 0.07, 0.1\}$ are considered. $N^{(c)}$ stands for the average sampling frequencies.

We then compare the empirical rejection rates of our test with those of 8 classical high-frequency jump tests constructed from equidistantly calendar-time-sampled observations (Table 3). These tests include BNS (Barndorff-Nielsen and Shephard, 2006), JO (Jiang and Oomen, 2008), LM (Lee and Mykland, 2008), ASJ (Aït-Sahalia and Jacod, 2009b), CPR (Corsi et al., 2010), PZ (Podolskij and Ziggel, 2010), MinRV and MedRV (Andersen et al., 2012). The parameter choices for all these tests are determined in accordance with the recommendations from their original literature.[11] Our analysis, in line with the Monte Carlo results of Dumitru and Urga (2012) and Maneesoonthorn et al. (2020), demonstrates that nearly all the tests constructed from equidistantly calendar-time-sampled observations suffer from size distortion and their results become highly unstable under the assumed additive Gaussian-t

[11] The parameter choices of the competing tests are reported in Online Appendix B.1.

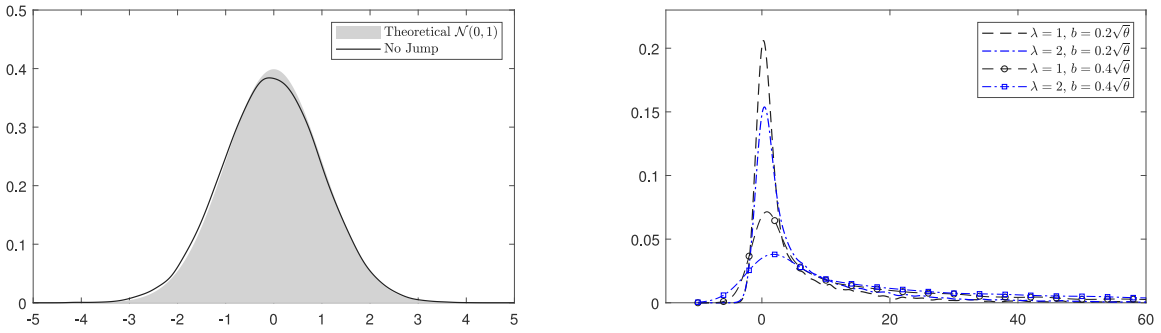

Fig. 6. Finite-sample distributions of the standardized test statistic $T_{c,\epsilon}$ in the presence of noise. We plot the finite-sample distribution under the null (solid line) and compare it with the simulated standard normal (shaded area). Jumps are simulated with a compounded Poisson process with the intensity λ, and their sizes follow a double exponential distribution with location parameter 0 and scale parameter b. We consider different parameter choices: (i) $\lambda = 1$, $b = 0.2\sqrt{\theta}$ (dash), (ii) $\lambda = 2$, $b = 0.2\sqrt{\theta}$ (dash-dot), (iii) $\lambda = 1$, $b = 0.4\sqrt{\theta}$ (dash-circle), and (iv) $\lambda = 2$, $b = 0.4\sqrt{\theta}$ (dash-square). In all cases, we select the pre-averaging window $k_n = \lfloor \theta\sqrt{N} \rfloor = 46$ with $\theta = 0.3$, the PDS barrier width $c = 5\sigma(\bar{r}_i)$, and the censoring parameter $\epsilon = 0.05$.

Table 3
Finite-sample size and power (%) of other tests.

Nominal size: 5%										
	Int. (sec)	N_{spl}	BNS	JO	LM	ASJ	CPR	PZ	MinRV	MedRV
Panel A No jump	5	4680	0.33	6.19	98.40	99.98	33.39	89.61	0.00	0.00
	15	1560	0.42	5.23	71.12	99.42	18.20	52.91	0.00	0.12
	30	780	3.30	5.25	46.82	76.71	13.02	30.25	0.99	2.53
	60	390	5.16	5.75	30.82	29.50	8.47	19.57	3.35	5.30
	120	195	6.46	8.08	17.73	10.60	7.86	16.76	4.78	6.91
	180	130	6.90	8.95	15.10	7.52	8.05	15.96	5.29	8.16
	300	78	7.65	10.87	12.12	4.84	8.97	15.58	5.34	8.98
Panel B Moderate jumps	5	4680	30.08	31.95	15.80	97.25	10.93	12.40	16.91	26.45
	15	1560	36.42	36.33	24.78	94.76	21.13	20.59	32.89	36.17
	30	780	33.20	33.89	32.00	77.29	28.64	28.39	28.95	33.25
	60	390	28.25	28.25	36.63	45.44	29.58	37.36	24.96	28.61
	120	195	21.64	20.90	32.73	24.43	24.18	30.47	20.07	23.51
	180	130	17.40	17.16	28.97	16.57	19.34	25.51	16.34	19.42
	300	78	13.83	11.33	20.56	11.44	15.93	19.09	13.32	14.74
Panel C Large jumps	5	4680	56.35	59.11	43.00	95.12	32.58	36.63	42.72	53.55
	15	1560	60.94	61.18	52.39	95.31	47.32	47.58	58.68	60.84
	30	780	59.06	59.05	58.46	83.73	54.54	55.57	54.79	58.43
	60	390	54.36	54.57	62.78	59.83	56.25	62.88	50.60	54.90
	120	195	46.78	46.16	58.81	36.76	50.12	56.74	44.11	49.20
	180	130	41.21	40.86	54.93	26.94	44.83	51.91	39.32	44.13
	300	78	34.07	33.19	45.73	16.14	38.76	43.76	33.66	37.49

