Jumps versus bursts: distinguishing sources of extreme risk in financial markets*

Xiaolu Zhao[†] Seok Young Hong[‡] Oliver Linton[§]

June 4, 2026

Abstract

Extreme price movements in financial markets can arise from different mechanisms, yet empirical methods often fail to distinguish between them. We develop a new approach to separate discrete price jumps from volatility bursts using ultra-high-frequency data. Our method employs an endogenous threshold sampling scheme that localizes large price movements in real time and remains robust to market microstructure noise and trading frictions. Applying the method to twenty years of data on Dow Jones Industrial Average stocks, we show that the two phenomena have distinct drivers. Price jumps are primarily associated with liquidity shocks, reflected in increases in trading volume, bid-ask spreads, and order imbalance, whereas volatility bursts are more closely linked to news arrivals. Distinguishing these sources of extreme risk provides new insights into market dynamics and has implications for risk monitoring, trading strategies, and asset pricing models.

JEL classification: G12, G14, C14

Keywords: Price jumps; Volatility bursts; Market microstructure noise; Endogenous sampling; High-frequency trading; News sentiment.

* This paper has benefited greatly from the valuable feedback of Kay Giesecke (Department Editor), the Associate Editor, and two anonymous referees. We also thank Torben Andersen, Dobrislav Dobrev, Jean Jacod, Aleksey Kolokolov, Ingmar Nolte, Manh Cuong Pham, Roberto Renò, Viktor Todorov and Christina Dan Wang for valuable comments and suggestions. We extend our gratitude to participants at the 2021 Society for Financial Econometrics (SoFiE) conference in San Diego, United States, the 2025 World Congress of the Econometric Society in Seoul, Republic of Korea, and the Workshop on Volatility, Jumps and Bursts in Lancaster, United Kingdom as well as seminar participants at Nanyang Technological University, Singapore. Xiaolu Zhao gratefully acknowledges financial support from the National Science Foundation of China (Grant 12271363), and Seok Young Hong gratefully acknowledges financial support from the Ministry of Education (MOE) of the Republic of Singapore (Tier 1 Seed Grant 023705-00001).

[†]University of Nottingham, United Kingdom; Email: vera.zhao@nottingham.ac.uk

[‡]Nanyang Technological University, Singapore; Email: seokyoung.hong@ntu.edu.sg

[§]Corresponding author. University of Cambridge, United Kingdom; Email: obl20@cam.ac.uk

1 Introduction

Extreme price movements are a defining feature of modern financial markets. Within seconds, prices can change dramatically, concentrating risk and triggering rapid responses by traders, market makers, and risk managers. Understanding the origin of these movements is therefore central to financial decision making. Yet a fundamental empirical challenge remains: large price movements can arise from different mechanisms. Some represent discrete price jumps, genuine discontinuities in the price process often associated with liquidity shocks or abrupt trading imbalances. Others reflect volatility bursts, periods of unusually intense but continuous price fluctuations that frequently accompany the arrival of new information. In practice, however, these phenomena are difficult to distinguish in observed data.

Most empirical methods detect jumps using prices sampled at predetermined intervals. While convenient, such interval-based sampling schemes impose observation times that are unrelated to the evolution of the price process. Consequently, genuine jumps may be detected with delay or missed entirely, while sequences of rapid but continuous price changes can accumulate within a sampling interval and be mistakenly classified as jumps. These problems are particularly acute in modern electronic markets, where trading is irregular and market microstructure frictions−such as bid-ask bounce, price discreteness, and intermittent trading−affect observed prices. This paper develops a new approach to distinguish discrete price jumps from volatility bursts using ultra-high-frequency data. Our method replaces conventional fixed-interval sampling with an endogenous threshold sampling scheme, in which observations occur whenever cumulative price changes exceed a small threshold. Because sampling times adapt to the realized price path, large discontinuities trigger observations immediately. This feature allows the method to localize price jumps precisely while reducing distortions caused by arbitrary sampling intervals. The approach is simple to implement and remains robust to market microstructure noise.

This paper contributes to the literature in three ways. First, we introduce a noise-robust jump detection procedure based on endogenous threshold sampling. The method exploits a simple empirical contrast: market microstructure noise generates frequent but small price fluctuations, whereas genuine price jumps are rare but large. Using this contrast, we construct a test statistic that remains reliable even when observed prices are contaminated by noise and trading frictions. We establish the asymptotic properties of the estimator and derive a central limit theorem for statistical inference. Second, we develop a framework that separates discrete price jumps from volatility bursts. By combining local jump detection statistics with daily variation measures, we obtain a proxy for volatility bursts after controlling for discrete discontinuities. This decomposition provides a clearer characterization of extreme price variation than existing approaches that tend to confound these mechanisms. Third, we provide new empirical evidence on the sources of extreme price movements using twenty years of ultra-high-frequency data for the 30 stocks in the Dow Jones Industrial Average. Our results show that the proposed method detects large price jumps with high power while maintaining low misclassification rates relative to conventional interval-based tests. We also find that volatility bursts represent a distinct component of market risk, accounting on average for approximately 7-8% of realized volatility across stocks. Separating these two forms of extreme price variation also clarifies their economic origins. Using high-frequency trading variables and news sentiment data, we show that price jumps are primarily associated with liquidity shocks, reflected in increases in trading volume, bid-ask spreads, and order imbalance. In contrast, volatility bursts are more closely linked to news arrivals and information shocks. Because liquidity shocks and news arrivals often occur simultaneously, empirical analyses that do not distinguish these mechanisms may generate mixed or conflicting conclusions regarding the drivers of extreme price movements.

Our study relates to a large literature on jump detection in high-frequency financial data. Early contributions develop nonparametric methods for identifying price jumps using high-frequency returns. For example, Barndorff-Nielsen and Shephard (2006) propose tests based on bipower variation that separate continuous volatility from discontinuous jump components, while Lee and Mykland (2008) introduce a nonparametric test that detects the timing and magnitude of jumps. These approaches, together with related methods based on power variations, truncation, or thresholding, form the basis of a widely used econometric toolkit for jump detection. Subsequent research expands this framework in several directions, including tests for co-jumps across assets (Jacod and Todorov (2009)), studies of related aspects of high-frequency return dynamics such as financial leverage effects and jump risk in asset pricing (Aït-Sahalia et al. (2013); Aït-Sahalia et al. (2017); Kalnina and Xiu (2017); Chong and Todorov (2024)), and applications to find evidence of potential manipulation by large traders (Scaillet et al. (2020)). Other work develops jump-robust measures of volatility and related quantities for forecasting and factor models (Fan and Wang (2007); Bollerslev and Todorov (2011); Patton and Sheppard (2015); Aït-Sahalia and Xiu (2016); Giesecke and Schwenkler (2019); Li et al. (2019); Pelger (2019, 2020); Li and Linton (2022); Ding et al. (2024)).

A growing body of evidence also highlights limitations of interval-based jump detection methods. Simulation and empirical studies show that different jump tests may yield inconsistent results and can be sensitive to sampling frequency and market microstructure noise. For example, some methods may detect an excessive number of jumps when applied to very high-frequency data, reflecting the influence of measurement noise rather than genuine discontinuities. Our approach contributes to this literature by adopting an endogenous sampling design that allows observations to be triggered by the evolution of the price process itself. By allowing sampling times to adapt to the price path, the proposed method improves the localization of price discontinuities and enables a clearer distinction between discrete jumps and volatility bursts.

The remainder of the paper proceeds as follows. Section 2 introduces the theoretical framework and endogenous sampling scheme. Section 3 presents the estimators used to measure price variation. Section 4 develops the jump detection procedures and asymptotic theories. Section 5 reports simulation evidence comparing the proposed method with existing approaches. Section 6 provides empirical illustrations using IBM stock data. Section 7 presents empirical results for the Dow Jones Industrial Average stocks. Section 8 examines the economic drivers of jumps and volatility bursts. Section 9 concludes. The Internet Appendix contains simulation results, additional empirical results, and proofs.

2 Theoretical foundation

We suppose that the latent log-price X_t is an Itô-semimartingale with jumps defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ admitting the following representation:

$$dX_t = \mu_t dt + \sigma_t dW_t + dJ_t, \tag{1}$$

where W_t is a standard Brownian motion, J_t is a finite activity jump process, and μ_t and σ_t are $(\mathcal{F}_t)$-adapted processes that are locally bounded and càdlàg, respectively. Here, the latent price refers to a noise-free process underlying the observed price and may, but need not, coincide with an efficient price. We have $dJ_t = U_t dR_t$, where R_t is the counting process for jumps and U_t is the jump size. One parameter of interest is the total quadratic variation of the latent price over the time interval [0,t], denoted $[X, X]_t$, which is almost surely equal to

$$[X, X]_t = \int_0^t \sigma_s^2 \; ds + \sum_{s\leq t} U_s^2. \tag{2}$$

The observations are recorded at times $\{t_j\}_j$, a sequence of endogenous stopping times generated by the hitting time scheme defined below in Section 2.1. We denote by Y_{t_j} the log-price we actually observe at times t_j. Their discrepancy from the latent price X_{t_j} stems from the presence of microstructure noise arising from, for example, bid/ask spread and data errors. The literature on market microstructure noise is extensive, and a widely adopted practice assumes an additive model, see Hasbrouck (1993), Zhang et al. (2005), Hansen and Lunde (2006), Bandi and Russell (2008), Park et al. (2016), and Li et al. (2022). Recent studies propose more general frameworks for microstructure noise; see, among others, Jacod et al. (2017) and Da and Xiu (2021). In the same spirit, we adopt a set of assumptions that arises naturally from our endogenous sampling scheme. These assumptions do not hinge on a specific structural model for microstructure noise and thus offer greater flexibility for econometric modeling. Section 2.2 shows that they are implied by, and are satisfied under, a broad class of standard microstructure-noise models.

2.1 The sampling scheme

We introduce the sampling scheme we adopt in this paper. How observation times are defined is crucial for the asymptotic behavior of volatility estimators and for their limiting variance[1]; see, e.g., Aït-Sahalia and Mykland (2003); Mykland and Zhang (2006); Oomen (2006); Hayashi et al. (2011); Zhang et al. (2022). We allow the sampling times to be both random and endogenous, which can account for key features of high-frequency financial data. The advantage and usefulness of random and endogenous sampling schemes (relative to deterministic equidistant sampling) have been widely recognized.

Let n be the parameter that defines the observation frequency and drives the asymptotics. For a single trading day $[0, 1]$ and a selected threshold δ, a set of event times $t_j \equiv t_{n,j},\ j \geq 0$ is defined in terms of absolute cumulative price changes exceeding δ. Under such an endogenous sampling scheme, a price jump will trigger an event immediately, thus enabling timely detection and precise localization of intraday price jumps (Section 4). Additionally, with δ sufficiently small (practical choice of threshold and justification discussed in Section 4.3), the problem of misclassifying *accumulated continuous price changes* as price jumps is substantially mitigated. For the asymptotic derivations, we suppose that the sequence of thresholds $\delta = \delta_n \to 0$ as $n \to \infty$, which is consistent with the infill asymptotics (and our practical selection) and is a natural choice for the econometric analysis of high-frequency data.

Formally, we model the event times as the hitting times defined as $t_{n,0} = 0$ and

$$t_{n,j+1} := \inf\{t > t_{n,j};\ |Y_t - Y_{t_{n,j}}| \geq \delta_n\}; \qquad j \geq 0. \tag{3}$$

The resulting time points form a sequence of strictly increasing stopping times. The number of price events, i.e. the number of 'hits' or 'crosses' up to time t is written

$$N_t := \max\{j \geq 0;\ t_{n,j} \leq t\}. \tag{4}$$

Since the sampling times are defined in terms of the observed price Y, the scheme is feasible and is straightforward to implement. It can be traced back to the barrier model of Cho and Frees (1988)

[1] Specifically, deriving a central limit theorem can be challenging in complex yet arguably more realistic settings.

for price discreteness and has since been discussed in the stochastic time-sampling literature; see, among others, Fukasawa and Rosenbaum (2012), Robert and Rosenbaum (2012), Li et al. (2013), Potiron and Mykland (2017), Hong et al. (2023), and Pelletier and Wei (2024).

Remark. Under (1) and (3), consecutive event-time changes in Y need not equal δ_n even when there are no jumps. Only when microstructure noise is absent does Y coincide with X, in which case the changes equal δ_n without jumps by continuity of the sample path. Under the scheme (3), data are collected whenever the absolute cumulative change in the observed price surpasses the barrier/threshold of size δ_n. So the resulting samples $t_{n,j}$ can be thought of as tick times - the transaction times with a price change (or quote-revision times), and therefore they belong to some full transaction record $\mathcal{T}_n := \{t^*_{1,j}, t^*_{2,j}, \ldots\}$ during the day $[0, 1]$. For notational convenience, we may suppress the index n in $t_{n,j}$ when no confusion arises.

We impose the following conditions on the jump behavior of the price process and on the sampling mesh. These assumptions are mild and arise naturally under endogenous sampling as sufficient (though not necessary) conditions that allow us to proceed without imposing a specific structural model for microstructure noise. Section 2.2 shows that they hold for a broad class of standard microstructure-noise models.

Assumption A. *(i) For some real constants $\xi_{k,t} \in [0, \infty)$, $k = 1, 2$, we have for all t that*

$$\frac{1}{N_t}\sum_{j=1}^{N_t}\left(\frac{|Y_{t_{j+1}} - Y_{t_j}|}{\delta_n}\right)^k \;:=\; \frac{1}{N_t}\sum_{j=1}^{N_t} a_{n,j}^k \;=\; \xi_{k,t} + o_p(\delta_n). \tag{5}$$

(ii) When there is no jump on $[0, t]$[2], we have

$$\frac{1}{N_t}\sum_{i=1}^{N_t-1}\sum_{j=i+1}^{N_t}\left(a_{n,i} - a_{n,j}\right)^2 \xrightarrow{p} 0. \tag{6}$$

Remark. These are natural regularity requirements under the hitting-time construction. Together with Assumption C below, Assumption A links the observed price Y and the latent price X through the threshold δ_n. Condition (i) ensures that overshoots around the barrier remain well behaved. In the noise-free and jump-free case $Y \equiv X$, we have $\xi_{1,t} = 1 = \xi_{2,t}$; under no jump, condition (ii) implies $\xi^2_{1,t} = \xi_{2,t}$. Condition (ii) posits that the noise remains stable across event times when there is no jump, which is consistent with microstructure frictions being locally stable in practice.

[2] That is, on the event Ω^C_t, defined in (22). The convergence in probability is understood in restriction to Ω^C_t.

Assumption B. *The mesh of the sampling points (3) satisfies the following: as $\delta_n \to 0$,*

$$\max_{1\le j\le N_t} |t_{n,j+1} - t_{n,j}| \;=\; O_p(\delta_n^2). \tag{7}$$

Remark. Under infill asymptotics, it is required to assume the sampling mesh $\max_j |t_{n,j+1} - t_{n,j}|$ tends to zero. Additionally, one further imposes a rate condition to obtain asymptotic results beyond consistency, such as limiting distributions, see Jacod and Shiryaev (2013). For example, Hayashi et al. (2011) (*the mixed renewal scheme*), Fukasawa and Rosenbaum (2012), Li et al. (2014), and Phillips and Yu (2023) (*flat price trading*) consider the cases for which the mesh is $O_p(\delta_n^2)$, $o_p(\delta_n^2)$, $o_p(\delta_n^{4/3+\epsilon})$, and $o(\delta_n)$, respectively, depending on the estimator and the limiting result of interest. Assumption B is fully consistent with this established practice. Its compatibility with the endogenous sampling scheme in the presence of microstructure noise is discussed in Section 2.2.

Assumption C. *The latent price process X satisfies the following:*

$$\sum_{j=1}^{N_t} \left\{ \delta_n^2 - \left(\Delta X_{t_{j+1}}\right)^2 1_{\left\{|\Delta X_{t_{j+1}}| \le \delta_n\right\}} \right\} = o(\delta_n). \tag{8}$$

Remark. At first glance, Assumption C might appear somewhat technical, but it is a mild regularity condition used in lieu of adopting a specific structural model for microstructure noise. Since the event times t_j are generated from the observed price Y, we need a condition linking the latent process X to the threshold δ_n. By construction, $\sum_{j=1}^{N_t} (\Delta X)^2 1\{|\Delta X| \le \delta_n\} \le N_t \delta_n^2 \le \sum_{j=1}^{N_t} (\Delta Y)^2$, and Assumption C ensures the gap between the first two terms is asymptotically small, which we use to establish our asymptotic theory. The assumption holds in many standard settings, for instance, when $X_t = \sigma W_t$ and $Y_t = X_t + \epsilon_t$ with noise ϵ_t of modest size. Section 2.2 provides detailed justification and shows that widely used microstructure noise models satisfy this assumption.

2.2 Detailed discussion and validity of Assumptions

As noted above, Assumption B is rather widely used in stochastic sampling. In our setting, it is a high-level mesh condition on the endogenous duration process induced by (3). The hitting-time construction itself does not fix the maximal duration rate; rather, Assumption B imposes the standard mesh rate condition needed for the limiting theory. This rate is natural under threshold sampling because the duration scale for a continuous Itô price to move by δ_n is of order δ_n^2. Specifically, under this condition, the continuous component of the latent price has $O_p(\delta_n)$ increments, while the noise increment considered below is asymptotically negligible. In this sense, Assumption B is compatible with our endogenous sampling scheme in the presence of noise considered here.

Assumptions A and C, which emerge as a natural consequence of endogenous sampling, are new to the best of our knowledge. They are high-level conditions that are sufficient but not necessary. They systematically link and relate the latent price X, the observed price Y, and the threshold δ, while avoiding any structural restrictions on microstructure noise.

In this subsection, we further discuss the validity and generality of Assumptions A and C. In particular, we show that they are *weaker than*, and hence *implied and satisfied by*, a broad class of canonical microstructure noise models in the high-frequency econometrics literature.

Example 1. The additive noise model

The additive noise model, which has been widely used and extensively studied in the literature, specifies the following structure for the microstructure noise:

$$Y_{t_j} = X_{t_j} + \delta_n^\alpha v_{t_j}, \tag{9}$$

where v_{t_j} is square-integrable and $\alpha \geq 0$. For the fixed noise case ($\alpha = 0$), see for example, Bandi and Russell (2008), Jacod et al. (2017), Chen and Mykland (2017), and Li and Linton (2022). For the shrinking noise case ($\alpha > 0$), see, among many others, Kalnina and Linton (2008), Zhang (2011), Aït-Sahalia and Jacod (2014), and Da and Xiu (2021).

Remark. Suppose the latent price process X follows (1) and is sampled according to (3). Let the observed price Y be defined by (9), where v_{t_j} is mean-square Hölder-continuous with exponent $\beta \geq 0$ (with $\beta = 0$ means having a bounded variance). Then, (9) always satisfies Assumptions A and C under Assumption B and $\alpha + 2\beta > 2$; see the Internet Appendix for the proof.

The constant α determines the speed at which the noise decays and affects the limiting behavior of the volatility signature plot. As an illustrative example, we consider the regular sampling case where $t_j = j/n$ and $\delta_n = n^{-1/2}$. Then, as the sampling frequency $n = N$ goes up, the left end of the volatility signature plot (i) shoots sharply upward when $\alpha = 0$ (or close to zero), (ii) mildly increases and remains comparable to the signal when $\alpha = 1$, and (iii) becomes relatively flat when $\alpha > 1$. The effective value of α depends on the nature of the high-frequency data under study, and can therefore differ across assets, venues, and time periods, see Da and Xiu (2021) for further insights and relevant discussions.

Example 2. The Roll bid-ask model

The Roll model (see Hasbrouck (1993), for example) is a seminal and classic framework for analyzing the impact of the bid-ask spread on asset prices:

$$Y_{t_j} \;=\; X_{t_j} + \varepsilon_{t_j}, \qquad \varepsilon_{t_j} = \frac{s}{2} q_{t_j} \tag{10}$$

where $q_{t_j} \in \{1, -1\}$ satisfies $\mathbb{E}(q_{t_j}) = 0$ and $\mathbb{E}(q_{t_j} q_{t_{j-1}}) = -\rho$ for some $\rho \in [0, 1]$, and s is the bid-ask spread. Under Assumption B and $s = s_n = o(\delta_n^2)$, Assumptions A and C are satisfied.

Example 3. The rounding noise

Another popular model is to consider a pure rounding noise, e.g. Ball (1988), Rosenbaum (2009):

$$Y_{t_j} \;=\; [X_{t_j}/\beta]\beta \;=\; X_{t_j} + r_{t_j}; \qquad r_{t_j} \in \left[-\frac{\beta}{2}, \frac{\beta}{2}\right], \tag{11}$$

where $\beta > 0$ is the rounding level, so that $|r_{t_{j+1}} - r_{t_j}| \leq \beta$. This bound concerns the rounding error part, and the observed price change can still span multiple ticks when the changes in the latent price are large. Again, under Assumption B and $\beta = \beta_n = o_p(\delta_n^2)$, Assumptions A and C hold.

We have shown that our assumptions are satisfied by some widely used models in the literature under mild conditions. All proofs are enclosed in the Internet Appendix. In Section 4.3, we provide practical guidance on choosing the threshold and discuss its implications for these conditions.

3 Duration-Based Estimators

We now discuss three price-duration-based estimators we consider in this paper, namely DV, DV^C and RV. They can be readily constructed, and are straightforward to implement. Later, we show how they can be utilized to conduct a jump test.

The first estimator DV. The first duration-based estimator is given by

$$DV_t(\delta_n) \;=\; N_t \delta_n^2. \tag{12}$$

This estimator has been studied in the literature, see e.g. Andersen et al. (2008), Fukasawa and Rosenbaum (2012), and Hong et al. (2023). In particular, in the absence of both jumps and microstructure noise (so that $X \equiv Y$), Hong et al. (2023) establish asymptotic mixed normality under the hitting time sampling scheme. They further discuss how specific structural forms of bid-ask spread and time discreteness affect the limiting bias and variance. (17) in Proposition 1 below generalizes their result to allow for jumps and a general form of microstructure noise.

The second estimator DV^C. In this paper we propose a novel, different version of the duration-based estimator denoted DV^C, whose difference from DV can be interpreted as reflecting the market microstructure effect. The estimator is given by

$$DV_t^C(\delta_n) \;=\; \left[\frac{1}{N_t}\sum_{j=1}^{N_t}\left(\frac{|Y_{t_{j+1}} - Y_{t_j}|}{\delta_n}\right)\right]^2 \cdot DV_t(\delta_n) \tag{13}$$

As defined earlier, Y is the actual *observed price* modulated by the presence of microstructure noise. Writing $|Y_{t_{j+1}} - Y_{t_j}|/\delta_n = a_{n,j}$, the estimator can be re-written as $DV_t^C(\delta_n) = \frac{1}{N_t}(\sum_{j=1}^{N_t} a_{n,j})^2\delta_n^2$.

The third estimator RV. The last estimator we consider is the popular realized volatility (RV) based on the prices observed at $\{t_j\}$:

$$RV_t(\delta_n) \;=\; \sum_{j=1}^{N_t}\left(Y_{t_{j+1}} - Y_{t_j}\right)^2. \tag{14}$$

While we define (14) in terms of the hitting/event times t_j, it could also be defined using the full transaction record $\mathcal{T}_n = \{t_1^*, t_2^*, \dots\}$. In that case, the limiting distribution will depend on how the transaction record $\{t_j^*\}$ is defined. The RV has a price duration interpretation as follows:

$$RV_t(\delta_n) \;=\; \sum_{j=1}^{N_t}\delta_n^2\left(\frac{|Y_{t_{j+1}} - Y_{t_j}|}{\delta_n}\right)^2 \;=\; \delta_n^2\sum_{j=1}^{N_t} a_{n,j}^2. \tag{15}$$

Remark. Unlike the latent price X, the observed price Y is not necessarily continuous even when there are no jumps. However, when there is no microstructure noise so that X and Y are identical, the consecutive differences in X and Y will always be δ_n in the absence of jump. Consequently, in the absence of both jumps and noise, all three estimators (12), (13) and (14) coincide.

The deviation of DV^C and RV from the first estimator DV can be viewed as reflecting the contribution of microstructure noise, such as the bid/ask spread and price discretization. Moreover, the difference between DV_t^C and RV_t captures the contribution of price discontinuities over $[0, t]$. Therefore, we later propose a noise-robust jump detection procedure based on this difference. For later reference, note that the difference

$$RV_t(\delta_n) - DV_t^C(\delta_n) \;=\; \frac{\delta^2}{N_t}\left(N_t\sum_{j=1}^{N_t} a_{n,j}^2 - \left(\sum_{j=1}^{N_t} a_{n,j}\right)^2\right) \;=\; \frac{\delta^2}{N_t}\sum_{i=1}^{N_t-1}\sum_{j=i+1}^{N_t}\left(a_{n,i} - a_{n,j}\right)^2, \tag{16}$$

which is non-negative. For any “regular events” not caused by a price jump, $a_{n,i}$ and $a_{n,j}$ are close in the presence of noise (Assumption A). If there are volatility bursts, which are composed of large continuous price changes, the difference between an $a_{n,i}$ from a burst period and a regular $a_{n,j}$ can be large. Hence, the daily statistic we derive later may also reflect the influence of volatility bursts.

3.1 Asymptotic Properties of the Estimators

The purpose of price jump tests is to assess whether there is a discontinuous component, dJ_t over the interval of interest $[0, t]$. Using the estimators presented above, we propose an easy-to-apply nonparametric procedure that detects large jumps from the price process in the presence of noise. We first present the limiting behavior of DV, DV^C and RV.

Proposition 1. *Suppose that Assumptions A, B, and C hold, and let the latent log price X be an Itô-semimartingale with jumps, i.e. (1). Then, for all $t \leq 1$, under the sampling scheme in (3) the estimators DV_t, DV_t^C, RV satisfy the following as $\delta_n \to 0$:*

$$DV_t(\delta_n) = N_t \cdot \delta_n^2 \xrightarrow{p} \langle X, X \rangle_t \equiv \int_0^t \sigma_s^2 \, ds \tag{17}$$

$$DV_t^C(\delta_n) = \left(\frac{1}{N_t} \sum_{j=1}^{N_t} \frac{|Y_{t_{j+1}} - Y_{t_j}|}{\delta_n} \right)^2 N_t \delta_n^2 \xrightarrow{p} \xi_{1,t}^2 \int_0^t \sigma_s^2 ds \tag{18}$$

$$RV_t(\delta_n) = \sum_{j=1}^{N_t} \left(Y_{t_{j+1}} - Y_{t_j} \right)^2 \xrightarrow{p} \xi_{2,t} \int_0^t \sigma_s^2 \, ds + \xi_{2,t} \sum_{s \leq t} U_s^2, \tag{19}$$

where $\xi_{1,t}$ and $\xi_{2,t}$ are the constants defined in (5).

Remark. Proposition 1 shows that RV provides an appropriate measure of the total quadratic variation, while DV and DV^C provide appropriate measures of the integrated variance of the price process. The first limit (17) generalizes the result in Hong et al. (2023) by allowing for jumps and MMS. Note that DV is not affected by microstructure noise, in the sense that its probability limit depends only on the latent price X. Furthermore, DV and DV^C are unaffected by the presence of jumps. This is similar to the jump-robustness of truncation-type estimators in the literature.

An immediate consequence of Proposition 1 is that

$$DV_t^C(\delta_n) - DV_t(\delta_n) \xrightarrow{p} (\xi_{1,t}^2 - 1) \int_0^t \sigma_s^2 ds, \tag{20}$$

which can be viewed as representing the effect of the market microstructure noise. Note that when $\xi_{2,t} = 1$, the right hand side of (19) becomes the quadratic variation $\int_0^t \sigma_s^2 \, ds + \sum_{s \leq t} U_s^2 \equiv [X, X]_t$. Let $r_k = Y_{t_{k+1}} - Y_{t_k}$, where the kth event is triggered by a price jump. Then, it follows that

$$RV_t(\delta_n) - DV_t^C(\delta_n) = \frac{N_t - 1}{N_t} (|r_k| - |r_{k-1}|)^2 + o_p(\delta_n^2), \tag{21}$$

which can be interpreted as the large price jumps extracted from RV. Accordingly, we use the difference between DV^C and RV to identify large price jumps in the presence of noise on a daily basis. The choice of the threshold δ is important in practice, and we elaborate on this in Section 4.3. We discuss local (intraday) jump detections in Section 4.2.

4 Noise Robust Jump Detection Procedure

The question of whether the path of the price process $X_t(\omega)$ has jumps over $[0, t]$ is equivalent to asking whether the realization ω belongs to either

$$\begin{aligned} \Omega_t^C &= \{\omega : \ t \mapsto X_t(\omega) \text{ is continuous over } [0, t]\} \qquad \text{or} \\ \Omega_t^J &= \{\omega : \ t \mapsto X_t(\omega) \text{ is discontinuous at least at one point over } [0, t]\}. \end{aligned} \tag{22}$$

As in other jump detection procedures, the aim is to test the null hypothesis of *no jump*; that is, if $\omega \in \Omega^C$. The main idea we adopt is to construct a testing procedure based on comparing the limiting behavior of the *continuous component* and the *jump component* of the price process.

4.1 A daily jump test

In view of (18) and (19) we define the following scale-free "ratio statistic":

$$\mathcal{S}_{t,n} \; := \; \frac{DV_t^C(\delta_n)}{RV_t(\delta_n)} \tag{23}$$

We may expect this statistic to be more stable than the simpler "linear statistic", Aït-Sahalia and Jacod (2014). Note that the probability limit of the statistic is given by

$$\mathcal{S}_{t,n} \xrightarrow{p} \begin{cases} 1 & \text{if } \omega \in \Omega_t^C \\ \frac{\xi_{1,t}^2 \int_0^t \sigma_s^2 \; ds}{\xi_{2,t} \int_0^t \sigma_s^2 \; ds + \xi_{2,t} \sum_{s \le t} U_s^2} & \text{if } \omega \in \Omega_t^J, \end{cases} \tag{24}$$

which is because $\xi_1^2 = \xi_2$ under the null of no jump (Assumption A).

This leads to a jump test based on $\mathcal{S}_{t,n}$ and the consistency of the test, as we show below. We first establish the joint central limit theorem for DV^C and RV. We have the following result:

Theorem 1. *Under the conditions assumed in Proposition 1 and the sampling scheme (3), the estimators DV^C and RV satisfy the following: as $\delta_n \to 0$,*

$$\left(\delta_n^{-1}\left(DV_t^C(\delta_n) - \xi_{1,t}^2 \int_0^t \sigma_s^2 \; ds\right), \delta_n^{-1}\left(RV_t(\delta_n) - \xi_{2,t} \int_0^t \sigma_s^2 \; ds - \xi_{2,t} \sum_{s \le t} U_s^2\right)\right) \;\longrightarrow\; (\mathcal{W}_{1,t}, \mathcal{W}_{2,t} + \mathcal{Y}_t) \tag{25}$$

where the convergence is $\mathcal{F}$-stably in law. The continuous centered Gaussian martingale $\mathcal{W}_{1,t}$ and $\mathcal{W}_{2,t}$ are defined on an extension of the space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \ge 0}, \mathbb{P})$ with the conditional variance given by $(2/3)\xi_1^4 \int_0^t \sigma_s^2 \; ds$ and $(2/3)\xi_2^2 \int_0^t \sigma_s^2 \; ds$, respectively. Furthermore, $\mathcal{Y}_t = \sum_{T_p \le t} 2\xi_{2,t}\eta_p U_{T_p}$, where T_p's form an enumeration of stopping times at which jump occurs, and $(\eta_p)_{p \ge 1}$ is a sequence of identically distributed random variables with density $f_\eta(x) \;=\; \int_0^\infty \frac{1}{\sqrt{2\pi t}} \sum_{k=-\infty}^{\infty} (\exp(-(x-4k)^2/(2t)) - \exp(-(x+2-4k)^2/(2t)))dt 1_{[-1,1]}(x)$.

The result above reflects the random nature of the sampling times. For example, when observations are equally spaced (regular sampling), η is instead a mixture of normal distributions (Jacod, 2008, Theorem 2.11). The limit components are pairwise uncorrelated, as shown in the proof.

We propose the following robust jump testing statistic:

$$Z_{n,t} \;=\; \frac{1 - \frac{DV_t^C(\delta_n)}{RV_t(\delta_n)}}{\sqrt{\frac{\frac{2}{3}(\widehat{\xi}_{2,t}^2 + \widehat{\xi}_{1,t}^4)\mathcal{D}_n(\delta_n)}{RV_t^2}}}, \tag{26}$$

where $\mathcal{D}_n(\delta_n) = N_t\delta_n^4$, and $\widehat{\xi}_{k,t} = \frac{1}{N_t}\sum_{j=1}^{N_t}(|Y_{t_{j+1}} - Y_{t_j}|/\delta_n)^k$. Building on the above, in Theorem 2 below we show that $Z_{n,t}$ is asymptotically standard normal distributed under the null of no jump. In doing so, we make use of the following stable convergence in law:

$$\delta_n^{-1}\left(1 - \frac{DV_t^C(\delta_n)}{RV_t(\delta_n)}\right) = \delta_n^{-1}\left(\frac{RV_t(\delta_n) - DV_t^C(\delta_n)}{RV_t(\delta_n)}\right) \longrightarrow \frac{\mathcal{W}_{2,t} - \mathcal{W}_{1,t}}{\xi_{2,t}\int_0^t \sigma_s^2\, ds}, \tag{27}$$

when the price *is continuous (no jump)*, i.e., in restriction to the set Ω_t^C.

Theorem 2. *Suppose the latent log price X follows (1) and is sampled at $\{t_{n,j}\}$ defined in (3). Under the null hypothesis of no jump, $Z_{n,t}$ is asymptotically standard normally distributed*[3]*. I.e., as $\delta_n \to 0$, for each $t \in (0,1]$,*

$$\mathcal{L}(Z_{n,t}|A) \longrightarrow N(0,1), \tag{28}$$

if $A \in \mathcal{F}, A \subset \Omega_t^C, \mathbb{P}(A) > 0$, where $\mathcal{L}(Z_{n,t}|A)$ denotes the distribution of $Z_{n,t}$ under the conditional probability $\mathbb{P}(\cdot|A)$. Furthermore, the tests with critical region $\mathcal{C}_n$ based on (28) have the strong asymptotic size α for the null hypothesis Ω_t^C, and are consistent for the alternative Ω_t^J.

Jump variation (JV) quantifies the magnitude of detected jump-type variation as a fraction of daily realized variation. For our daily test, we define it as

$$JV_d = \frac{RV - DV^C}{RV}. \tag{29}$$

4.2 A local jump test

We consider a local jump statistic motivated by Andersen, Bollerslev, and Dobrev (2007) (hereafter ABD). Let N_t be the number of endogenous price events on day t and let $r_{q,t} = Y_{t_q} - Y_{t_{q-1}}$ be the q-th event return, for $q = 1, \ldots, N_t$. We define the local statistic by the standardized event return

$$Z_{q,t}^l = \frac{r_{q,t}}{\sqrt{\Delta_t \cdot DV_t^C}}. \tag{30}$$

where $\Delta_t = 1/N_t$. This uses DV_t^C as a noise-included, jump-robust estimate, thus avoiding sparse sampling and achieving precise localization of price jumps, which trigger price events immediately as they occur. When the q-th event contains no price jump, $Z_{q,t}^l = O_p(1)$ by Assumption A and the consistency of the DV^C statistic in Proposition 1. On the other hand, when there is a fixed-size jump, $|Z_{q,t}^l| \to \infty$, which gives the consistency intuition for the local test.

Following ABD, we use the Gaussian approximation to calibrate the local critical value. A local jump is detected when $|Z_{q,t}^l| > \Phi^{-1}(1 - \beta_t/2) = c_t$, where Φ is the standard normal cdf, $\beta_t = 1 - (1-\alpha)^{1/N_t}$, and α is the daily significance level. Under the ABD approximation, this

[3] In the sense of *stable convergence in law in restriction to the set A, see Aït-Sahalia and Jacod (2014, page 95).*

choice controls the daily FWER, since $\mathbb{P}(\max_{1\leq q\leq N_t}|Z^l_{q,t}|\leq c_t)=(1-\beta_t)^{N_t}=1-\alpha$. Since FDR is bounded above by FWER, this also yields conservative FDR control. Beyond this approximation, in our endogenous event-time case, Assumption A implies that $\max_q|Z^l_{q,t}|=O_p(1)$, while the critical value c_t diverges; hence the rule is asymptotically conservative. Furthermore, since $|Z^l_{q,t}|$ diverges at the $1/\delta_n$ rate, the test is consistent against fixed-size local jumps.