This table reports the finite-sample size and size-adjusted power (%) of 10,000 simulations of 8 classical tests at 5% nominal level: BNS (Barndorff-Nielsen and Shephard, 2006), JO (Jiang and Oomen, 2008), LM (Lee and Mykland, 2008), ASJ (Aït-Sahalia and Jacod, 2009b), CPR (Corsi et al., 2010), PZ (Podolskij and Ziggel, 2010), MinRV and MedRV (Andersen et al., 2012). All these tests are constructed on observations equidistantly sampled with various intervals in calendar time: 5, 15, 30, 60, 120, 180 and 300 s, and "N_{spl}" stands for the sampling frequencies.

mixture noise and rounding errors. This noise significantly distorts their finite-sample null distributions, particularly at higher sampling frequencies. It might be interesting to see that the size of the JO test is close to the nominal level. However, a closer examination reveals that this is caused by two cancelling distortions due to the mixture of Gaussian and t-distributed noise specification, see Online Appendix B.3 for details. While sparse sampling can alleviate size distortion, it also substantially weakens the power of these tests.

For more appropriate benchmarks when noise is present, we also consider some noise-robust versions of classical tests (Table 4) constructed from tick-level or finely sampled observations: the noise-adjusted PZ (Podolskij and Ziggel, 2010), LM12 (Lee and Mykland, 2012), and ASJL (Aït-Sahalia et al., 2012). Similar to our test, all these noise-robust tests rely on the pre-averaging approach of Jacod et al. (2009) to "pre-filter" the noise-contaminated observations.[12] The "optimal" tuning parameters for those

[12] With a simplified i.i.d. noise specification, Jiang and Oomen (2008) propose an analytically modified form of JO. However, it cannot achieve comparable performance under the simulated Gaussian-t mixture noise.

Table 4
Finite-sample size and power (%) of other noise-robust tests.

Nominal size: 5%	Int. (sec)	N_{spl}	PZ*	LM12	ASJL
Panel A: No jump	Tick	23 400	5.29	5.03	5.12
	5	4 680	4.96	8.83	8.79
Panel B: Moderate jumps	Tick	23 400	38.57	22.70	38.22
	5	4 680	30.38	18.79	17.66
Panel C: Large jumps	Tick	23 400	64.78	40.76	63.50
	5	7 680	56.49	31.38	41.96

This table reports the finite-sample size and size-adjusted power (%) of 10,000 simulations of 3 noise-robust tests at 5% nominal level: Noise-adjusted PZ (Podolskij and Ziggel, 2010), LM12 (Lee and Mykland, 2012), and ASJL (Aït-Sahalia et al., 2012). All these tests are constructed on tick-level and 5-second-sampled observations. The tuning parameters for those tests are selected by minimizing the absolute distance between the nominal size and the empirical size with the simulated tick-level noise-contaminated observations.

Table 5
Finite-sample size and power (%) of truncation-based filtering technique.

Nominal size: 5%		Panel A No Jump (with FWER control)		Panel B	Panel C
Ticks	N_{spl}	Šidák	Bonferroni	Moderate jumps	Large jumps
1	23 400	100.00	100.00	9.64	29.57
5	4 680	100.00	100.00	15.33	39.49
15	1 560	99.97	99.96	23.41	49.80
30	780	83.41	82.53	31.92	57.67
60	390	37.31	35.72	38.12	62.78
120	195	15.54	15.18	37.22	61.23
180	130	10.19	10.04	33.09	57.26
300	78	6.99	7.10	26.45	50.42

This table reports the finite-sample size and size-adjusted power (%) of 10,000 simulations of the truncation-based jump filtering technique. Observations are sampled at various multiples of ticks, where "N_{spl}" stands for the corresponding sampling frequencies. The truncation thresholds are constructed from the localized pre-averaged bipower variation of Podolskij and Vetter (2009) computed within each backward-looking window of 1800 ticks. The threshold parameter k is adjusted with both the Šidák and Bonferroni corrections, and $\varpi = 0.5$.

tests are selected by minimizing the absolute distance between the nominal size and the empirical size with the simulated tick-level noise-contaminated observations.[13] We find that, while these noise-robust tests reliably control size, their power performance slightly falls short of that achieved by our proposed test.