The influence of a special source of noise, called staleness, is discussed in detail in the simulation and empirical sections. For the local test, we define jump variation as follows: assuming K jumps are detected and r_k is the return of the kth price jump,

$$JV_l = \sum_{k=1}^{K} r_k^2 \Big/ RV. \tag{31}$$

In the remainder of the paper, we use DV^c to refer to the local jump test. We refer to the daily jump test in (28) as the DV^C test, given the role of DV^C in the daily test statistic $Z_{n,t}$.

4.3 Selection of threshold

The threshold parameter δ is important in our endogenous thresholding approach. From a practical point of view, it should be small enough for at least three reasons: 1) to render sufficient price events to average out large price jumps; 2) to stop continuous price changes from accumulating into false discontinuities, and to identify/record large continuous price changes such as those resulting from volatility bursts (for the daily test); 3) to effectively shrink the sampling interval to the minimal permitted by the market microstructure (MMS) noise. The last point brings about the lower bound of δ: it shall be large enough to guard against MMS noise elements such as bid/ask spread and price-discretization. Additionally, though our approach operates on observed prices, we expect the underlying latent price to have moved during any price event as well, despite possible bid/ask bounces. Accordingly, the threshold parameter is set as the average spread of the day plus 1 tick (i.e. 0.01 for the U.S. equities):

$$\delta_p = s + 1. \tag{32}$$

Here, δ_p is the threshold price change; the implied threshold return is

$$\delta_j = \frac{\delta_p}{P_{t_{j-1}}}, \tag{33}$$

where P_{t_j} is the stock price at t_j, so this varies over time. This threshold choice is supported by simulation evidence in Table 9 and empirical bootstrap evidence in Table 3, where alternative values of δ are discussed. Basically, when $\delta = s$, the test is severely oversized, and when $\delta \geq s + 2$, its power is diminished. $\delta = s + 1$ serves as an "optimal" point as we expected.

We next give a more formal justification for this choice. Consider the canonical model:

$$Y_t = X_t + \text{noise}_t, \tag{34}$$

where the noise includes both the bid-ask spread s (Example 2; (10)) and a rounding error with tick size τ (Example 3; (11))[4]. As shown in Section 2.2, this model satisfies Assumptions A and C.

Suppose a crossing was triggered at some t_j in the observed price Y. If there was no change in the latent price, i.e. $|\Delta X| = 0$, then the change in the observed price must have been entirely attributable to variations in the microstructure noise. Note that by the triangle inequality, we have

$$|\Delta Y| = |\Delta X + \Delta\text{noise}| < s + \tau. \tag{35}$$

Recall that the condition for a sampling event to occur is $|\Delta Y| \geq \delta$. Therefore, setting

$$\delta \geq s + \tau = s + 1 \tag{36}$$

ensures that the event was triggered by an *actual change in the latent price* X, rather than by pure variations in the microstructure noise. In other words, with this choice, the event could not have been triggered without an actual change in the latent price X. Combining the theoretical lower bound (36) with our earlier practical objective to keep the threshold as small as possible, we arrive at the choice of (32).

5 Simulation Study

In this section we assess the size and power of the daily test[5] DV^C and the local test DV^c. We include the nearest neighbor truncation methods, $MedRV$ and $MinRV$ of Andersen et al. (2012), to compare with DV^C, and the bi-power variation BV employed in ABD, to compare with DV^c. MED, MIN, and ABD (BV) were all star performers in a recent comprehensive review for jump detection methods by Maneesoonthorn et al. (2020). We further adopt subsampled versions of these estimators to reduce MMS noise as recommended by Andersen et al. (2012). For example, MED_{60}, MED_{30} and MED_5 are $MedRV$ estimates sampled every 60, 30, and 5 seconds respectively and subsampled every second; $SBV_{300,30}$ is BV estimates sampled every 5 minutes and subsampled every 30 seconds. By "subsampling" we mean moving the grid forward by certain number of seconds each time, arrive at a new subsample, and then take an average of all subsamples, thus making the estimator robust to noise. For example, if samples are taken every 60 seconds and

[4] Here X and Y are interpreted as price levels, rather than log-prices following the standard convention, so that the spread s and the tick size τ are naturally defined on the price scale.

[5] As noted in Section 4.2, the DV^C test refers to the daily jump test in (28).

subsampled every second then we have 60 subsamples to average from. In addition, we devise a tick-by-tick local test from Christensen, Oomen, and Podolskij (2014) (hereafter COP) based on returns preaveraged over rolling blocks. The preaveraging procedure is designed to mitigate the impact of noise. A bi-power variation estimator is applied on these preaveraged returns to arrive at noise- and jump-robust measures of variance, $PABV$. Local test statistics are given by standardizing preaveraged returns by $PABV$. The block length is determined by the θ parameter. COP recommends θ within [0.5,2], so we consider $PABV_1$ with $\theta = 1$ and $PABV_2$ with $\theta = 2$. A larger θ gives a wider block length.

Following Chernov et al. (2003), Huang and Tauchen (2005), Barndorff-Nielsen et al. (2008) and Hong et al. (2023), the true log-prices are generated by two stochastic volatility models: SV1F and SV2F. Details of the two baseline volatility schemes, as well as procedures to add MMS noise are relegated to Internet Appendix A. Later we will add one jump each day at a random time point to the two baseline models to arrive at SV1FJ and SV2FJ.[6] We have four jump sizes: "extra large" means that the jump variance accounts for 30% of daily integrated variance on average; "large" means jump variance accounts for 20%; "medium" accounts for 10%; and "small" for 5%. 5000 repetitions are performed in each scenario.

Consistent with the empirical data, we initially assume the average inter-trade duration to be 6 seconds (we later examine the detrimental effect of staleness by increasing it to 12 seconds). So we first simulate 23400 points for the underlying latent price process which is equivalent to a 6.5 trading hour session with one observation each second, and then randomly select transaction points based on a Bernoulli sampling scheme of probability 1/6. Previous studies often assume trades arrive every second. For example, Maneesoonthorn et al. (2020) simulate 21600 observations per trading day, equivalent to a 6-hour trading session with an inter-trade duration of 1 second. Clearly more frequent trading benefits all jump tests, but such a setup is far from reality and the role of time-discretization, or staleness, will be diminished. Tables containing all simulation results from this section are in Internet Appendix A.

5.1 Daily test

Under the more volatile SV2FJ model and compared with SV1FJ, misclassification percentages increase while powers decrease across all tests. Our daily test is particularly sensitive to changing volatility, as evident from Equation (16): when the difference between any two continuous returns

[6] We put in Internet Appendix A results for an additional model where a volatility jump component is added to SV1FJ, giving rise to SV1FJJ, where we randomly select a 10-minute interval and double the returns within which is equivalent to a 7.7% volatility jump that is broadly consistent with the size of the volatility bursts we find in Table 4. The performance of different estimators under SV1FJJ is consistent with those under SV2FJ.

a_i and a_j - regardless of their distance - is large, as is the case when one of them comes from volatility bursts, $RV - DV^C$ is inflated, and so does the final daily jump statistic in Equation (26). Indeed, our daily jump statistic includes impact from volatility bursts. Evidently, as we move from a one-factor to a two-factor volatility scheme, our daily statistic is severely oversized, manifesting the increased influence of rapidly changing local volatilities. MIN and MED mitigate jumps by taking the minimum or median of two or three consecutive returns respectively, and thus are more robust to local volatility bursts. Given a 6-second average inter-trade duration, the MED_{30} method performs the best in both size and power, and it will be employed as the main competitor to our daily test in the empirical applications.

MIN and MED become ineffective (size going to 1) when the sampling interval is 5 seconds because they are highly sensitive to zero returns stemming from price staleness or long inter-trade durations. An average transaction duration of 6 seconds and a sampling interval of 5 seconds will render an abundance of zero returns, which severely deflate their integrated volatility estimates and thus inflate the jump statistic, as explained in detail in Kolokolov and Renò (2024). When applying this class of estimators, we will need to stretch the sampling interval as staleness deteriorates to remedy "zeros". We will elaborate more on issues with staleness in Section 5.3.

5.2 Local test

Robust to changing volatilities, the local test proves to be more accurate and powerful than the daily one. Under the more volatile regime of SV2FJ, our local test is only slightly oversized when $\alpha \leq 10^{-3}$; while at 0.01 and 0.05 significance levels, it generates even fewer misclassifications than theoretically granted. In comparison, both $PABV$- and SBV-type tests are severely oversized under SV2FJ and their test sizes go up as sampling intervals decrease, which is a typical effect of staleness on conventional interval-based tests: local test statistics are computed as interval returns divided volatility and an abundance of zero returns will deflate the denominator, thus inflating the entire jump statistic, with more frequent sampling rendering more zeros.

Drastic changes of the local volatility cause more spurious detections by interval-based tests from misclassifying accumulated continuous price changes into large discontinuities when observed only at interval endings. This type of false detections is eliminated by our method due to a minimal threshold value (Equation (32)). Apart from size issues, to interval-based tests, their powers can be diminished when large jumps were undermined by continuous price changes of an opposite sign within the same interval. Despite smaller sizes (fewer misclassifications), our local test proves to be (almost) universally more powerful across all significance levels and jump sizes. The only exception is SBV_{30} but it has become practically ineffective due to high sizes (especially under SV2FJ).

It is interesting to observe how $PABV_1$ performs compared to SBV_{60}. Both tests are correctly sized when $\alpha = 10^{-5}$ or 10^{-4} under the less volatile SV1FJ regime, but the powers of $PABV_1$ are universally lower than those of SBV_{60} and they drop sharply as jump sizes decrease. Obviously, the pre-averaging procedure designed to mitigate noise takes its toll on the testing power, and the problem worsens as the jump size shrinks. But this procedure helps reduce size. Both zero returns and the accumulated price changes benefit from some averaging over the neighborhood to reduce false positives. Evidently, the sizes of $PABV$-type tests are universally lower than those of SBV's.

5.3 Staleness

When we increase the transaction interval to 12 seconds to accommodate staleness, MED_{30} becomes severely oversized under both volatility regimes. Our daily test is roughly equally oversized under the highly volatile SV2FJ scheme, but is much better sized under SV1FJ.

How does staleness affect the local detection methods? We see a universal increase in size across all tests when the average inter-trade duration increases from 6 seconds to 12 seconds, but our local test still maintains low misclassification rates under the less volatile SV1FJ scheme. $PABV$-type methods have a comparably good size as ours, and $PABV_2$ is even better sized than DV^c under SV2FJ. As argued by Kolokolov and Renò (2024), zero returns will in theory be followed by a big price move as compensation, but the pre-averaging procedure can offset this effect. $PABV_2$ has a larger block size than $PABV_1$, so its size is even smaller. Yet as mentioned before, the pre-averaging procedure will undermine the impact of jumps, which can be seen from their even lower powers than those under 6-second inter-trade duration, and the longer the block length the lower the power. $PABV$-type estimators are calculated on business time, i.e. block size is based on the number of observations not the number of seconds. Thus, with increased staleness, the block sizes are actually larger (in calendar time), resulting in the lowered powers.

Our endogenous thresholding approach operates on tick time, i.e. only price changes are likely to bring about price events, so it is robust to the abundant zero returns produced by staleness. Equivalently, under our endogenous sampling scheme, staleness affects the duration between price events rather than generating additional zero-return observations. But the subsequent larger price moves, especially when coupled with drastic changes in volatility, such as those under SV2FJ, could lead to more misclassifications. However, even in these extreme cases, our local test remains the best balance of size and power; and when volatility is more stable (as is often seen when staleness is high), our test is correctly sized with powers close to 100% for large price jumps.

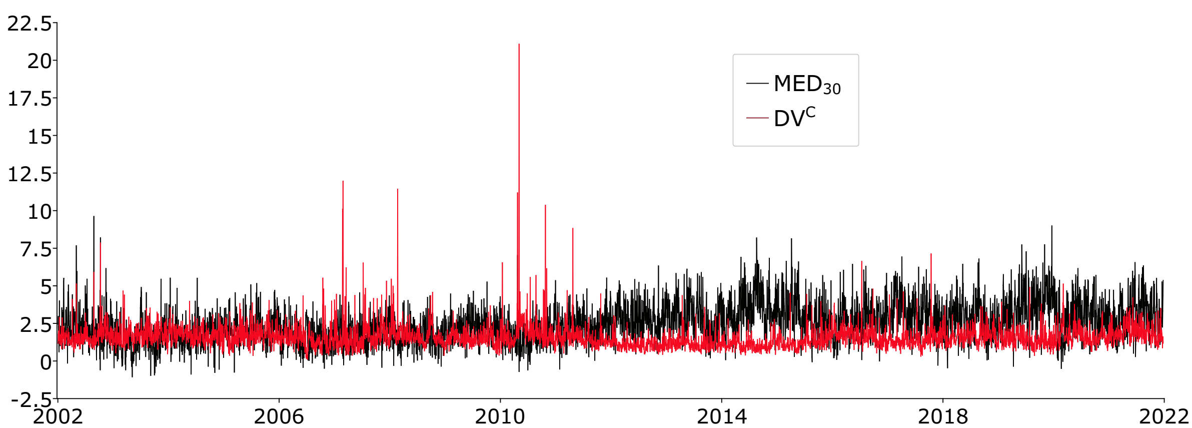
6 An empirical showcase: IBM

Empirically, we first take IBM stock as an example[7] to illustrate the types of misclassifications often produced by interval-based tests and show how our tests perform in these scenarios. Jump tests often disagree on jump days and we focus on the days when discrepancies in the statistics are the largest and visually inspect whether large price discontinuities are present.

6.1 Daily test

We take the best performer, MED_{30}, from Section 5, as comparison. Figure 1 plots the daily variations of jump statistics for MED_{30} and DV^C. There are three things to notice at a first glance: 1) the highest point with a jump statistic over 20 rendered by our daily test is the flash crash day, which will be analyzed in Section 6.2; 2) MED_{30} can generate negative jump statistics while our daily statistics are guaranteed to be positive since $RV > DV^C$ by construction; 3) MED_{30} jump statistics increase drastically and are much higher than ours from around 2012 onwards. We analyze the last point later and for now turn to the first half of the sample period and examine the two days which exhibit largest differences (apart from the flash crash day) in jump statistics: 26th February 2008 and 27th February 2007.

Figure 1: Daily test statistics: DV^C and MED_{30}



Figures 2 and 3 plot the intraday price changes as well as the whole price paths on those two days. Large price discontinuities are clearly present, correctly detected by our daily test with significant Z values of 11.43 and 10.08 respectively. The reason MED_{30} misses the jumps in Figure 2 is that it was a sequence of three consecutive jumps of sizes 0.42, 0.23 and 1.1 respectively, equivalent to 8.4, 4.6, and 22 times of the threshold δ_p. In this case, taking the median does not diminish the influence

[7]Details on data cleaning are relegated to Internet Appendix A

of jumps. Our test, on the other hand, averages jumps over all observations to arrive at the noise-included variance estimate DV^C, which is then fairly robust to small number of jumps. In Figure 3, we can see two sets of reversing jumps, roughly 0.34 and 0.5 in size respectively, equivalent to 8.5 and 12.5 times δ_p. Such reversal-type jumps are particularly difficult for interval-based tests to detect since they would have disappeared before observed at the end of a sampling interval. Under our endogenous thresholding approach, jumps trigger price events immediately as they occur, and feed into the final jump statistics with their entire impact preserved. On both days price jumps are followed by local bursts of volatility. Note that the reversal jumps discussed in this study are not the "bounce-back outliers" mentioned in Aït-Sahalia et al. (2011) which refer to pairs of large price changes of opposite signs but exactly the same size - the reversal jumps we discovered are of different sizes.

Figure 2: Price change and price paths on 26Feb2008: $DV^C(Z = 11.43)$ signals jump presence while $MED_{30}(Z = 0.35)$ does not - false negatives with consecutive jumps.

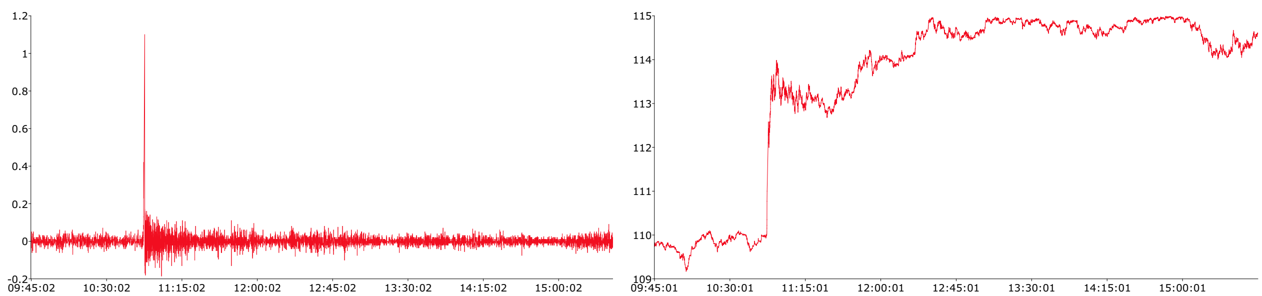
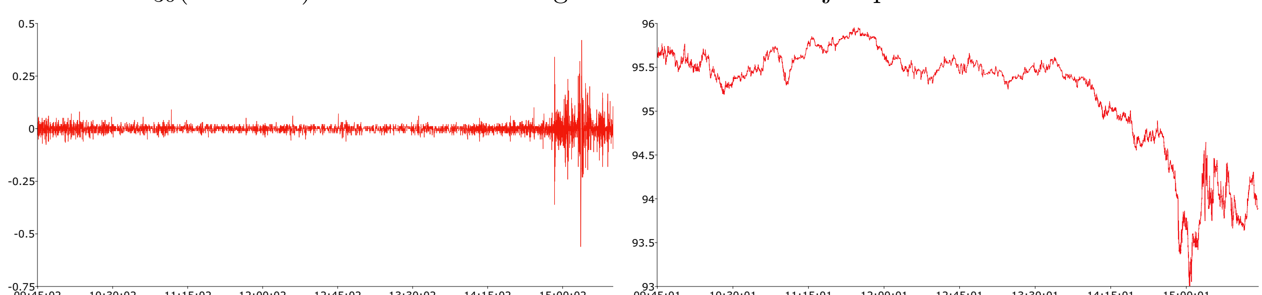


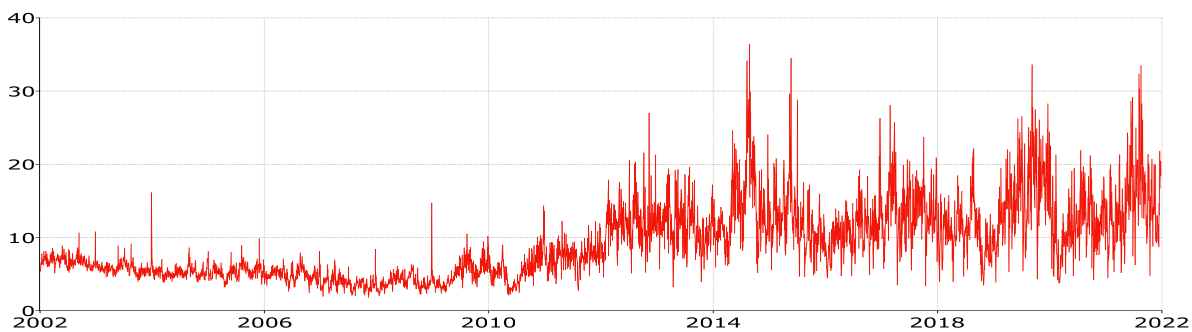
Figure 3: Price change and price paths on 27Feb2007: $DV^C(Z = 10.08)$ signals jump presence while $MED_{30}(Z = 0.35)$ does not - false negatives with reversal jumps.



6.1.1 Staleness

So what happened to the MED_{30} test statistics during the second half of the sample period? The answer lies in Figure 4: IBM's average inter-trade duration more than doubled to 12.76 seconds during 2012 to 2021, from an average 5.39 seconds during 2002 to 2011. As carefully elaborated by Kolokolov and Renò (2024), staleness intensifies the spurious detection problem faced by conventional interval-based jump tests due to their under-estimates of the integrated variance with the abundant zero returns produced by sparse sampling. One way to circumvent the staleness problem is to decrease the sampling frequency, which will unfortunately lead to more misclassifications of volatility and drift bursts into large price jumps, as we will see in Section 6.2.

Figure 4: Inter-trade duration (in seconds), IBM: severe staleness since 2012



We show through simulation how our tests are robust to staleness. For the daily test, as the average annualized integrated variance more than halved from 0.039 during 2002 to 2011, to 0.018 during 2012 to 2021, we can see an increased staleness is coupled with a decreased volatility. In that case, the effect of an increased staleness on our daily test is limited by a lower volatility and our test renders fewer misclassifications in simulation especially when the significance levels are low, i.e. $\alpha < 10^{-3}$. We can see this pattern in Figure 1: our test delivers fairly stable statistics across the whole period, as opposed to a sharp increase by MED_{30} when an increased staleness kicks in.

6.2 Local test

The local test aims to accurately detect price jumps as they occur and precisely pin down their locations. We first look at the accuracy issue by comparing our DV^c with SBV_{300}. We choose SBV_{300} as comparison as it maintains the lowest misclassification rate among all competing methods as evidenced by simulation, and a good robustness to staleness, as seen from its comparable size to our local test when time discreteness increases to 12 seconds. Figure 5 plots the jump variation (JV) statistics for SBV_{300} and DV^c when $\alpha = 10^{-5}$. Both the jump days and JV values differ between the two tests, and we select three days when the JV differences are among the largest[8].

[8] Plots for the three days, 30Jul2018, 14Nov2014, and 25Aug2008, when JV's by SBV_{300} are 62%, 51%, and 46% respectively, are of a similar pattern and are available from the authors upon request.

Figure 5: JV's for two local tests: DV^c and SBV_{300}

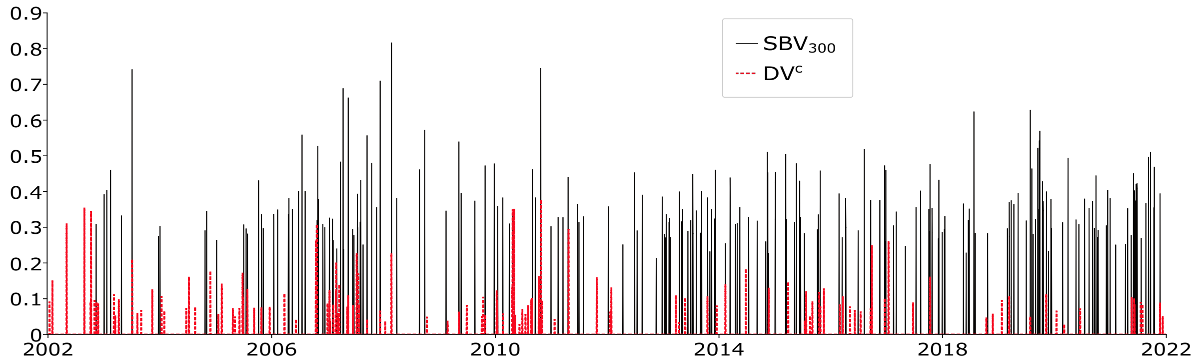


Figure 6: Price change and price paths on 16Oct2006: DV^c signals jump presence with $JV = 26\%$ while SBV_{300} does not - false negatives with reversal jumps.

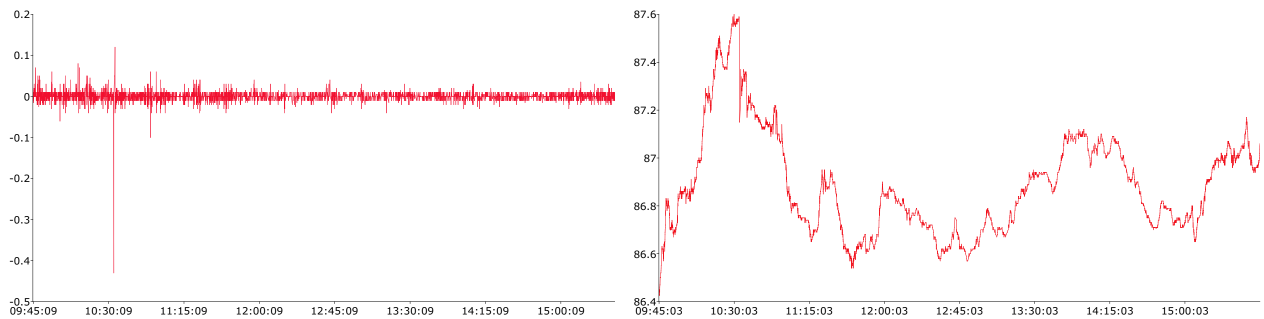


Figure 6 presents a typical day when our local test detects a big price jump with JV of 26%, while SBV_{300} does not detect any. From the price change plot it is easy to see a big negative price jump, and the price path on the right shows its being partially reversed by subsequent positive price changes. Observed at a 5-minute frequency, that reversal jump simply "disappeared".

Figures 7 and 8 plot two days when SBV_{300} identifies jumps and reports JV values over 50%, while DV^c does not detect any. Looking at the price path graphs on the left, in Figure 7, the stock price rapidly decreases by over 1 dollar around noon; while in Figure 8, around market open, the stock price first rises rapidly by 0.9 dollars and then drops by about 0.5 dollars. Are they true price discontinuities? Not really. Evident from the price change plots on the right, in both cases, the price changes constituting those "jumps" never exceed 15 ticks and are hardly distinguishable from adjacent price moves. Accordingly, the discontinuities detected by SBV_{300} are actually continuous price changes of roughly the same direction that cumulate into large price moves when observed at sparse sampling intervals. The misclassifications of drift bursts, as in Figure 7, and volatility bursts, as in Figure 8, into price jumps are typical of interval-based tests, and have been documented by Bajgrowicz, Scaillet, and Treccani (2016) (hereafter BST), COP, and Christensen, Oomen, and Renò (2022), among others. Our local method, in contrast, tests each price event individually, based on a low threshold that effectively stops continuous price changes from (overly) accumulating.

Figure 7: Price and price change paths on 11Aug2016: SBV_{300} detects 1 jump with $JV = 52\%$ while DV^c detects none - false positives from drift bursts.

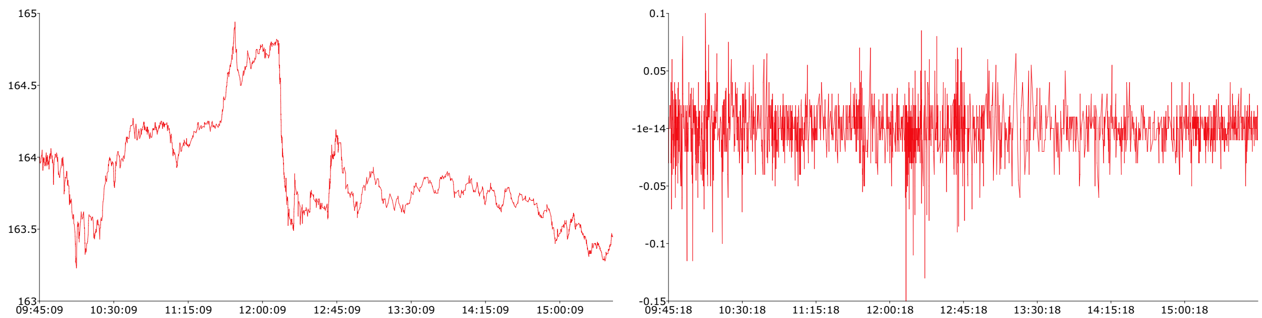
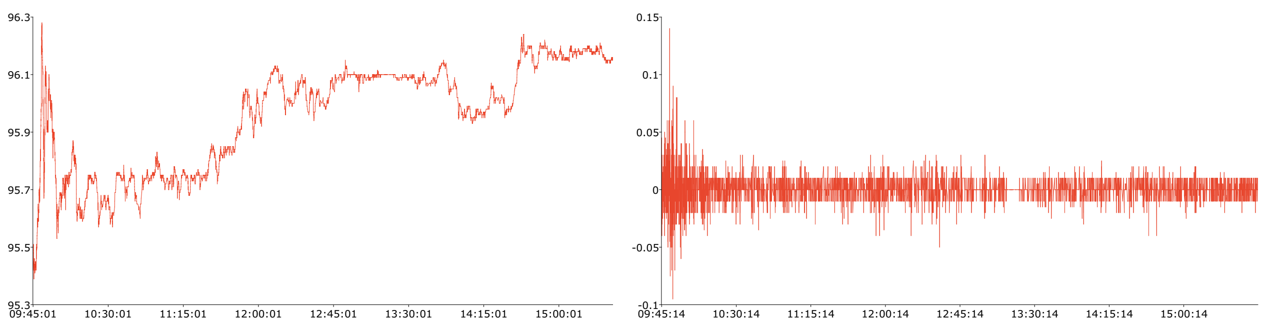


Figure 8: Price and price change paths on 16Apr2007: SBV_{300} detects 2 jumps with $JV = 69\%$ while DV^c detects none - false positives from volatility bursts.

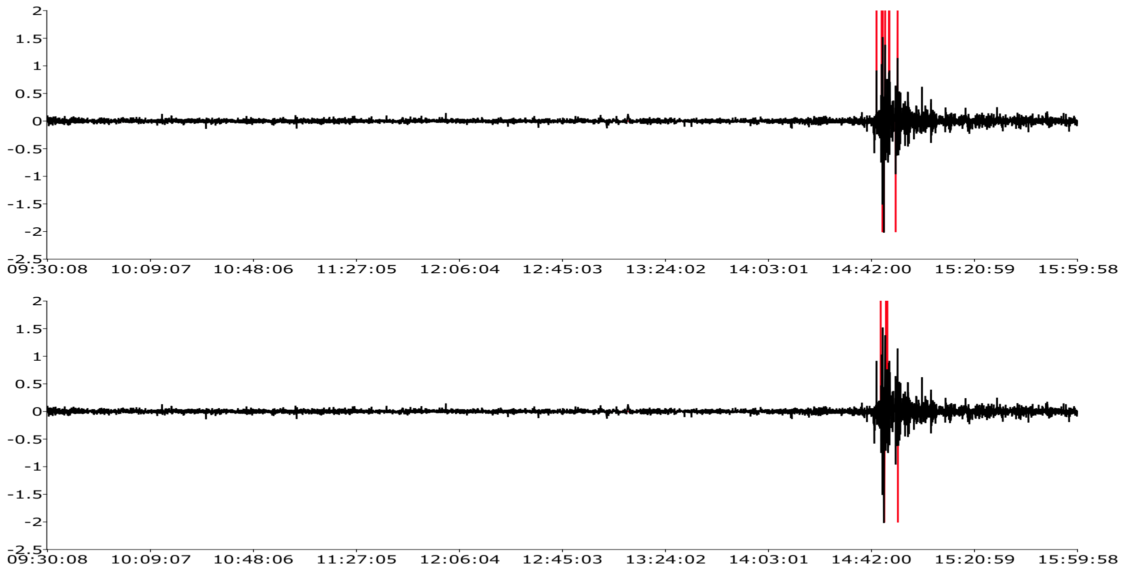


6.2.1 The Flash Crash

Apart from size and power of price jump tests, the precise localization of price jumps is equally important. On the one hand, immediate detections largely alleviate the misclassification problem; on the other, a precise jump location is the first step to either estimate other important risk factors, such as a high-frequency financial leverage, see Bibinger et al. (2019), or applications where precise matching is required: examples include a study on the origin of jumps as in Section 8, and the immediate impact of jumps on high-frequency option pricing.

Our endogenous thresholding approach naturally guarantees precise calibrations of price jumps to the finest time resolution as permitted by data availability because a price jump will trigger a price event immediately, and then a local jump test can be performed. We choose SBV_{30} as

Figure 9: Price changes on the flash crash day: DV^c detects 10 jumps (top plot) and SBV_{30} detects 5 (bottom plot), with $\alpha = 10^{-5}$.



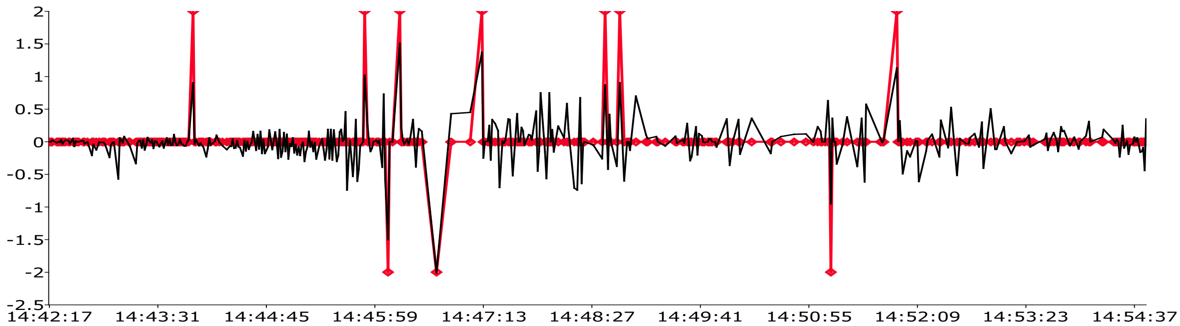
comparison here since, though severely oversized, it gives the highest power among all competing local detection methods in simulation.

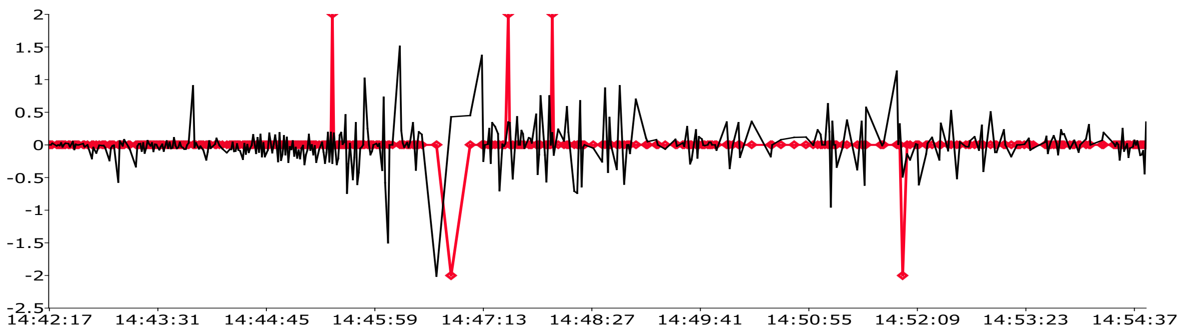
We focus on the historical incident of flash crash that occurred on the 6th May 2010, which has been widely studied in the literature, see e.g. Kirilenko et al. (2017) and Menkveld and Yueshen (2019). Figure 9 highlights price jumps identified by DV^c and SBV_{30}: 10 by the former and 5 by the latter. Jumps of smaller sizes can be detected by our method due to the low threshold value. Apart from missing some jumps, SBV_{30} detects them with time lags, as more clearly seen in Figure 10 where we zoom in on the price changes during the flash crash period. While our local test can pin down price jumps as they occur, interval-based tests will identify only the intervals that contain jumps without calibrating their occurrences to a higher time precision. It is also interesting to see how SBV_{30} missed the second and third jumps: they are a pair of reversal jumps. Lastly, we note that the flash crash incident involves both price jumps and volatility/drift bursts and they are clearly intertwined. Our local test can separate the most prominent price jumps from bursts.