Furthermore, we consider the widely employed jump filtering and detection technique from the recent literature (see, e.g., Aït-Sahalia et al., 2025; Aleti et al., 2025) as an alternative benchmark. This method is based on the sequential detection approach of Andersen et al. (2007b) and the thresholding technique of Mancini (2009). Specifically, returns are classified as jumps if their absolute value exceeds the threshold $k\sigma_{t_{n,i}}\Delta_{n,i}^{\varpi}$, where the spot volatility is typically estimated recursively on a backward-looking window with a jump- and noise-robust procedure. With the rolling spot volatility estimates, one periodically checks (e.g., every few transactions or every few minutes) whether a return has exceeded k standard deviations. Since our focus is to test for the existence of jumps over a fixed interval (a trading day), we adjust the threshold parameter k using both the Šidák and Bonferroni corrections — two widely used methods for controlling the family-wise error rate (FWER) — to maintain a nominal size of 5% for the overall procedure and address the multiple testing issue.[14]

Table 5 presents the finite-sample size and size-adjusted power of the truncation-based filtering technique, where we estimate the spot volatilities with the localized pre-averaged bipower variation of Podolskij and Vetter (2009) and evaluate returns every few ticks.[15] Although the filtering technique exhibits reliable size and power performance in the absence of noise with various choices of volatility estimators (see Table B.1 in Online Appendix B.2), our empirically calibrated simulations under noise contamination reveal that it tends to spuriously detect normal returns as jumps, particularly when sampled at a very high frequency. Furthermore, similar

[13] Note that the optimal tuning parameters are not empirically feasible in practice. Therefore, the results presented should be interpreted as upper bounds of the performance for these benchmark tests.

[14] We note that both the Šidák and Bonferroni corrections are highly conservative, as the standardized returns may exhibit serial correlation. A simulation-based procedure adopted by Christensen et al. (2022) can generate data-driven critical values from AR(1) processes that account for serial correlations. However, as the critical values need to be adjusted downwards to achieve the correct size, it results in more inflated rejection rates than those reported in Table 5, and is therefore omitted here.

[15] We also examine returns sampled in equidistant calendar-time intervals, with the adjustments for intraday volatility pattern incorporated, and observe similar finite-sample performance, see Table B.2 in Online Appendix B.2.

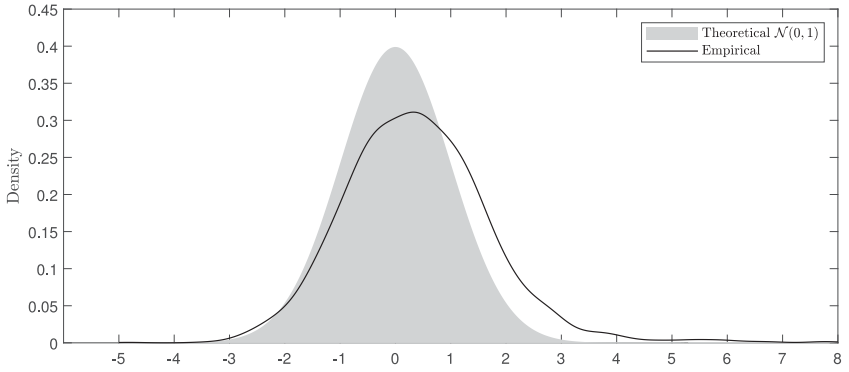

Fig. 7. Testing results for selected NYSE stocks in 2020. We plot the empirical distribution of the standardized test statistic for all 2530 stock-day pairs and, for comparison, the simulated standard normal distribution (shaded area). We use the PDS barrier width $c = 4\sigma(\tilde{r}_i)$, the censoring parameter $\epsilon = 0.05$, and the pre-averaging window $k_n = \lceil \theta\sqrt{N} \rceil$ with $\theta = 0.3$, which corresponds to the first row in Table 7.

to the empirical applications of Aït-Sahalia et al. (2025), we consider a wide range of k from 3.5 to 9, and test for both tick-time- and calendar-time-sampled returns across various sampling frequencies. We find that both inadequate downsampling and low k-values lead to considerable overrejection under the null, whereas further sparse sampling and more stringent truncation thresholds result in a loss of power under the alternatives (see Figs. B.1 and B.2 in Online Appendix B.2), which indicates the difficulty of balancing such trade-offs when employing the truncation-based filtering technique.

As illustrated in Table 2, our PDS-based test demonstrates robustness across various parameter choices: (i) barrier width c, (ii) censoring parameter ϵ, and (iii) pre-averaging window $k_n = \lceil \theta\sqrt{N} \rceil$, even when we consider such a complicated and realistic noise specification. Furthermore, our test remains competitive and, often superior, to those noise-robust tests with optimal parameter choices. While we refrain from providing optimal parameter choices, we offer recommended ranges for practitioners:

(i) Choose c as a multiple of the standard deviation of $\tilde{r}_i$, i.e., $c = K\sigma(\tilde{r}_i)$, with $3 \le K \le 10$.
(ii) Choose ϵ in $[0.03, 0.15]$.
(iii) Choose the pre-averaging window $k_n = \lceil \theta\sqrt{N} \rceil$ with $\theta \in [0.2, 0.8]$.

Through extensive simulation studies with different specifications of market frictions, we believe that the recommended parameter choices work reasonably well in finite samples when the number of intraday tick-level observations is no less than 10,000. Additional simulation results can be found in Online Appendix B.3.

5. Empirical analysis

In this section, we employ our new jump test on the high-frequency transaction data of 10 stocks listed on the New York Stock Exchange (NYSE): American Express (AXP), Boeing (BA), Disney (DIS), IBM, Johnson & Johnson (JNJ), JP Morgan (JPM), Merck (MRK), McDonald's (MCD), Procter & Gamble (PG), and Walmart (WMT). Our Trade and Quote (TAQ) dataset includes all transactions from 9:30 am to 4:00 pm on each trading day in 2020. As is standard in empirical research involving high-frequency financial data, we apply filters, as outlined in Barndorff-Nielsen et al. (2009), to eliminate obvious data errors, remove all transactions in the original record that are later corrected, canceled or otherwise invalidated, and retain only transactions from NYSE. Table 6 reports descriptive statistics of trades on these selected NYSE stocks, which include the number of trades, observed transaction prices in dollar terms, and intraday log-returns in basis points. Our PDS-based test utilizes the same tuning parameters as those in Section 4: the PDS barrier width $c = K\sigma(\tilde{r}_i)$ with K ranging from 4 to 6, the censoring parameter $\epsilon = 0.05$, and three pre-averaging windows $k_n = \lceil \theta\sqrt{N} \rceil$ with $\theta \in \{0.3, 0.4, 0.5\}$.

Table 7 reports the proportions of trading days with rejections in 2020, as determined by our PDS-based test. For the selected stocks, the proportions of trading days with identified jumps are no more than 20%, with only AXP and MCD identified to exhibit over 15% of trading days containing jumps. There is little variation in the rejection rates across different stocks, and the results are relatively stable with different parameter choices. For each stock, there is a slight decrease in the percentage of identified jumps when we employ a larger barrier width c for PDS, i.e., sample less frequently. To visualize the testing results for the selected stocks in 2020, we aggregate all stock-day outcomes, which yields a total of 2530 stock-day pairs. Fig. 7 illustrates the empirical distributions of the standardized test statistic (solid line) and compares it with the standard normal distribution $\mathcal{N}(0,1)$. Relative to the limiting distribution under the null hypothesis of no jump (shaded area), the empirical distribution of our test statistic deviates slightly towards the right side, but maintains a bell shape centered around 0.5.

To eliminate spurious detections due to the multiple testing issue, Bajgrowicz et al. (2016) propose a formal treatment of the over-identification bias with double asymptotics when the jump tests are applied over a sample of many days. We apply their thresholding methods to our results: (i) the universal threshold $\sqrt{2\ln 253}$, and (ii) the threshold based on the false discovery rate

Table 6
Descriptive statistics of daily trades on selected NYSE stocks.

Stock		AXP	BA	DIS	IBM	JNJ
Number of trades	Min	3171	10 556	9785	5047	6383
	Max	59 273	245 802	125 550	49 178	71 733
	Mean	19 351	55 314	37 962	17 818	22 966
	Std Dev	9205	38 352	20 390	8265	11 561
Transaction prices	Min	67.03	89.00	79.07	90.56	109.16
	Max	138.16	349.45	183.40	158.78	157.66
	Mean	100.88	181.46	121.78	123.16	143.33
	Std Dev	14.72	51.15	20.12	11.57	8.53
Intraday log-returns (1×10^{-4})	Min	−123.24	−163.29	−100.49	−143.49	−110.04
	Max	97.48	129.22	75.30	143.49	200.25
	Mean	0.00	0.00	0.00	0.00	0.00
	Std Dev	1.78	1.85	1.17	1.44	1.23
Stock		JPM	MRK	MCD	PG	WMT
Number of trades	Min	12 593	5787	3968	7516	9845
	Max	156 987	71 570	55 024	76 337	90 546
	Mean	44 738	22 833	16 096	23 224	26 148
	Std Dev	25 335	11 058	7838	10 422	13 032
Transaction prices	Min	76.92	65.26	124.23	94.31	102.00
	Max	141.10	92.14	231.91	146.92	153.60
	Mean	103.17	79.86	195.17	125.19	128.55
	Std Dev	14.28	4.55	21.74	11.63	12.04
Intraday log-returns (1×10^{-4})	Min	−103.80	−177.00	−154.39	−132.25	−305.08
	Max	103.80	117.50	142.80	207.58	190.19
	Mean	0.00	0.00	0.00	0.00	0.00
	Std Dev	1.06	1.26	1.74	1.33	1.13

This table contains summary statistics for the number of trades, observed transaction prices in dollars, and intraday log-returns in basis points for 10 selected NYSE stocks in 2020. Data are collected from the TAQ database which includes all transactions from 9:30 am to 4:00 pm in each trading day. We apply filters, as outlined in Barndorff-Nielsen et al. (2009), to eliminate clear data errors, remove all transactions in the original record that are later corrected, cancelled or otherwise invalidated, and keep transactions on NYSE only.