7 Comprehensive assessment: DJIA30 constituents

In this section, we assess the power of our tests with real data, use this to give a simple proxy of volatility bursts, and finally summarize reliable estimates of jump frequency and variation. We apply 20 years of ultra-high-frequency data for the 30 Dow Jones Industrial Average (DJIA) index

Figure 10: Zoom in on jumps identified by DV^c (top plot) and SBV_{30} (bottom plot).





constituents from 2002 to 2021, obtained from the TAQ database and are time-stamped to a second.[9] Apart from the data cleaning procedures outlined in Internet Appendix A, it is customary to remove data around market open and close due to abnormal trading activities. We follow Wang and Mykland (2014) and remove data within the first and last 15 minutes of a daily trading session and use ultra-high-frequency data from 9:45 to 15:45 only. The purpose of this procedure is not to alleviate the intraday seasonality pattern which we merely treat as abnormal changes in local volatility, the same as volatility bursts. As evidenced by the simulation results, our local tests are robust to drastic changes in local volatility, but the effect of volatility bursts will be included in the daily test statistics.

7.1 Empirical bootstrap test for power

In this section, we put all methods, daily and local, through a rigorous test for power, by the empirical bootstrap method of Bollerslev et al. (2008). First we select days when none of the tests detects a jump with $\alpha = 10\%$ for MED-type and our tests, and $\alpha = 10^{-5}$ for $PABV$- and SBV-

[9] Since the DJIA composition is constantly changing and our data was obtained in two batches, the end of 2012 and the end of 2021, not all stocks have a complete 20 years of data. Out of the 30 stocks included in our sample, AXP, BA, CAT, DIS, GE, HD, IBM, JNJ, JPM, KO, MCD, MMM, MRK, PG, WMT and XOM have complete data; AA, DD and UTX use data from 2012 and 2021 since they were no longer constituents by the end of 2021; CVX, GS, NKE, PFE, TRV, UNH, V, VZ, CSCO, INTC, and MSFT use data from 2013 to 2021 since they were not constituents at the end of 2012 when we first collected data.

type tests. Thus the sizes of the tests at significance levels equal or lower are zero. We call them no-event days. Then, we add one jump each no-event day, with a fixed size of 6, 8, 10 or 12 times the average spread of the day (s) to random time points to see whether the tests can detect it. We consider two scenarios: a single jump and a jump that reverses gradually in four consecutive steps, each of the size $2s$.

Tabulated in tables 1 and 2 are summary results, presented as averages over all stocks.[10] In each scenario, the powers of our local test go to 100%, as the jump sizes increase from $6s$ to $12s$ and α goes to 5%. Our daily test, which is sensitive to volatility variations, shows less power - 80% when jumps are large and at 5% significance level. The median and mean of the average daily spread among all stocks (over the no-event days included for this empirical bootstrap) is 1.73 ticks and 2.8 ticks, so the largest jump size, $12s$, is around 0.21 to 0.34 dollars on average. The power of our local test for this jump size ranges between 74% and 100%, regardless of significance levels and jump types, i.e. single or reversal. Evidenced by simulations, our local test is not oversized even at $\alpha = 5\%$. As the jump size decreases, the power goes down, and when it reaches $6s$, which is about 10 ticks, the detection rates are well below 25% regardless of α values.

For all the interval-based daily and local tests presented in Table 2, their powers depend crucially on the type of jumps they encounter. When there is a single price jump, their powers are comparably much lower than our local test, and when the jumps tend to reverse, as often observed in real data and documented by COP, the conventional tests become largely ineffective. This pattern, however, is broken by SBV_{60}: instead of a sharp drop in power regardless of jump sizes and significance levels, the power "gaps" between "singles" and "reversals" only gradually grow larger with increasing jump sizes, which is a manifestation of the misclassification problem faced by SBV_{60}. As shown in simulations, compared with other tests, SBV_{60} is severely oversized. The power from detecting small price discontinuities, i.e. $6s$, is driven largely by false positives. As the jump size goes up, the proportion of power generated by correct identifications increases, and so does the "gap".

Table 3 provides empirical bootstrap evidence (jump size $= 10s$) on threshold selection. $\delta = s{+}1$ provides the best balance of size and power: a lower value causes severe misclassification (over-size) problem and a higher threshold damages testing power.

7.2 Comparison of DV^C and DV^c: a proxy for volatility bursts

In this subsection, we will first examine to what extent our local and daily tests agree on jump days, and then study the meaning of their JV differences (see (29) and (31)). In Table 4 we tabulate

[10] Detailed results for individual stocks are in Internet Appendix A or available upon request.

in the first row the common jump proportions as the number of common days over the number of DV^c identified jump days. The denominator is chosen this way since the local test isolates price jumps more precisely. The second row records the average JV differences on common jump days, presented again as averages over all 30 stocks.[11] The values chosen for α^c (for the local test DV^c) and α^C (for the daily test DV^C) are the same as in simulations, with the latter generally higher to account for the inclusion of volatility burst effects in the daily statistic. When $\alpha^C = \alpha^c = 0.01$, the common detection proportion is only 0.58, i.e. 42% of the jumps identified by the local test at 1% significance level are not recognised by the daily test, though local jumps under $\alpha^c = 0.01$ are not typically large. As α^c goes down and α^C goes up, the common rates are higher, reaching 100% and no lower than 96.6% when the former is 10^{-5} or 10^{-4} and the latter is 5% or 10%.

What does the difference in JV mean? We interpret it as a proxy for volatility bursts, or more generally, abnormal local changes in volatility. This interpretation is suggested by (16) and formalized by Proposition 1. Let $IV_t := \int_0^t \sigma_s^2 ds$ and $Q_t := \sum_{s \le t} U_s^2$. Since Proposition 1 implies that RV_t and DV_t^C converge in probability to $\xi_{2,t}(IV_t + Q_t)$ and $\xi_{1,t}^2 IV_t$, respectively, the daily jump-variation measure satisfies

$$JV_t^d = \frac{RV_t - DV_t^C}{RV_t} \xrightarrow{p} \frac{Q_t}{IV_t + Q_t} + \left(1 - \frac{\xi_{1,t}^2}{\xi_{2,t}}\right) \frac{IV_t}{IV_t + Q_t}. \tag{37}$$

The first term is the discrete-jump contribution, while the second term is a continuous event-heterogeneity component. This second component is zero when normalized event returns are homogeneous, and becomes positive when event-time returns are locally dispersed, as occurs during short-lived periods of abnormally high local volatility. Since the local jump variation JV_t^l is designed to isolate locally identified discrete jumps and thus accounts for the first component, the difference $JV_t^d - JV_t^l$ provides an empirical proxy for abnormal local volatility variation.

As shown in Table 4, the average JV differences stay surprisingly stable across varied combinations of significance levels and individual JV differences in Internet Appendix A follow the same pattern: the variations in average volatility burst proportions typically stay within 1% to 2% across different jump sizes. This is no coincidence and points to the fact that volatility bursts account for a steady 7% to 8% of daily RV on average.

7.3 Jump frequency and variation

Let us now summarise in Table 5 the jump frequency (JF, number of jump days over total number of trading days) and jump variation (JV, calculated by the formulae in (29) and (31)) statistics which

[11] Results on individual stocks are in Internet Appendix A

are the focus of many prior studies. Presented here are averages over all 30 stocks; detailed results for each stock are reported in Internet Appendix A. We set our local test DV^c, which proves to be both accurate and powerful, as the benchmark and add a universal threshold (UT) significance level by BST, calculated as $\alpha_u = \Phi(-\sqrt{2logN})$, which asymptotically reduces the false discovery rate from multiple testing down to zero. Notice that $\alpha = 10^{-5}$ is even tighter than UT in our case.

JF increases with α while JV decreases with it. DV^c under $\alpha = UT$ gives an average 3.6% JF, equivalent to one jump every 5.6 weeks, and 12% JV, or a 0.4% (= 3.6% × 12%) when averaged over all trading days. Simulation evidence shows that our local test is (severely) undersized, and higher α's up to 5% increase false discovery rates to only 0.3% under SV1FJ and 1.6% under the more volatile SV2FJ, but higher α's will lower the average size of detected jumps. At $\alpha = 5\%$, JF increases to once every 7 trading days but JV decreases to 7.8%, signalling the inclusion of more medium-sized jumps. Yet the average JV over all trading days is still only 1.08% ($= 0.139 \times 0.078$). This result is consistent with COP who find JV accounts for only 1% of total RV (QV) on average. Yet our results emphasize that price jumps remain an important risk factor. On jump days, large jumps account for over 12% of total RV on average, and JV remains over 7% as jump sizes get smaller. We also confirm that jumps are indeed rare, with JF ranging from once every 6.45 weeks ($\alpha = 10^{-5}$) to every 6 trading days ($\alpha = 0.1$), depending on the jump size.

$PABV_1$ finds an average 4.8 ($= 0.019 \times 252$) jumps per year (with $\alpha = 10^{-5}$), which is exactly the same result as by BST who also uses a pre-averaged version of the Bipower variation estimator, though applied on a different dataset between 2006 and 2008. This number is lower than our result with DV^c (at $\alpha = 10^{-5}$ DV^c still finds 7.8 jumps per year on average). This discrepancy is not surprising. Evident from the simulation results and the empirical power test, $PABV$-type methods maintain low false detection rates at the expense of testing power due to a pre-averaging procedure: the average block lengths given by Δ is about 6.6 minutes for $PABV_1$ and 13.2 minutes for $PABV_2$. SBV_{300} maintains a low size as well, thus generating a reasonable JF of 4.9%, but its JV goes to 36.5%, tripling that by DV^c and obviously inflated by volatility bursts due to sparse sampling at a 5-minute frequency.

Our daily statistics also include the influence of volatility bursts. A low significance level, such as $\alpha = 0.1\%$, makes DV^C-detected days more likely to contain genuine price jumps, apart from volatility bursts. At $\alpha = 0.1\%$, DV^C gives a 5.6% JF and a 18.8% JV. With α between 10^{-4} and 10^{-3}, our local test renders comparable JF's between 4.1% and 5.9%, which when matched with $JF = 5.6\%$ gives a linearly interpolated JV around 10.42%. By linearly interpolated JV, we mean the local-test JV evaluated at the daily-test JF by linear interpolation between the two nearest local

JF,JV points. The difference in JVs, at an equal JF, is 8.38% (=0.188-0.1042), which represents the influence of bursts, is consistent with the results presented in Table 4. MED-generated JF's and JV's, on the other hand, are heavily inflated by staleness, similar to those by SBV_{60}.

Lastly, we have performed a runs test as in Scaillet et al. (2020) on the 20 years of jump detections, to assess jump and burst clustering behaviors. Results are in Internet Appendix A. We find some clustering behavior (about 1/3 of stocks under examination) from large price jumps, and prevalent volatility burst clustering[12], consistent with the classic volatility clustering phenomenon. False positives can yield an invalid conclusion of jump clustering as discussed in BST. Indeed, we find price jumps cluster more as their sizes go down (and thus are mixed up with bursts).

8 A study on the origins of jumps and bursts

Numerous studies have investigated how news announcements can be related to the occurrence of price jumps; see, among others, Evans (2011), Boudt and Petitjean (2014), Aït-Sahalia and Xiu (2016), Bollerslev et al. (2018), Gürkaynak et al. (2020), Jeon et al. (2022), Erdemlioglu and Yang (2023), Aït-Sahalia et al. (2024). Many studies indeed find that news induces price jumps. However, as Bajgrowicz et al. (2016) (BST) pointed out, most price jumps identified by conventional interval-based tests are spurious and are more likely to be volatility bursts instead. They further find that a majority of news releases do not cause jumps but generate volatility bursts which were loosely defined as jumps detected at low frequency but not at high frequency.

We have constructed an empirical proxy for volatility bursts in Section 7.2 as the difference between our daily and local jump statistics. Following the same logic, we proxy volatility bursts by our daily statistics after controlling for our local statistics, in order to study their origins. We confirm the conclusion by BST that news causes volatility bursts. With regard to price jumps, we arrive at a similar conclusion as with Boudt and Petitjean (2014) that spread and volume are the key drivers of price jumps, but we do not discover an amplifying effect from news. However, Boudt and Petitjean (2014) applied the interval-based test of Lee and Mykland (2008), though at a relatively high frequency of 2 minutes, so it is likely that their findings on jumps involve volatility bursts as well. This concern brings about the unique and effective structure of our regression models in Section 8.2, where we separate the causes of jumps from those of bursts by controlling for the latter in the regression for the former (and vice versa), so that the differing degrees of influence from news sentiments and trading variables on the two extreme risk factors are untangled. Otherwise,

[12] In assessing volatility burst clustering behavior, we remove price jumps first, and then apply the daily tests on the remaining returns and prices.

both the preliminary analysis in Section 8.1 and the simpler version of regression models in Section 8.2 would suggest news and high-frequency trading are both key drivers of jumps and bursts - a similar conclusion as in Boudt and Petitjean (2014).

Importantly, in performing regressions, we keep only days with either jumps (local test at 0.1% significance) or bursts (daily test at 10% significance)[13], removing all no-event days to ensure effective interpretations of regression results. The numbers of days remaining for each stock are included in Table 7.

8.1 Preliminary analysis

We collect news information from the Dow Jones Edition of the RavenPack database for equities. 4[14] of the 30 stocks miss large chunks of news data so we have 26 stocks left in this section. To assess the immediate impact of news on jumps and bursts, we retain only news released within trading hours, so news on weekends and holidays are removed. Additionally, due to three abnormal spikes of news release at 15:41, 15:46, and 15:51, news released within the last 20 minutes near the market close, that is, after 15:40:00, is removed[15]. Correspondingly, trading data within the last 20 minutes of the trading session are removed as well, together with the first 15 minutes after market open (same as in other empirical sections), that is, before 9:45, due to possible drastic price moves resulting from news accumulated during previous non-trading hours.

RavenPack composes Relevance (REL), Event Sentiment Score (ESS), and Event Novelty Score (ENS) information for all news. Following Shi et al. (2016) and Reed et al. (2020), we retain only the 100% relevant news ($REL = 100$) for each firm and compose a news sentiment score (NSS) for the ith news release during the day as below:

$$NSS_i = \frac{ESS_i - 50}{50} \cdot \frac{ENS_i}{100}, \tag{38}$$

given that ESS centers around 50 with positive news scoring above 50 and negative news below 50. ENS measures how "novel" the news is. Given that multiple sources may release the same piece of information at different times, only the first release gets a score of 100. However, we view repeated mention as a sign of importance, which may induce larger increase of volatility.

Further, to measure the total news impact during day t, we take the sum of absolute NSS_i to arrive at the absolute daily sentiment (ADS):

$$ADS_t = \sum_{i=1}^{W} |NSS_i|, \tag{39}$$

[13]For those stocks without a complete 20 years of data, we use 1% significance level for local tests.
[14]AA, DD, TRV, and UTX.
[15]The average absolute sentiment scores are 0.22, 0.12, and 0.21 respectively, for periods 9:30 to 15:40, 15:41 to 16:00, and weekends and holidays.

where W is the total number of news releases during day t. Here we do not distinguish between positive and negative sentiments, since both can induce price jumps and volatility bursts.[16]

Table 6 summarizes average values of news and trading variables based on jump identifications by our daily and local tests, including MED_{30} as comparison. S_j is the average absolute daily sentiment (ADS_t) score on jump days and S_n on no-jump days. For high-frequency trading variables, we consider volume (v), order imbalance (o), and bid/ask spread (s), where order imbalance is calculated as the absolute difference between bid size and ask size. We take percentage differences between the "local" and daily average values of the above three variables, with "local" defined as within 60 seconds before, either the occurrence of a local price jump identified by DV^c with $\alpha^c = 10^{-5}$, or the largest price event during a jump day identified by DV^C or MED_{30} with $\alpha^C = 0.1\%$, resulting in dv, do, and ds, where $dv = \frac{v^l - v^d}{v^d}$ for example, with v^l and v^d being the local and daily average volumes; do and ds are calculated similarly.

The results are very interesting. When classified by MED_{30}, average news sentiment scores are not significantly different between jump and no-jump days, while by our daily and local tests, average S_j's are 54.5% and 84.4% higher than average S_n's. Specifically, only 23.1% stocks produce higher jump-day sentiment scores when classified by MED_{30}, while the number increases to 92.3% by our tests. High-frequency trading variables show the same pattern, averaging much larger than 1 and evidencing that trading activities are elevated before jumps. 61.5%, 65.4% and 57.7% stocks have higher than unity dv, do and ds by the daily test, while they rise to 76.9%, 76.9% and 84.6% by the local test: average values by the local test almost double those by the daily one. Additionally, the average news sentiment score is 22% higher on local jump days than on daily jump ones.

Overall, it seems both news and high-frequency trading exert more influence on jumps than on bursts (proxied by the daily statistics). We need to resort to regression analysis to distinguish their differing influences on the two risk factors.

8.2 Regression analysis

In applying our daily and local jump statistics to regression models, we refrain from subjective classifications but instead use the original Z statistics directly; as to local jumps, we use the highest absolute Z value within the day, which is the largest standardized price event return (in absolute value). We remove days where neither local nor daily test statistic is significant. Additionally, due to the abundant no-news days, we retain only days with non-zero news sentiment scores.

[16] In Internet Appendix A, we perform additional regressions for the daily statistic, and separately for positive and negative local statistics, where we distinguish between positive and negative news and also test for the equality of their coefficients. We do not find consistently significant asymmetric news or jump effects for the stocks under examination.

The four regressions are abbreviated as below:[17]

$$Z_t^{l(d)} \;=\; \beta_0 + \beta_1 S_t + \beta_2 dv_t + \beta_3 do_t + \beta_4 ds_t + (\beta_5 Z_t^{d(l)}) + \beta_6 Vol_t + \epsilon_t, \tag{40}$$

where Z_t^d and Z_t^l are the absolute values of daily and local jump test statistics respectively, $S_t = ADS_t$, $dv_t = \frac{v_t^l - v_t^d}{v_t^d}$, $do_t = \frac{o_t^l - o_t^d}{o_t^d}$, $ds_t = \frac{s_t^l - s_t^d}{s_t^d}$, and Vol_t is a control variable for the daily integrated variance calculated using the NPDV estimator in Hong et al. (2023) with a threshold range of 2 to 4 times the spread. The standardized daily and local statistics used here Z_t^d and Z_t^l are both $O_p(1/\delta_n)$ under the jump alternatives.[18] We therefore use them as standardized signals to compare driver patterns across the daily and local regressions. Note that integrated volatility is free of the influence of both MMS noise and jumps. In brackets on the right-hand side is a control variable added for the second setup, for example, we control for Z^l in the second regression for Z^d.