Table 7
Empirical rejection rates (%) for selected NYSE stocks.

k_n	$c/\sigma(\tilde{r}_i)$	AXP	BA	DIS	IBM	JNJ	JPM	MRK	MCD	PG	WMT
$\theta=0.3$	4	17.00	10.67	9.88	13.04	13.83	10.67	10.67	16.60	13.44	11.07
	5	15.81	10.67	9.49	12.25	11.86	10.67	10.28	16.21	11.86	10.67
	6	14.23	9.88	9.09	12.25	11.46	10.28	9.88	15.02	11.86	11.07
$\theta=0.4$	4	16.21	9.88	10.28	13.44	12.65	10.67	9.88	15.81	12.65	12.25
	5	15.02	9.88	9.49	12.25	12.25	10.67	9.49	15.42	11.86	11.07
	6	14.23	9.49	9.49	11.46	11.07	9.88	9.49	14.23	11.46	10.28
$\theta=0.5$	4	15.81	10.28	10.28	12.25	13.04	10.28	10.28	15.81	12.25	11.46
	5	14.23	9.09	9.88	12.25	11.46	9.49	9.09	15.02	11.86	10.67
	6	13.44	8.70	9.09	11.46	11.07	9.88	9.09	14.62	11.07	10.67

This table reports the proportions of days with jumps identified by the PDS-based test for 10 NYSE stocks in 2020. We use three pre-averaging windows $k_n=\lceil\theta\sqrt{N}\rceil$ with $\theta\in\{0.3,0.4,0.5\}$, different PDS barrier widths $c=K\sigma(\tilde{r}_i)$, i.e., the integer multiple of the standard deviation of pre-averaged returns, with K ranging from 4 to 6, and the censoring parameter $\epsilon=0.05$. The total number of trading days is 253.

(FDR).[16] The adjusted results of our test for all selected stocks are reported in Table 8. It is noteworthy that our testing results are fairly robust to the control of spurious detections, which underscores the empirical reliability of our PDS-based test.

The empirical results for alternative calendar-time-sampling-based and noise-robust tests — both with and without the control of spurious detections — are reported in Online Appendix B.4. We find that the outcomes of calendar-time-based tests vary substantially with the sampling frequency, whereas the noise-robust tests constructed from all available tick-level observations generally identify fewer days with jumps. Especially, the truncation-based detection method flags jumps on over 60% of trading days when returns are sampled every 30 s, which echoes the findings of Aït-Sahalia et al. (2025), but this proportion drops to about 20% when the data are down-sampled to 5-minute intervals. Our proposed test yields results comparable to certain noise-robust benchmarks — such as the ASJL test constructed from tick-level data — but exhibits superior robustness to the control of spurious detections.

[16] For the vector of one-side test statistics $(S_1,S_2,\ldots,S_N)'$ which converge to i.i.d. standard normal random variables under the null, the universal threshold is $\sqrt{2\ln N}$ (Bajgrowicz et al., 2016). The data-adaptive FDR threshold is determined from the observed p-value distribution by the Benjamini–Hochberg procedure.

Table 8
Adjusted empirical rejection rates (%) for selected NYSE stocks.

	k_n	$c/\sigma(\bar{r}_i)$	AXP	BA	DIS	IBM	JNJ	JPM	MRK	MCD	PG	WMT
Panel A Universal threshold	$\theta = 0.3$	4	15.02	9.49	9.09	12.25	12.25	7.91	9.49	14.62	11.07	9.49
		5	13.83	9.88	8.70	11.07	10.28	8.30	9.09	14.62	9.88	8.70
		6	12.65	8.70	8.30	11.07	10.28	7.91	8.70	13.44	9.88	9.09
	$\theta = 0.4$	4	14.23	8.70	9.49	12.25	11.07	8.30	8.70	14.23	10.28	9.88
		5	13.44	8.70	8.70	11.46	11.07	8.30	8.30	13.83	9.88	9.09
		6	13.04	8.70	8.70	10.67	9.88	7.91	8.30	12.65	9.49	8.70
	$\theta = 0.5$	4	13.83	9.09	9.09	11.46	11.86	8.30	9.09	13.83	10.28	9.49
		5	12.65	8.30	9.09	11.07	10.28	7.51	7.91	13.44	9.88	8.70
		6	11.86	7.91	8.30	10.67	9.88	7.51	8.30	13.04	9.49	9.09
Panel B FDR threshold	$\theta = 0.3$	4	13.44	9.09	8.70	11.86	11.07	7.11	9.09	13.44	9.88	8.70
		5	12.65	9.49	8.70	11.07	9.88	7.51	8.70	13.44	9.09	8.30
		6	11.86	8.30	8.30	10.67	9.88	7.51	8.30	12.65	9.09	8.30
	$\theta = 0.4$	4	13.04	8.70	9.09	11.86	10.28	7.51	8.30	13.04	9.88	9.09
		5	12.25	8.30	8.30	11.07	10.67	7.51	8.30	12.65	9.09	8.70
		6	12.25	8.70	8.30	9.88	9.49	7.51	7.91	11.86	9.09	8.30
	$\theta = 0.5$	4	12.65	9.09	8.30	11.07	11.07	7.51	9.09	12.25	9.09	9.09
		5	11.46	7.91	8.70	10.67	9.49	7.11	7.91	12.25	9.09	7.91
		6	11.07	7.91	8.30	10.28	9.49	6.72	7.91	12.25	9.09	8.30

This table reports the proportions of days with jumps identified by the PDS-based test for 10 NYSE stocks in 2020, with the control of spurious detections using (i) the universal threshold and (ii) the FDR threshold of Bajgrowicz et al. (2016). We use three pre-averaging windows $k_n = \lceil \theta\sqrt{N} \rceil$ with $\theta \in \{0.3, 0.4, 0.5\}$, different PDS barrier widths $c = K\sigma(\bar{r}_i)$, i.e., the integer multiple of the standard deviation of pre-averaged returns, with K ranging from 4 to 6, and the censoring parameter $\epsilon = 0.05$. The total number of trading days is 253.

6. Conclusions

This paper introduces a novel nonparametric high-frequency jump test for a discretely observed Itô semimartingale. Our approach utilizes a path-dependent sampling scheme for the tick-level price observations. The key intuition behind the construction of our test is that, different from a continuous price increase or decrease over a certain time interval, a discontinuous shift with a larger magnitude will always trigger an exit-time event and induce a disproportionately large threshold exceedance under infill asymptotics. Additionally, a two-step noise reduction technique is designed to alleviate the impact of weakly dependent market microstructure noise. Through extensive simulations, we validate the reliable finite-sample performance of our test under empirically realistic specifications for price observations, which is convincingly superior to a comprehensive collection of "classical" methods. The Monte Carlo results demonstrate that the performance of our test is robust to various aggregation levels and tuning parameter choices. An empirical analysis of NYSE-traded stocks provides strong statistical evidence for jumps across all selected stocks, and the results are robust to the correction of spurious detections. This methodology stands as the first exploration of the duration-based approach to test for jumps, which offers a robust and easy-to-implement tool for researchers and practitioners.

Acknowledgments

We thank the Co-Editor Torben Andersen, the Associate Editor, and two anonymous referees for their constructive comments and suggestions. We also thank Leopoldo Catania (discussant), Carsten Chong, Kim Christensen, Dobrislav Dobrev, Roxana Halbleib, Seok Young Hong, Aleksey Kolokolov, Sébastien Laurent, Manh Pham, Roberto Renò, George Tauchen, Stephen Taylor, Viktor Todorov, Giovanni Urga, as well as participants at the CFE-CMStatistics 2020, 6th KoLaMaFr Workshop on Financial Econometrics, XXIII Workshop on Quantitative Finance, Conference on Intrinsic Time in Finance, Economics of Financial Technology Conference 2022, EcoSta 2022, QFFE 2022, IAAE 2022, Econometric Society Asian Meeting 2022, SoFiE 2022, AsianFA 2022, Workshop on Volatility, Jumps and Bursts, Econometric Society Australasian Meeting 2022, 25th Dynamic Econometrics Conference, and seminars at various institutions, for helpful comments and suggestions. Shifan Yu acknowledges financial support from the Economic and Social Research Council (ESRC) North West Social Science Doctoral Training Partnership (ES/P000665/1). The MATLAB code example can be founded at https://github.com/Shifan-Yu/Jumps_Endogenous_Sampling.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2025.106132.

References

Aït-Sahalia, Y., Jacod, J., 2009a. Estimating the degree of activity of jumps in high frequency data. Ann. Stat. 37 (5A), 2202–2244.
Aït-Sahalia, Y., Jacod, J., 2009b. Testing for jumps in a discretely observed process. Ann. Stat. 37 (1), 184–222.
Aït-Sahalia, Y., Jacod, J., Li, J., 2012. Testing for jumps in noisy high frequency data. J. Econometrics 168 (2), 207–222.