Table 7 presents the coefficient t values of the two regressions for Z^d: on the left are t values for the basic setup (without the variable in brackets in Equation (40)), and on the right Z^l is added as an additional control variable. Results in the left panel are roughly consistent with the preliminary analysis in Section 8.1 that both news and trading variables, especially bid/ask spread, contribute to the daily jump statistic which includes both price jumps and volatility bursts. This is consistent with the findings of Jeon et al. (2022) and Baker et al. (2024), who report the influence of news on large daily price moves. After controlling for local jumps using our local test statistic Z^l, however, coefficients for high-frequency trading variables plummet, while the news sentiment coefficients remain significant on average. Specifically, the average volume coefficient turns significantly negative, average order imbalance coefficient becomes insignificant, and the average spread coefficient drops the most. News coefficients drop too, but 96.1% remain positive and 57.7% still significant at 5% level.

Two conclusions can be drawn here. First, news sentiment is the main driver of volatility bursts, proxied by our daily statistic after controlling for the local; second, trading variables, especially volume and spread, are highly correlated with local jumps: the addition of Z^l causes a multicollinearity problem and "absorbs" the variables that are most closely related.

Further evidence is provided in Table 8, which presents regression results for Z^l. The left panel shows the no-control case and consistent with Section 8.1, all news and trading variables contribute significantly to local jumps. After adding volatility bursts, proxied by Z^d, as control, the average news coefficient drops to negative. Volume and bid/ask spread coefficients remain

[17] Regression coefficients for interaction terms between news and trading variables are insignificant so interaction terms are removed. Regression results with interaction terms are available from the authors upon request.

[18] Additionally, we include in Internet Appendix A an illustrative IBM example comparing the two standardized statistics. In the example, they are comparable in scale and exhibit substantial agreement on jump days, consistent with Table 4.

significant on average, with the former 65.4% and the latter 42.3% significant individually. Note that in the no control case, integrated volatility does not contribute significantly to local jumps (unlike bursts), and when volatility burst is added as control, its coefficient turns significantly negative. As expected, volatility bursts are a part of integrated volatility while jumps are not.

In all, we find news sentiment is more related to volatility bursts while trading activities, represented by volume and spread, are the main drivers of jump occurrences.

9 Conclusion

This paper develops an easy-to-implement approach to separate price jumps and volatility bursts from ultra-high-frequency data in the presence of noise and staleness. We utilize endogenous sampling via a threshold on the price dimension, enabling timely localization of large discontinuities. We provide practical guidance for choosing the threshold and large-sample theory for the proposed statistics. Simulation and empirical evidence show that our local jump test attains high power with low misclassification, reducing false detections commonly faced by interval-based tests. For DJIA stocks, large jumps are infrequent, occurring about once every one to two months, yet on jump days they account for about 12% of daily RV on average and a meaningful share of quadratic variation.

Our daily statistic reflects both price jumps and volatility bursts. Comparing daily and local jump statistics, we find that volatility bursts account for a stable 7% to 8% of daily RV on average. By controlling for local price jumps in our regression model, we find news is more closely related to volatility bursts, whereas high-frequency trading variables, especially volume and bid/ask spread, are the main drivers of price jumps. Our results are consistent with the findings of Bajgrowicz et al. (2016), and complement existing research (e.g., Gürkaynak et al. (2020), Jeon et al. (2022), Baker et al. (2024)) that examine large price swings in low frequency contexts by providing additional micro-level insights. For example, what appears to be a single jump at low frequency may break down into intraday bursts or rapid reversals. While we provide evidence on the origins of these two risk factors, their implications for asset pricing and forecasting remain topics for future research.

By distinguishing price jumps from volatility bursts, our approach provides a useful tool for understanding and managing extreme market events, with applications in trading and market making, risk management, and asset pricing. It may also inform market design (e.g., circuit breakers) and market surveillance, including the detection of insider trading or potential price manipulation. Using this distinction in intraday risk monitoring, algorithmic trading, or policy event studies may help practitioners and policymakers assess sudden price swings and better interpret market dynamics.

References

Aït-Sahalia, Y., J. Fan, R. J. Laeven, C. D. Wang, and X. Yang (2017). Estimation of the continuous and discontinuous leverage effects. *Journal of the American Statistical Association 112*(520), 1744–1758.

Aït-Sahalia, Y., J. Fan, and Y. Li (2013). The leverage effect puzzle: Disentangling sources of bias at high frequency. *Journal of Financial Economics 109*(1), 224–249.

Aït-Sahalia, Y. and J. Jacod (2014). *High-Frequency Financial Econometrics*. Princeton University Press.

Aït-Sahalia, Y., C. Li, and C. Li (2024). So many jumps, so few news. Working paper, National Bureau of Economic Research. NBER Working Paper Series.

Aït-Sahalia, Y. and P. A. Mykland (2003). The effects of random and discrete sampling when estimating continuous-time diffusions. *Econometrica 71*, 483–549.

Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics 160*(1), 160–175.

Aït-Sahalia, Y. and D. Xiu (2016). Increased correlation among asset classes: Are volatility or jumps to blame, or both? *Journal of Econometrics 194*(2), 205–219.

Andersen, T. G., T. Bollerslev, and D. Dobrev (2007). 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. *Journal of Econometrics 138*(1), 125–180.

Andersen, T. G., D. Dobrev, and E. Schaumburg (2008). Duration-based volatility estimation. Working Paper, Northwestern University.

Andersen, T. G., D. Dobrev, and E. Schaumburg (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics 169*(1), 75–93.

Bajgrowicz, P., O. Scaillet, and A. Treccani (2016). Jumps in high-frequency data: Spurious detections, dynamics, and news. *Management Science 62*(8), 2198–2217.

Baker, S. R., N. Bloom, S. Davis, and M. Sammon (2024). What triggers stock market jumps? Working paper, National Bureau of Economic Research. NBER Working Paper.

Ball, C. A. (1988). Estimation bias induced by discrete security prices. *The Journal of Finance 43*(4), 841–865.

Bandi, F. M. and J. R. Russell (2008). Microstructure noise, realized variance, and optimal sampling. *The Review of Economic Studies 75*(2), 339–369.

Barndorff-Nielsen, O. E., P. Hansen, A. Lunde, and N. Shephard (2008). Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica 76*(6), 1481–1536.

Barndorff-Nielsen, O. E. and N. Shephard (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics 4*(1), 1–30.

Bibinger, M., C. Neely, and L. Winkelmann (2019). Estimation of the discontinuous leverage effect: Evidence from the NASDAQ order book. *Journal of Econometrics 209*(2), 158–184.

Bollerslev, T., T. H. Law, and G. Tauchen (2008). Risk, jumps, and diversification. *Journal of Econometrics 144*(1), 234–256.

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

Bollerslev, T. and V. Todorov (2011). Estimation of jump tails. *Econometrica 79*(6), 1727–1783.

Boudt, K. and M. Petitjean (2014). Intraday liquidity dynamics and news releases around price jumps: Evidence from the DJIA stocks. *Journal of Financial Markets 17*, 121–149.

Chen, R. Y. and P. A. Mykland (2017). Model-free approaches to discern non-stationary microstructure noise and time-varying liquidity in high-frequency data. *Journal of Econometrics 200*(1), 79–103.

Chernov, M., A. R. Gallant, E. Ghysels, and G. Tauchen (2003). Alternative models for stock price dynamics. *Journal of Econometrics 116*(1-2), 225–257.

Cho, D. C. and E. W. Frees (1988). Estimating the volatility of discrete stock prices. *The Journal of Finance 43*(2), 451–466.

Chong, C. H. and V. Todorov (2024). Volatility of volatility and leverage effect from options. *Journal of Econometrics 240*(1), 105669.

Christensen, K., R. Oomen, and M. Podolskij (2014). Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics 114*(3), 576–599.

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

Da, R. and D. Xiu (2021). When moving-average models meet high-frequency data: Uniform inference on volatility. *Econometrica 89*(6), 2787–2825.

Ding, Y., Y. Li, G. Liu, and X. Zheng (2024). Stock co-jump networks. *Journal of Econometrics 239*(2), 105420.

Erdemlioglu, D. and X. Yang (2023). News arrival, time-varying jump intensity, and realized volatility: Conditional testing approach. *Journal of Financial Econometrics 21*(5), 1519–1556.

Evans, K. P. (2011). Intraday jumps and US macroeconomic news announcements. *Journal of Banking & Finance 35*(10), 2511–2527.

Fan, J. and Y. Wang (2007). Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of the American Statistical Association 102*(480), 1349–1362.

Fukasawa, M. and M. Rosenbaum (2012). Central limit theorems for realized volatility under hitting times of an irregular grid. *Stochastic Processes and Their Applications 122*(12), 3901–3920.

Giesecke, K. and G. Schwenkler (2019). Simulated likelihood estimators for discretely observed jump-diffusions. *Journal of Econometrics 213*(2), 297–320.

Gürkaynak, R., B. Kısacıkoğlu, and J. H. Wright (2020). Missing events in event studies. *Econometrica 110*(12), 3871–3912.

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

Hasbrouck, J. (1993). Assessing the quality of a security market: A new approach to transaction-cost measurement. *The Review of Financial Studies 6*, 191–212.

Hayashi, T., J. Jacod, and N. Yoshida (2011). Irregular sampling and central limit theorems for power variations: The continuous case. *Annales de l'IHP Probabilités et statistiques 47*(4), 1197–1218.

Hong, S. Y., I. Nolte, S. Taylor, and X. Zhao (2023). Volatility estimation and forecasts based on price durations. *Journal of Financial Econometrics 21*(1), 106–144.

Huang, X. and G. Tauchen (2005). The relative contribution of jumps to total price variance. *Journal of Financial Econometrics 3*(4), 456–499.

Jacod, J. (2008). Asymptotic properties of realized power variations and related functionals of semimartingales. *Stochastic Processes and Their Applications 118*(4), 517–559.

Jacod, J., Y. Li, and X. Zheng (2017). Statistical properties of microstructure noise. *Econometrica 85*(4), 1133–1174.

Jacod, J. and A. Shiryaev (2013). *Limit Theorems for Stochastic Processes*. Springer-Verlag.

Jacod, J. and V. Todorov (2009). Testing for common arrivals of jumps for discretely observed multidimensional processes. *Annals of Statistics 37*(4), 1792–1838.

Jeon, Y., T. H. McCurdy, and X. Zhao (2022). News as sources of jumps in stock returns: Evidence from 21 million news articles for 9000 companies. *Journal of Financial Economics 145*(2), 1–17.

Kalnina, I. and O. Linton (2008). Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error. *Journal of Econometrics 147*(1), 47–59.

Kalnina, I. and D. Xiu (2017). Nonparametric estimation of the leverage effect: A trade-off between robustness and efficiency. *Journal of the American Statistical Association 112*(517), 384–396.

Kirilenko, A., A. S. Kyle, M. Samadi, and T. Tuzun (2017). The flash crash: High frequency trading in an electronic market. *The Journal of Finance 72*(3), 967–998.

Kolokolov, A. and R. Renò (2024). Jumps or staleness? *Journal of Business & Economic Statistics 42*(2), 516–532.

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

Li, J., V. Todorov, G. Tauchen, and H. Lin (2019). Rank tests at jump events. *Journal of Business & Economic Statistics 37*(2), 312–321.

Li, M. Z. and O. B. Linton (2022). A ReMeDi for microstructure noise. *Econometrica 90*, 367–389.

Li, Y., G. Liu, and Z. Zhang (2022). Volatility of volatility: Estimation and tests based on noisy high frequency data with jumps. *Journal of Econometrics 229*(2), 422–451.

Li, Y., P. A. Mykland, E. Renault, L. Zhang, and X. Zheng (2014). Realized volatility when sampling times are possibly endogenous. *Econometric Theory 30*(3), 580–605.

Li, Y., Z. Zhang, and X. Zheng (2013). Volatility inference in the presence of both endogenous time and microstructure noise. *Stochastic Processes and Their Applications 123*(7), 2696–2727.

Maneesoonthorn, W., G. M. Martin, and C. S. Forbes (2020). High-frequency jump tests: Which test should we use? *Journal of Econometrics 219*(2), 478–487.

Menkveld, A. J. and B. Z. Yueshen (2019). The flash crash: A cautionary tale about highly fragmented markets. *Management Science 65*(10), 4470–4488.

Mykland, P. A. and L. Zhang (2006). ANOVA for diffusions and itô processes. *Annals of Statistics 34*(4), 1931–1963.

Oomen, R. C. (2006). Properties of realized variance under alternative sampling schemes. *Journal of Business & Economic Statistics 24*(2), 219–237.

Park, S., S. Y. Hong, and O. B. Linton (2016). Estimating the quadratic covariation matrix for asynchronously observed high frequency stock returns corrupted by additive measurement error. *Journal of Econometrics 191*(2), 325–347.

Patton, A. J. and K. Sheppard (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *The Review of Economics and Statistics 97*(3), 683–697.

Pelger, M. (2019). Large-dimensional factor modeling based on high-frequency observations. *Journal of Econometrics 208*(1), 23–42.

Pelger, M. (2020). Understanding systematic risk: A high-frequency approach. *Journal of Finance 75*(4), 2179–2220.

Pelletier, D. and W. Wei (2024). A stochastic price duration model for estimating high-frequency volatility. *Journal of Financial Econometrics 22*(5), 1372–1396.

Phillips, P. C. B. and J. Yu (2023). Information loss in volatility measurement with flat price trading. *Empirical Economics 64*, 2957–2999.

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

Reed, A. V., M. Samadi, and J. S. Sokobin (2020). Shorting in broad daylight: Short sales and venue choice. *Journal of Financial and Quantitative Analysis 55*(7), 2246–2269.

Robert, C. Y. and M. Rosenbaum (2012). Volatility and covariation estimation when microstructure noise and trading times are endogenous. *Mathematical Finance 22*(1), 133–164.

Rosenbaum, M. (2009). Integrated volatility and round-off error. *Bernoulli 15*(3), 687–720.

Scaillet, O., A. Treccani, and C. Trevisan (2020). High-frequency jump analysis of the bitcoin market. *Journal of Financial Econometrics 18*(2), 209–232.

Shi, Y., W. Liu, and K. Ho (2016). Public news arrival and the idiosyncratic volatility puzzle. *Journal of Empirical Finance 37*, 159–172.

Wang, C. D. and P. A. Mykland (2014). The estimation of leverage effect with high-frequency data. *Journal of the American Statistical Association 109*(505), 197–215.

Zhang, C., J. Li, and T. Bollerslev (2022). Occupation density estimation for noisy high-frequency data. *Journal of Econometrics 227*(1), 189–211.

Zhang, L. (2011). Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics 160*(1), 33–47.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association 100*(472), 1394–1411.

Table 1: Empirical power of our local and daily jump tests

	DV^c					DV^C		
α	10^{-5}	10^{-4}	10^{-3}	0.01	0.05	10^{-3}	0.01	0.05
			single					
$6s$	.00	.01	.03	.13	.23	.00	.00	.08
$8s$	.08	.20	.35	.55	.71	.00	.01	.32
$10s$	.41	.56	.74	.90	.97	.00	.16	.56
$12s$	.75	.89	.97	1.00	1.00	.11	.41	.80
			gradual reversal					
$6s$	.00	.00	.03	.12	.23	.00	.00	.09
$8s$	.07	.19	.34	.54	.70	.00	.01	.32
$10s$	.40	.55	.73	.90	.97	.00	.16	.56
$12s$	.74	.89	.97	1.00	1.00	.11	.41	.80

Notes: There is one jump per no-event day at a random time point, with jump sizes in multiples of spread s ranging from $4s$ to $14s$. Jumps are *single*s, or reverse gradually in four consecutive steps, each of the size $2s$.

Table 2: Empirical power for competing daily and local jump tests

	MED_{30}			MED_{60}			SBV_{300}		SBV_{60}		$PABV_1$		$PABV_2$	
α	10^{-3}	.01	.05	10^{-3}	.01	.05	10^{-5}	10^{-4}	10^{-5}	10^{-4}	10^{-5}	10^{-4}	10^{-5}	10^{-4}
								single						
$6s$	.01	.03	.22	.00	.01	.06	.01	.03	.14	.25	.01	.02	.00	.00
$8s$	.05	.17	.50	.01	.06	.20	.02	.05	.36	.47	.05	.08	.01	.01
$10s$	.17	.43	.72	.06	.17	.38	.05	.09	.58	.68	.12	.16	.03	.04
$12s$	.37	.64	.86	.15	.34	.57	.11	.17	.76	.83	.24	.30	.06	.09
								gradual reversal						
$6s$	.00	.00	.03	.00	.00	.00	.00	.02	.05	.14	.00	.00	.00	.00
$8s$	.01	.01	.09	.00	.00	.02	.00	.02	.13	.21	.00	.00	.00	.00
$10s$	.03	.05	.15	.00	.00	.01	.01	.03	.25	.33	.00	.00	.00	.00
$12s$	.05	.11	.26	.00	.01	.03	.01	.03	.33	.42	.00	.01	.00	.00

Notes: There is one jump per no-event day at a random time point, with jump sizes in multiples of spread s ranging from $4s$ to $14s$. Jumps are *single*s, or reverse gradually in four consecutive steps, each of the size $2s$.

Table 3: Empirical bootstrap size and power under different threshold δ values

	DV^c					DV^C		
α	10^{-5}	10^{-4}	10^{-3}	0.01	0.05	10^{-3}	0.01	0.05
	Size							
s	.02	.04	.08	.16	.25	.36	.56	.78
$s+1$	.00	.00	.00	.00	.00	.00	.00	.00
$s+2$	.00	.00	.00	.00	.00	.00	.00	.00
$s+3$	.00	.00	.00	.00	.00	.00	.00	.00
	Power ($10s$)							
s	1.00	1.00	1.00	1.00	1.00	.87	.99	1.00
$s+1$	.43	.57	.73	.89	.97	.00	.17	.57
$s+2$	.14	.22	.34	.47	.58	.00	.04	.16
$s+3$	.07	.11	.17	.28	.39	.00	.03	.09

Notes: δ ranges from s to $s+3$ (in ticks). One jump of size $10s$ is added to each no-event day to test for power.