Aït-Sahalia, Y., Kimmel, R., 2007. Maximum likelihood estimation of stochastic volatility models. J. Financial Econ. 83 (2), 413–452.
Aït-Sahalia, Y., Li, C.X., Li, C., 2025. So Many Jumps, So Little News. Working Paper.
Aït-Sahalia, Y., Mykland, P.A., Zhang, L., 2005. How often to sample a continuous-time process in the presence of market microstructure noise. Rev. Financial Stud. 18 (2), 351–416.
Aït-Sahalia, Y., Mykland, P.A., Zhang, L., 2011. Ultra high frequency volatility estimation with dependent microstructure noise. J. Econometrics 160 (1), 160–175.
Aleti, S., Bollerslev, T., Siggaard, M., 2025. Intraday market return predictability culled from the factor zoo. Manag. Sci. (forthcoming).
Andersen, T.G., Bollerslev, T., 1997. Intraday periodicity and volatility persistence in financial markets. J. Empir. Finance 4 (2–3), 115–158.
Andersen, T.G., Bollerslev, T., Diebold, F.X., 2007a. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. Rev. Econ. Stat. 89 (4), 701–720.
Andersen, T.G., Bollerslev, T., Dobrev, D., 2007b. No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications. J. Econometrics 138 (1), 125–180.
Andersen, T.G., Bondarenko, O., Kyle, A.S., Obizhaeva, A.A., 2018. Intraday Trading Invariance in the E-Mini S&P 500 Futures Market. Working Paper.
Andersen, T.G., Dobrev, D., Schaumburg, E., 2008. Duration-Based Volatility Estimation. Working Paper.
Andersen, T.G., Dobrev, D., Schaumburg, E., 2012. Jump-robust volatility estimation using nearest neighbor truncation. J. Econometrics 169 (1), 75–93.
Andersen, T.G., Su, T., Todorov, V., Zhang, Z., 2024. Intraday periodic volatility curves. J. Amer. Statist. Assoc. 119 (546), 1181–1191.
Andersen, T.G., Thyrsgaard, M., Todorov, V., 2019. Time-varying periodicity in intraday volatility. J. Amer. Statist. Assoc. 114 (528), 1695–1707.
Bajgrowicz, P., Scaillet, O., Treccani, A., 2016. Jumps in high-frequency data: Spurious detections, dynamics, and news. Manag. Sci. 62 (8), 2198–2217.
Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2009. Realized kernels in practice: Trades and quotes. Econ. J. 12 (3), 1–32.
Barndorff-Nielsen, O.E., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. J. Financial Econ. 2 (1), 1–37.
Barndorff-Nielsen, O.E., Shephard, N., 2006. Econometrics of testing for jumps in financial economics using bipower variation. J. Financial Econ. 4 (1), 1–30.
Bibinger, M., Winkelmann, L., 2015. Econometrics of co-jumps in high-frequency data with noise. J. Econometrics 184 (2), 361–378.
Bollerslev, T., Li, S.Z., Zhao, B., 2020. Good volatility, bad volatility, and the cross section of stock returns. J. Financial Quant. Anal. 55 (3), 751–781.
Bollerslev, T., Todorov, V., Xu, L., 2015. Tail risk premia and return predictability. J. Financial Econ. 118 (1), 113–134.
Caporin, M., Kolokolov, A., Renò, R., 2017. Systemic co-jumps. J. Financial Econ. 126 (3), 563–591.
Chernozhukov, V., Chetverikov, D., Kato, K., 2013. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. Ann. Stat. 41 (6), 2786–2819.
Chernozhukov, V., Chetverikov, D., Kato, K., 2019. Inference on causal and structural parameters using many moment inequalities. Rev. Econ. Stud. 86 (5), 1867–1900.
Christensen, K., Oomen, R., Renò, R., 2022. The drift burst hypothesis. J. Econometrics 227 (2), 461–497.
Corsi, F., Pirino, D., Renò, R., 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. J. Econometrics 159 (2), 276–288.
Cremers, M., Halling, M., Weinbaum, D., 2015. Aggregate jump and volatility risk in the cross-section of stock returns. J. Financ. 70 (2), 577–614.
Dimitriadis, T., Halbleib, R., 2022. Realized quantiles. J. Bus. Econom. Statist. 40 (3), 1346–1361.
Dumitru, A.-M., Urga, G., 2012. Identifying jumps in financial assets: A comparison between nonparametric jump tests. J. Bus. Econom. Statist. 30 (2), 242–255.
Durrett, R., 2019. Probability: Theory and Examples. Cambridge University Press, Fifth ed..
Engle, R.F., Russell, J.R., 1998. Autoregressive conditional duration: A new model for irregularly spaced transaction data. Econometrica 66 (5), 1127–1162.
Fan, J., Yao, Q., 2003. Nonlinear Time Series: Nonparametric and Parametric Methods. Springer.
Fukasawa, M., 2010. Realized volatility with stochastic sampling. Stochastic Process. Appl. 120 (6), 829–852.
Fukasawa, M., Rosenbaum, M., 2012. Central limit theorems for realized volatility under hitting times of an irregular grid. Stochastic Process. Appl. 122 (12), 3901–3920.
Gerhard, F., Hautsch, N., 2002. Volatility estimation on the basis of price intensities. J. Empir. Finance 9 (1), 57–89.
Gloter, A., Jacod, J., 2001. Diffusions with measurement errors. I. Local asymptotic normality. ESAIM Probab. Stat. 5, 225–242.
Hansen, P.R., Lunde, A., 2006. Realized variance and market microstructure noise. J. Bus. Econom. Statist. 24 (2), 127–161.
Harris, L., 1986. A transaction data study of weekly and intradaily patterns in stock returns. J. Financial Econ. 16 (1), 99–117.
Hautsch, N., Podolskij, M., 2013. Preaveraging-based estimation of quadratic variation in the presence of noise and jumps: Theory, implementation, and empirical evidence. J. Bus. Econom. Statist. 31 (2), 165–183.
Hong, S.Y., Nolte, I., Taylor, S.J., Zhao, X., 2023. Volatility estimation and forecasts based on price durations. J. Financial Econ. 21 (1), 106–144.
Huang, X., Tauchen, G., 2005. The relative contribution of jumps to total price variance. J. Financial Econ. 3 (4), 456–499.
Ibragimov, I.A., 1962. Some limit theorems for stationary processes. Theory Probab. Appl. 7 (4), 349–382.
Ibragimov, I.A., Linnik, Y.V., 1971. Independent and Stationary Sequences of Random Variables. Wolters-Noordhoff.
Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., Vetter, M., 2009. Microstructure noise in the continuous case: The pre-averaging approach. Stochastic Process. Appl. 119 (7), 2249–2276.
Jacod, J., Li, Y., Zheng, X., 2017. Statistical properties of microstructure noise. Econometrica 85 (4), 1133–1174.
Jacod, J., Li, Y., Zheng, X., 2019. Estimating the integrated volatility with tick observations. J. Econometrics 208 (1), 80–100.
Jiang, G.J., Oomen, R.C., 2008. Testing for jumps when asset prices are observed with noise – a "swap variance" approach. J. Econometrics 144 (2), 352–370.
Jiang, G.J., Yao, T., 2013. Stock price jumps and cross-sectional return predictability. J. Financial Quant. Anal. 48 (5), 1519–1544.
Jing, B.-Y., Kong, X.-B., Liu, Z., Mykland, P., 2012. On the jump activity index for semimartingales. J. Econometrics 166 (2), 213–223.
Khaniyev, T., Kucuk, Z., 2004. Asymptotic expansions for the moments of the Gaussian random walk with two barriers. Statist. Probab. Lett. 69 (1), 91–103.
Koike, Y., 2017. Time endogeneity and an optimal weight function in pre-averaging covariance estimation. Stat. Inference Stoch. Process. 20 (1), 15–56.
Kolokolov, A., Renò, R., 2024. Jumps or staleness? J. Bus. Econom. Statist. 24 (2), 516–532.
Laurent, S., Shi, S., 2020. Volatility estimation and jump detection for drift-diffusion processes. J. Econometrics 217 (2), 259–290.
Lee, S.S., Mykland, P.A., 2008. Jumps in financial markets: a new nonparametric test and jump dynamics. Rev. Financial Stud. 21 (6), 2535–2563.
Lee, S.S., Mykland, P.A., 2012. Jumps in equilibrium prices and market microstructure noise. J. Econometrics 168 (2), 396–406.
Li, Z.M., Linton, O., 2022. A ReMeDI for microstructure noise. Econometrica 90 (1), 367–389.
Li, J., Liu, Y., 2021. Efficient estimation of integrated volatility functionals under general volatility dynamics. Econometric Theory 37 (4), 664–707.
Li, Y., Mykland, P.A., Renault, E., Zhang, L., Zheng, X., 2014. Realized volatility when sampling times are possibly endogenous. Econometric Theory 30 (3), 580–605.
Liu, L.Y., Patton, A.J., Sheppard, K., 2015. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. J. Econometrics 187 (1), 293–311.
Lorden, G., 1970. On excess over the boundary. Ann. Math. Stat. 41 (2), 520–527.
Lotov, V.I., 1996. On some boundary crossing problems for Gaussian random walks. Ann. Probab. 24 (4), 2154–2171.
Mancini, C., 2009. Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. Scand. J. Stat. 36 (2), 270–296.
Maneesoonthorn, W., Martin, G.M., Forbes, C.S., 2020. High-frequency jump tests: Which test should we use? J. Econometrics 219 (2), 478–487.
Nolte, I., Xu, Q., 2015. The economic value of volatility timing with realized jumps. J. Empir. Finance 34, 45–59.
Pelger, M., 2020. Understanding systematic risk: A high-frequency approach. J. Financ. 75 (4), 2179–2220.
Podolskij, M., Vetter, M., 2009. Bipower-type estimation in a noisy diffusion setting. Stochastic Process. Appl. 119 (9), 2803–2831.