Table 4: Portions of commonly detected days and difference in JV

α^c	10^{-5}			10^{-4}			10^{-3}			0.01		
α^C	.01	.05	.10	.01	.05	.10	.01	.05	.10	.01	.05	.10
common	.865	.989	1.000	.792	.966	.995	.701	.921	.982	.577	.847	.949
JV difference	.078	.072	.071	.080	.072	.070	.081	.072	.068	.080	.071	.067

Notes: *common* is the number of commonly detected jump days divided by the total number of DV^c detected days; JV difference is the JV by DV^C minus JV by DV^c; α^c is the significance level for DV^c and α^C for DV^C.

Table 5: Summary on jump frequency (JF) and jump variation (JV)

JF	JV	JF	JV	JF	JV	JF	JV	JF	JV	JF	JV
DV^C (0.001)		DV^C (0.01)		DV^C (0.05)		DV^C (0.1)		MED (0.001)		MED (0.01)	
.056	.188	.120	.154	.271	.128	.416	.114	.315	.256	.501	.216
$DV^c(10^{-5})$		DV^c (UT)		$DV^c(10^{-4})$		$DV^c(10^{-3})$		DV^c (0.01)		DV^c (0.05)	
.031	.128	.036	.120	.041	.115	.059	.102	.094	.088	.139	.078

JF	JV	JF	JV	Δ	JF	JV	Δ	JF	JV		
$SBV_{300}(10^{-5})$		$SBV_{60}(10^{-5})$		$PABV_1(10^{-5})$			$PABV_2(10^{-5})$			DV^c (0.1)	
.049	.365	.275	.130	7.8	.019	.160	15.6	.007	.215	.170	.074

Notes: JF is the number of jump days over the total number of trading days; JV is calculated by the formulae in equations (29) and (31) on jump days; significance levels are in brackets. Δ for $PABV$'s is the average block lengths in minutes.

Table 6: Summary statistics for news and trading variables under different detection methods

	MED		DV^C					DV^c				
	S_j	S_n	S_j	S_n	dv	do	ds	S_j	S_n	dv	do	ds
AXP	.47	.41	.71	.42	6.57	3.17	2.47	.97	.41	10.84	3.31	3.10
BA	.52	.56	.74	.54	4.13	2.14	1.65	.84	.54	6.53	3.55	2.33
CAT	.49	.41	.64	.42	5.63	1.11	2.94	.52	.42	4.41	0.79	2.34
CVX	.36	.43	.56	.39	0.34	0.28	0.62	.65	.41	0.57	0.56	1.09
DIS	.44	.41	.62	.41	1.22	1.04	0.87	.59	.42	2.69	2.47	1.74
GE	.47	.53	1.07	.51	2.80	0.33	0.82	1.30	.51	14.00	2.36	2.55
GS	.42	.51	.77	.44	0.53	0.19	1.50	.57	.44	0.02	0.60	1.85
HD	.59	.48	.45	.51	0.55	4.48	1.22	.26	.52	0.79	13.11	1.31
IBM	.45	.54	.74	.51	2.75	0.54	1.69	.85	.52	4.71	1.22	2.65
JNJ	.44	.46	.62	.45	2.21	1.17	1.42	.66	.45	4.32	2.45	2.27
JPM	.49	.58	.61	.57	0.92	0.33	0.60	.95	.56	4.15	0.87	2.09
KO	.41	.40	.41	.40	2.62	0.99	1.48	.48	.40	6.95	2.60	1.70
MCD	.47	.40	.55	.41	3.20	1.77	1.62	1.01	.41	5.83	4.41	2.62
MMM	.37	.39	.52	.38	8.24	2.78	4.09	.57	.38	8.99	3.34	5.20
MRK	.38	.42	.51	.41	3.89	1.36	2.61	.72	.41	9.98	1.71	5.04
NKE	.38	.57	.71	.50	0.60	3.08	0.67	.81	.51	1.81	5.35	0.82
PFE	.49	.60	.97	.55	1.82	0.18	0.48	1.17	.55	2.56	0.88	0.69
PG	.39	.40	.40	.40	3.59	1.68	1.58	.54	.40	6.75	2.27	3.68
UNH	.43	.49	.58	.45	0.58	1.57	1.01	.43	.46	0.64	0.95	1.09
V	.32	.67	.64	.56	0.44	1.36	0.72	.97	.54	1.56	6.09	1.39
VZ	.35	.41	.63	.39	0.62	1.41	0.56	.76	.39	1.36	2.16	1.09
WMT	.39	.44	.77	.42	6.50	3.41	1.80	1.90	.42	14.29	4.31	3.49
XOM	.42	.45	.56	.43	1.95	0.77	0.75	.71	.44	4.99	0.90	1.64
CSCO	.27	.42	.98	.30	0.30	6.54	0.22	1.52	.31	0.77	12.64	0.46
INTC	.31	.64	1.12	.42	0.40	0.85	0.18	1.07	.45	0.48	1.09	0.17
MSFT	.39	.52	.72	.36	0.33	1.01	0.25	.74	.48	0.90	3.38	0.63
avg.	.42	.48	.68	.44	2.41	1.68	1.30	.83	.45	4.65	3.21	2.04

Notes: S_j is the average ADS over jump days and S_n its counterpart over no-jump days. Jump days are identified by MED_{30}, DV^C or DV^c with $\alpha = 0.1\%$ for daily tests and $\alpha = 10^{-5}$ for local tests. $dv = \dfrac{v^l - v^d}{v^d}$, where v^l and v^d are local and daily average volumes, with "local" defined as within 60 seconds before, either the occurrence of a local price jump or the largest price event during a jump day; do (for order imbalance) and ds (for spread) are calculated similarly.

Table 7: Regression for daily test statistic Z^d

	S	dv	do	ds	vol	Z^l	S	dv	do	ds	vol	obs.
	no control for local jumps					control for local jumps						
AXP	1.98	0.86	-1.65	5.30	2.21	39.57	0.99	-1.05	-1.21	-0.78	3.31	690
BA	3.98	0.62	1.43	3.58	3.09	44.48	2.69	-2.09	0.70	1.18	4.93	1104
CAT	2.74	-0.45	14.19	3.61	-1.53	28.56	1.79	-2.32	5.80	-0.25	0.65	798
CVX	2.43	3.38	0.17	0.82	-1.10	14.59	2.79	-0.90	0.78	0.49	-1.61	256
DIS	1.98	-2.79	2.96	2.87	0.43	26.71	3.19	-1.85	0.52	-3.02	1.00	555
GE	2.01	0.18	0.19	0.41	3.81	21.15	-1.43	-1.79	0.52	-0.75	3.16	281
GS	0.69	0.09	-0.74	4.73	0.82	24.40	0.52	-0.46	-0.12	-0.17	1.16	600
HD	-0.52	0.98	-0.01	7.30	1.20	12.48	0.09	-1.14	0.81	3.57	2.29	283
IBM	5.24	0.75	-0.49	7.85	3.72	50.18	3.12	-4.97	-0.71	3.66	5.76	1465
JNJ	2.15	7.47	-0.37	3.71	4.30	28.06	0.70	-0.68	-0.71	1.59	5.43	728
JPM	4.78	-3.31	0.58	6.06	12.62	37.79	2.32	-5.41	0.84	0.52	13.62	1320
KO	-0.90	-1.19	2.77	1.38	2.59	22.15	0.25	-2.91	0.01	1.10	3.62	755
MCD	3.72	-1.21	0.27	1.56	1.65	23.94	3.21	-2.40	0.84	-2.03	2.79	751
MMM	0.46	-1.57	0.01	8.28	3.28	47.90	3.79	-0.82	1.11	6.01	4.81	529
MRK	4.31	-0.33	5.28	12.03	5.21	36.25	0.40	-2.92	2.71	2.15	6.20	788
NKE	1.26	1.17	0.55	0.63	0.20	22.52	2.53	1.01	-1.05	-0.20	0.76	210
PFE	2.66	-1.09	-0.73	2.42	4.49	17.31	2.12	-1.89	-0.98	2.74	4.13	138
PG	-0.00	0.41	1.17	4.05	4.87	44.00	0.46	-3.99	0.22	2.14	7.53	716
UNH	0.32	0.66	0.01	2.80	0.90	17.70	0.64	-0.11	-1.16	1.94	1.03	211
V	0.88	0.03	4.69	1.83	2.26	13.05	0.11	-0.39	0.27	0.91	2.95	192
VZ	1.29	-0.07	0.99	2.12	4.30	29.13	1.73	-1.62	0.72	1.44	5.97	231
WMT	4.47	3.10	-0.30	2.32	3.27	24.90	2.85	0.25	-0.23	0.03	5.22	828
XOM	2.83	0.52	-0.51	8.85	3.03	51.89	3.52	-3.78	0.26	0.82	7.94	1385
CSCO	4.26	0.67	-0.96	1.09	3.62	12.26	1.62	0.46	-1.23	0.81	5.53	97
INTC	3.62	1.26	0.17	3.63	5.13	15.17	2.99	0.27	0.46	2.94	4.11	176
MSFT	5.38	1.07	-0.03	7.99	3.44	23.35	4.84	-0.80	0.43	1.97	4.47	504
avg.	2.38	0.43	1.14	4.12	2.99	28.06	1.84	-1.63	0.37	1.11	4.11	600

Notes: $Z_t^d = \beta_0 + \beta_1 S_t + \beta_2 dv_t + \beta_3 do_t + \beta_4 ds_t + (\beta_5 Z_t^l) + \beta_6 Vol_t + \epsilon_t$, where Z_t^d and Z_t^l are the absolute values of daily and local (take the largest absolute Z statistic for no-jump days) test statistics respectively and Vol_t is a control variable for the daily integrated variance. In the left panel are t values for coefficients of a basic regression model for Z^d without Z^l, and in the right panel are results when Z^l is included to control for local jumps. The last column records the numbers of trading days (with news) remaining, after removing no-jump and no-burst days.

Table 8: Regression for local test statistic Z^l

	S	dv	do	ds	vol	Z^d	S	dv	do	ds	vol
	no control for volatility bursts					control for volatility bursts					
AXP	1.72	1.73	-1.17	6.88	0.47	39.57	0.11	1.83	0.37	4.37	-2.49
BA	2.95	2.34	1.26	3.57	0.18	44.48	-0.40	3.08	0.19	1.17	-3.83
CAT	2.07	1.66	13.57	5.32	-2.79	28.56	0.18	2.82	4.40	3.88	-2.42
CVX	0.55	6.03	-0.60	0.67	0.12	14.59	-1.47	4.96	-0.97	0.16	1.17
DIS	-0.16	-2.08	3.47	6.56	-0.30	26.71	-2.50	0.02	1.88	6.63	-0.95
GE	3.70	1.64	-0.16	1.12	2.34	21.15	3.40	2.42	-0.51	1.28	-1.05
GS	0.45	0.59	-0.92	6.85	0.00	24.40	-0.05	0.74	-0.56	4.87	-0.82
HD	-0.98	3.18	-1.10	7.00	-1.08	12.48	-0.84	3.23	-1.36	2.98	-2.23
IBM	4.19	4.76	-0.07	7.04	0.30	50.18	0.04	6.87	0.52	1.27	-4.38
JNJ	2.30	11.04	0.18	3.60	0.74	28.06	1.08	7.86	0.64	1.31	-3.37
JPM	4.39	0.60	-0.00	7.88	4.33	37.79	1.34	4.30	-0.61	5.00	-6.52
KO	-1.73	1.71	4.38	0.83	-0.36	22.15	-1.50	3.16	3.38	-0.05	-2.54
MCD	1.96	0.91	-0.56	4.71	-0.68	23.94	-0.65	2.27	-0.97	4.89	-2.35
MMM	-1.30	-1.35	-0.52	6.20	1.34	47.90	-3.98	0.17	-1.23	-2.77	-3.73
MRK	5.13	1.83	4.54	13.35	1.79	36.25	2.77	3.44	0.58	5.75	-3.77
NKE	-0.11	0.74	1.32	0.87	-0.24	22.52	-2.19	-0.46	1.59	0.64	-0.78
PFE	1.76	-0.05	-0.23	1.08	2.58	17.31	-0.80	1.54	0.69	-1.67	-1.97
PG	-0.28	2.90	1.24	3.43	1.13	44.00	-0.53	4.93	0.45	-0.06	-5.77
UNH	-0.11	0.93	0.95	2.02	0.32	17.70	-0.56	0.67	1.50	-0.25	-0.60
V	1.15	0.45	6.47	1.68	0.19	13.05	0.75	0.60	4.21	0.57	-1.88
VZ	0.56	0.76	0.74	1.65	1.73	29.13	-1.28	1.79	-0.30	-0.53	-4.38
WMT	3.51	4.43	-0.20	3.50	-1.03	24.90	0.75	3.16	-0.00	2.61	-4.18
XOM	0.95	3.36	-0.81	10.27	-1.96	51.89	-2.30	5.04	-0.68	5.13	-7.58
CSCO	4.01	0.49	-0.26	0.75	0.27	12.26	0.96	-0.06	0.81	-0.18	-3.94
INTC	2.16	1.42	-0.17	2.21	3.12	15.17	-0.87	0.71	-0.46	-0.80	-1.10
MSFT	2.78	2.25	-0.45	9.01	0.48	23.35	-1.56	2.13	-0.62	4.41	-2.87
avg.	1.60	2.01	1.19	4.54	0.50	28.06	-0.39	2.58	0.50	1.95	-2.86

Notes: $Z_t^l = \beta_0 + \beta_1 S_t + \beta_2 dv_t + \beta_3 do_t + \beta_4 ds_t + (\beta_5 Z_t^d) + \beta_6 Vol_t + \epsilon_t$, where Z_t^l and Z_t^d are the absolute values of local (take the largest absolute Z statistic for no-jump days) and daily test statistics respectively and Vol_t is a control variable for the daily integrated variance. In the left panel are t values for coefficients of a basic regression model for Z^l without Z^d, and in the right panel are results when Z^d is included to control for volatility bursts.

Table 9: Simulation evidence: size and power under different δ values

	DV^c (SV2FJ)					DV^C (SV1FJ)				
α	10^{-5}	10^{-4}	10^{-3}	0.01	0.05	10^{-5}	10^{-4}	10^{-3}	0.01	0.05
					size					
s	.017	.037	.074	.160	.277	.508	.648	.879	.999	1.000
$s+1$	.000	.001	.002	.006	.016	.000	.000	.000	.037	.290
$s+2$	.000	.000	.001	.002	.003	.000	.000	.000	.000	.004
$s+3$	.000	.000	.000	.000	.001	.000	.000	.000	.000	.000
$s+4$	.000	.000	.000	.000	.001	.000	.000	.000	.000	.000
$s+5$	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
					extra large jump					
s	.996	.998	.998	.999	1.000	.999	.999	.999	1.000	1.000
$s+1$	.950	.970	.986	.994	.997	.562	.735	.895	.966	.991
$s+2$	.785	.848	.904	.954	.977	.070	.190	.425	.765	.941
$s+3$	.551	.647	.760	.851	.912	.006	.030	.131	.402	.777
$s+4$	.373	.464	.583	.722	.823	.001	.005	.031	.184	.507
$s+5$	.240	.319	.436	.578	.706	.000	.000	.008	.072	.307
					large jump					
s	.994	.996	.998	.999	.999	.994	.999	.999	1.000	1.000
$s+1$	.926	.952	.970	.985	.993	.173	.372	.617	.883	.964
$s+2$	.714	.788	.857	.912	.947	.001	.016	.093	.369	.757
$s+3$	.442	.547	.659	.776	.853	.000	.000	.008	.102	.390
$s+4$	.250	.342	.452	.605	.725	.000	.000	.001	.023	.173
$s+5$	.144	.202	.298	.446	.579	.000	.000	.000	.006	.070
					medium jump					
s	.944	.960	.978	.986	.994	.820	.959	.998	1.000	1.000
$s+1$	.640	.730	.814	.882	.929	.001	.019	.148	.488	.826
$s+2$	.257	.364	.494	.641	.745	.000	.000	.001	.030	.241
$s+3$	.107	.155	.244	.385	.516	.000	.000	.000	.002	.039
$s+4$	.049	.075	.126	.215	.332	.000	.000	.000	.000	.006
$s+5$	.022	.037	.066	.127	.207	.000	.000	.000	.000	.001
					small jump					
s	.653	.740	.822	.888	.939	.643	.814	.981	1.000	1.000
$s+1$	.173	.250	.369	.525	.642	.000	.000	.014	.195	.580
$s+2$	.040	.063	.115	.206	.311	.000	.000	.000	.001	.050
$s+3$	.012	.020	.042	.084	.142	.000	.000	.000	.000	.003
$s+4$	.004	.009	.018	.040	.072	.000	.000	.000	.000	.000
$s+5$	.002	.004	.007	.019	.036	.000	.000	.000	.000	.000

Notes: Threshold values range from s (average bid/ask spread in ticks) to $s+5$ ticks. "Extra large" means jump variance accounts for 30% of daily integrated variance on average; "large" means for 20%; "medium" for 10% and "small" for 5%. For details of the simulation setup refer to Section 5.

Jumps versus bursts: distinguishing sources of extreme risk in financial markets. Internet Appendix (June 4, 2026)

Internet Appendix A: Supplementary results

I Data cleaning criteria

The raw data is cleaned using the methods of Barndorff-Nielsen, Hansen, Lunde and Shephard (2009) and Hong, Nolte, Taylor and Zhao (2023). Data entries, trades and quotes, that meet one or more of the following conditions are deleted: 1) entries outside of the normal 9:30am to 4pm daily trading session; 2) entries with either bid, ask or transaction price equal to zero; 3) transaction prices that are above the ask price plus the bid/ask spread or below the bid price minus the bid/ask spread; 4) entries with negative bid/ask spread; 5) entries with spread larger than 50 times the median spread of the day; 6) entries for which the mid-quote deviates by more than 10 mean-absolute-deviations from a rolling centered median (excluding the observation under consideration) of 50 observations (25 observations before and 25 after). When multiple transaction, bid or ask prices have the same time stamp, the median price is used. We match trades with corresponding bid and ask quotes using a refined Lee and Ready algorithm as outlined in Nolte (2008), which yields the bid/ask spreads.