Podolskij, M., Ziggel, D., 2010. New tests for jumps in semimartingale models. Stat. Inference Stoch. Process. 13 (1), 15–41.
Potiron, Y., Mykland, P.A., 2017. Estimation of integrated quadratic covariation with endogenous sampling times. J. Econometrics 197 (1), 20–41.
Reiß, M., 2011. Asymptotic equivalence for inference on the volatility from noisy observations. Ann. Stat. 39 (2), 772–802.
Rogozin, B.A., 1964. On the distribution of the first jump. Theory Probab. Appl. 9 (3), 450–465.
Tse, Y.-K., Yang, T.T., 2012. Estimation of high-frequency volatility: An autoregressive conditional duration approach. J. Bus. Econom. Statist. 30 (4), 533–545.
Ubukata, M., Oya, K., 2009. Estimation and testing for dependence in market microstructure noise. J. Financial Econ. 7 (2), 106–151.
Varneskov, R.T., 2017. Estimating the quadratic variation spectrum of noisy asset prices using generalized flat-top realized kernels. Econometric Theory 33 (6), 1457–1501.
Vetter, M., Zwingmann, T., 2017. A note on central limit theorems for quadratic variation in case of endogenous observation times. Electron. J. Stat. 11 (1), 963–980.
Wood, R.A., McInish, T.H., Ord, J.K., 1985. An investigation of transactions data for NYSE stocks. J. Financ. 40 (3), 723–739.
Wu, C.-F.J., 1986. Jackknife, bootstrap and other resampling methods in regression analysis. Ann. Stat. 14 (4), 1261–1295.
Xiu, D., 2010. Quasi-maximum likelihood estimation of volatility with high frequency data. J. Econometrics 159 (1), 235–250.
Yan, S., 2011. Jump risk, stock returns, and slope of implied volatility smile. J. Financial Econ. 99 (1), 216–233.
Zhang, L., Mykland, P.A., Aït-Sahalia, Y., 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. J. Amer. Statist. Assoc. 100 (472), 1394–1411.
Zhou, B., 1996. High-frequency data and volatility in foreign-exchange rates. J. Bus. Econom. Statist. 14 (1), 45–52.