II Simulated models and main results

The one-factor model (SV1F) is:

$$dX_t = \sigma_t dW_t, \tag{41}$$

$$\sigma_t = \exp(\beta_0 + \beta_1 \tau_t), \tag{42}$$

$$d\tau_t = \alpha \tau_t dt + dB_t. \tag{43}$$

The SV parameters are the same as in Hong, Nolte, Taylor and Zhao (2023). $\beta_0 = -4.311$, $\beta_1 = 0.05934$, $\alpha = -0.011$, and $corr(dW_t, dB_t) = -0.3$. The initial value of τ_t is drawn from its unconditional distribution, which is $N(0, -0.5/\alpha)$.

And the two-factor model (SV2F) is:

$$dX_t = \sigma_t dW_t, \tag{44}$$

$$\sigma_t = \text{s-exp}(\beta_0 + \beta_1 \tau_{1t} + \beta_2 \tau_{2t}), \tag{45}$$

$$d\tau_{1t} = \alpha_1 \tau_{1t} dt + dB_{1t}, \tag{46}$$

$$d\tau_{2t} = \alpha_2 \tau_{2t} dt + (1 + \phi \tau_{2t}) dB_{2t}, \tag{47}$$

The spline-exponential function is the usual exponential function with a knot point of 150% annualized volatility. In addition, $\beta_0 = -4.442$, $\beta_1 = 0.04$, $\beta_2 = 0.635$, $\alpha_1 = -0.005501$, $\alpha_2 = -1.3863$, and $\phi = 0.25$. The correlations among the Wiener processes are $corr(dW_t, dB_{1t}) = corr(dW_t, dB_{2t}) = -0.3$ and $corr(dB_{1t}, dB_{2t}) = 0$. The persistent first factor is initialized each day by drawing from its unconditional distribution while the strongly mean-reverting second factor is simply started at zero. The initial price $P_0 = 50$.

Finally we add MMS noise to the true latent prices the same way as in Hong, Nolte, Taylor and Zhao (2023). We assume the bid/ask spread, s, to be positively related to the instantaneous volatility, σ_t, and $s = 8\sigma_t$, where s is in ticks and σ_t is an annualized number so that the average spread is 2 ticks given an annual volatility of 25%, which is consistent with empirical data. Trades sit on the bid or the ask side of the book randomly with equal probability. Bid and ask prices are generated by subtracting from or adding to the latent price half of the spread. Price discreteness is introduced by rounding the resulting prices to the nearest cent.

III Control type-I error: the local jump test

The local test aims to single out the impact of price jumps only, in contrast, the daily statistics contain influence from volatility bursts. Accordingly, we focus solely on the misclassification rate of local jump tests in this section, including SBV- and $PABV$-type estimators for comparison. If a local jump test is accurate, the detected jump must be the largest price change of the day. Table 10 summarizes our findings, to save space presented as an average over all 30 stocks. Hit calculates the proportion of days when the identified price jump is exactly the largest price change, and $size$ gives the average size of detected jumps as multiples of the threshold δ. Due to the interval-based nature of $PABV$ and SBV tests, a "hit" will mean the detected jump interval contains the largest price change.

As expected, our local test stands out with a hit rate at 100% for large jumps ($\alpha = 10^{-5}$ or 10^{-4}) of average sizes 8.7 to 9.5 times the threshold (δ), and at 99% and 98% for smaller jumps of size 6.2 to 7.9 times δ. Given the mean value (over all 30 stocks) of the average daily bid/ask spread is 2.9 ticks, the large jumps are of size 0.3 to 0.4 dollars on average (given $\delta = s + 0.01$). Our local test is essentially standardizing every price event by the noise-included variance estimate (after properly scaled to account for multiple testing). As long as the price jump is large enough, it will trigger a price event immediately, regardless of cumulative returns from the past price path (within the event). Thus, our local test is roughly equivalent to selecting the largest absolute price change. But exceptions do exist. A miss will arise when the largest absolute price change is not big enough and when partially offset by the afore-cumulated price changes of an opposite sign could let a smaller price event emerge as the winner. This is also more likely to happen when the average bid/ask spread, thus the threshold value is high, making it both more difficult to trigger an event

Table 10: Hit rates and size (in multiples of δ) of detected jumps

hit	size (δ)	hit	size (δ)	hit	size (δ)	hit	size (δ)	hit	size (δ)
DV^c (10^{-5})		DV^c (10^{-4})		DV^c (10^{-3})		DV^c (0.01)		DV^c (0.05)	
1.00	9.52	1.00	8.73	0.99	7.87	0.99	6.90	0.98	6.18
SBV_{300} (10^{-5})		SBV_{60} (10^{-5})		SBV_{30} (10^{-5})		$PABV_1$ (10^{-5})		$PABV_2$ (10^{-5})	
0.40	3.58	0.32	2.98	0.28	2.60	0.53	5.96	0.49	6.42

Notes: for DV^c, *hit* calculates the portion of days when the detected jump is exactly the maximum price change of the day; and *size* is the average size of detected jumps, recorded as multiples of the threshold value (δ). For other local tests, *hit* records the portions of days when the detected "interval" includes the maximum price change of the day; and *size* records the average size (in multiples of δ) of the biggest price change over the detected "interval". α values are in brackets. Mean and median values of average daily bid/ask spread over all 30 stocks are 2.9 and 2.0 ticks.

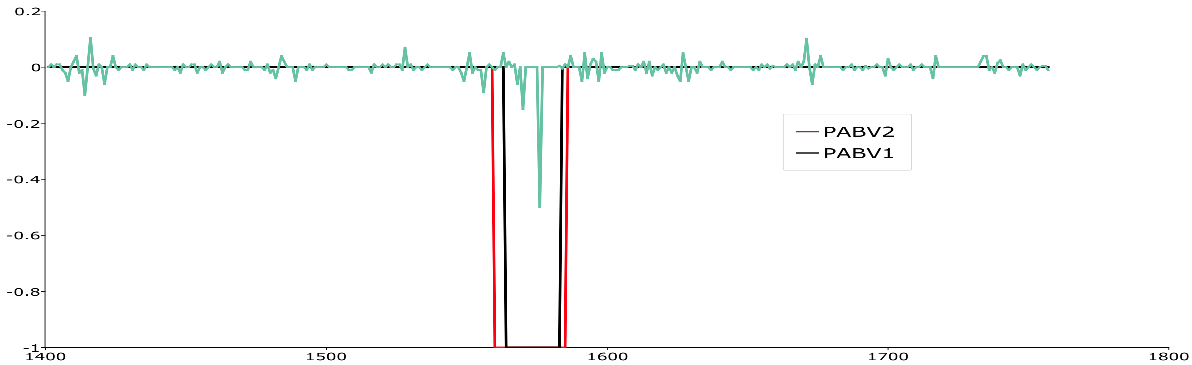
and less timely to stop continuous price changes from cumulating. From table 10, we see this slight rise in the misclassification rate, which can be roughly construed as the distance between the hit rate and 1, from 0 to 2% when the jump size identified by our local test gets smaller.

For interval-based tests, hit rates will be lower, accompanied by smaller average jump sizes. SBV and $PABV$ tests display different features. For the former class of local tests, false positives arise mainly from drift and volatility bursts, as well as increased staleness, as evident from Section 6.2; while the latter often see inflated continuous price changes when large jumps are averaged over the entire neighborhood, thus leading to misclassifications.

However, the pre-averaging procedure also guarantees that, once a pre-averaged return is classified as a jump, it must be truly big enough, which is why the identified jump size is considerably larger by $PABV$ than by SBV, and larger by $PABV_2$ than by $PABV_1$. The higher the θ parameter, the larger the block size. A pre-averaged return is calculated as a weighted average of all returns within the block, with the middle return bearing the highest weight under the weighting function used by COP (and us). Due to overlapping blocks, a large price discontinuity will inflate all pre-averaged returns around its neighborhood, resulting in false positives, as evident in Figure 11, where all observations within the highlighted intervals are classified as jumps. We use only the largest price change over the "halo" interval to calculate the "hit" and "size" statistics in Table 10. We can see this halo effect is stronger by $PABV_2$ due to its wider block size.

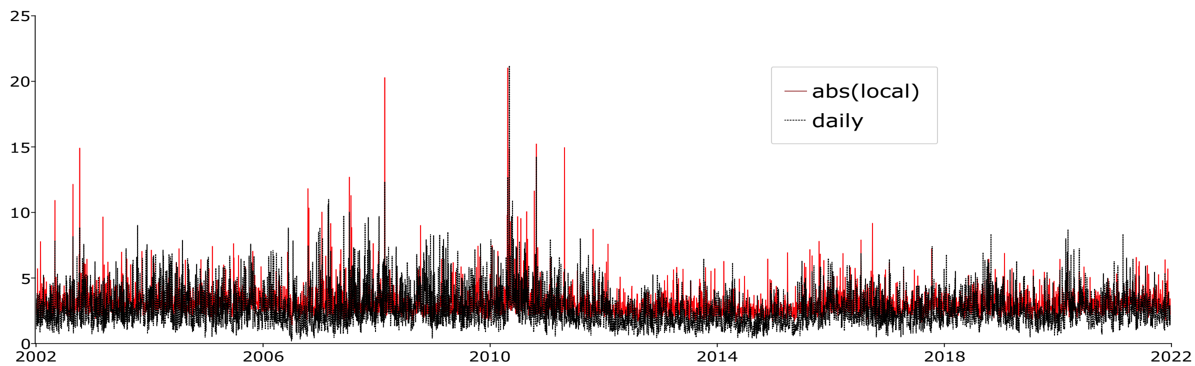
Meanwhile, the impact of a price jump is undermined by averaging over all returns within the block, leading to possible false negatives, which is why $PABV_2$ has both a higher misclassification rate and a higher average jump size: some price jumps go undetected with heavier averaging

Figure 11: Compare $PABV_1$ and $PABV_2$, the first detected jump from AA



Notes: Preaveraged returns within the highlighted regions are all classified as jumps. Compared to $PABV_1$, $PABV_2$ generates a stronger "halo effect".

Figure 12: Regression variables: (absolute) local statistics and daily statistics, IBM



Notes: the standardized (absolute) local and daily statistics are comparable in scale and exhibit substantial agreement on jump days.

but those that can survive tend to be larger, especially when many price jumps reverse through subsequent price changes of an opposite sign.

IV Additional simulation evidence with volatility jumps

V Additional empirical results

V.1 Clustering

V.2 Signed news and jumps

VI Empirical results: individual stocks

VI.1 Control type-I error: hit & size among local tests

VI.2 Empirical power

VI.3 A proxy for volatility bursts

VI.4 Jump size and frequency

Table 11: Daily jump tests: size and power under two volatility schemes and varied jump sizes

	SV2FJ					SV1FJ				
α	10^{-5}	10^{-4}	10^{-3}	0.01	0.05	10^{-5}	10^{-4}	10^{-3}	0.01	0.05
					size					
DV^C	.002	.007	.039	.213	.476	.000	.000	.000	.037	.290
MIN_{60}	.000	.000	.000	.001	.015	.000	.000	.000	.000	.000
MED_{60}	.000	.000	.001	.013	.057	.000	.000	.000	.000	.001
MIN_{30}	.000	.000	.000	.004	.028	.000	.000	.000	.000	.003
MED_{30}	.000	.000	.002	.017	.082	.000	.000	.000	.000	.011
MIN_5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MED_5	.999	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
					extra large jump					
DV^C	.509	.644	.784	.898	.953	.562	.735	.895	.966	.991
MIN_{60}	.067	.155	.308	.551	.771	.007	.065	.330	.793	.964
MED_{60}	.319	.445	.616	.787	.892	.345	.630	.887	.983	.996
MIN_{30}	.244	.371	.541	.725	.850	.265	.538	.810	.948	.979
MED_{30}	.578	.677	.788	.885	.936	.881	.954	.981	.991	.995
MIN_5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MED_5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
					large jump					
DV^C	.261	.401	.593	.794	.907	.173	.372	.617	.883	.964
MIN_{60}	.013	.037	.099	.286	.526	.000	.000	.028	.266	.693
MED_{60}	.102	.183	.327	.540	.728	.021	.093	.337	.742	.946
MIN_{30}	.065	.129	.254	.464	.675	.007	.066	.278	.674	.895
MED_{30}	.270	.389	.538	.717	.837	.290	.539	.803	.949	.981
MIN_5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MED_5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
					medium jump					
DV^C	.048	.118	.282	.551	.777	.001	.019	.148	.488	.826
MIN_{60}	.000	.001	.007	.052	.195	.000	.000	.000	.007	.113
MED_{60}	.007	.022	.062	.189	.374	.000	.000	.006	.084	.357
MIN_{30}	.003	.011	.036	.130	.304	.000	.000	.003	.084	.350
MED_{30}	.029	.066	.157	.328	.527	.001	.013	.082	.356	.702
MIN_5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MED_5	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
					small jump					
DV^C	.011	.033	.124	.366	.633	.000	.000	.014	.195	.580
MIN_{60}	.000	.000	.001	.010	.064	.000	.000	.000	.000	.011
MED_{60}	.000	.002	.011	.055	.172	.000	.000	.000	.004	.048
MIN_{30}	.000	.001	.004	.032	.119	.000	.000	.000	.003	.069
MED_{30}	.003	.010	.032	.109	.262	.000	.000	.002	.037	.206
MIN_5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MED_5	.999	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Notes: “extra large” means jump variance accounts for 30% of daily integrated variance on average; “large” means for 20%; “medium” for 10% and “small” for 5%.

Table 12: Local jump tests: size and power under two volatility schemes and varied jump sizes

	SV2FJ					SV1FJ				
α	10^{-5}	10^{-4}	10^{-3}	0.01	0.05	10^{-5}	10^{-4}	10^{-3}	0.01	0.05
					size					
DV^c	.000	.001	.002	.006	.016	.000	.000	.000	.001	.003
$PABV_1$	.001	.007	.016	.037	.072	.000	.000	.000	.003	.012
$PABV_2$	.001	.001	.006	.014	.030	.000	.000	.000	.000	.004
SBV_{300}	.003	.009	.018	.066	.161	.000	.000	.001	.011	.053
SBV_{60}	.007	.015	.045	.128	.286	.000	.000	.002	.019	.070
SBV_{30}	.016	.035	.088	.226	.422	.000	.001	.005	.037	.130
					extra large jump					
DV^c	.976	.986	.992	.996	.998	.998	.998	.999	.999	.999
$PABV_1$	.702	.787	.862	.923	.953	.857	.919	.962	.981	.988
$PABV_2$	.170	.243	.349	.490	.606	.140	.234	.381	.562	.693
SBV_{300}	.233	.338	.494	.682	.820	.219	.375	.569	.761	.876
SBV_{60}	.983	.991	.996	.999	1.000	.999	.999	.999	.999	.999
SBV_{30}	.998	.999	.999	1.000	1.000	.999	.999	1.000	1.000	1.000
					large jump					
DV^c	.926	.952	.970	.985	.993	.992	.995	.997	.998	.999
$PABV_1$	.402	.510	.630	.749	.826	.452	.610	.753	.868	.922
$PABV_2$	.058	.090	.154	.249	.348	.026	.062	.131	.246	.367
SBV_{300}	.090	.154	.263	.439	.617	.056	.128	.262	.469	.653
SBV_{60}	.908	.947	.974	.990	.996	.995	.999	.999	.999	.999
SBV_{30}	.991	.994	.998	.999	1.000	.999	.999	1.000	1.000	1.000
					medium jump					
DV^c	.640	.730	.814	.882	.929	.826	.904	.956	.983	.991
$PABV_1$	.083	.127	.197	.310	.420	.042	.088	.173	.306	.441
$PABV_2$	.009	.014	.032	.064	.105	.001	.004	.012	.033	.071
SBV_{300}	.017	.034	.074	.172	.325	.005	.015	.045	.128	.268
SBV_{60}	.491	.610	.737	.856	.932	.594	.744	.860	.937	.973
SBV_{30}	.857	.908	.950	.982	.994	.974	.989	.995	.999	1.000
					small jump					
DV^c	.173	.250	.369	.525	.642	.157	.288	.476	.683	.811
$PABV_1$	.012	.024	.050	.097	.162	.002	.008	.017	.047	.097
$PABV_2$	.002	.004	.011	.023	.047	.000	.001	.001	.005	.014
SBV_{300}	.007	.015	.032	.088	.205	.001	.002	.008	.037	.108
SBV_{60}	.123	.193	.305	.487	.663	.079	.156	.295	.477	.647
SBV_{30}	.400	.519	.656	.804	.895	.457	.604	.755	.877	.937

Notes: "extra large" means jump variance accounts for 30% of daily integrated variance on average; "large" means for 20%; "medium" for 10% and "small" for 5%.

Table 13: 12 seconds inter-trade duration: the effect of staleness

daily tests										
α	10^{-5}	10^{-4}	10^{-3}	0.01	0.05	10^{-5}	10^{-4}	10^{-3}	0.01	0.05
	SV2FJ					SV1FJ				
	size									
DV^C	.006	.027	.119	.376	.653	.000	.000	.011	.217	.617
MIN_{60}	.000	.000	.000	.006	.052	.000	.000	.000	.000	.015
MED_{60}	.000	.000	.003	.026	.111	.000	.000	.000	.001	.025
MIN_{30}	.006	.029	.125	.396	.657	.004	.031	.158	.505	.818
MED_{30}	.010	.037	.119	.317	.548	.003	.021	.104	.362	.684
	extra large jump									
DV^C	.591	.722	.842	.927	.968	.661	.820	.934	.978	.993
MIN_{60}	.106	.214	.407	.666	.843	.045	.199	.539	.892	.981
MED_{60}	.383	.531	.685	.841	.931	.501	.753	.928	.988	.996
MIN_{30}	.676	.802	.901	.960	.984	.922	.968	.986	.994	.998
MED_{30}	.821	.888	.938	.973	.986	.979	.988	.994	.997	.998
	large jump									
DV^C	.351	.511	.701	.864	.941	.313	.518	.754	.937	.981
MIN_{60}	.021	.062	.162	.400	.659	.001	.013	.121	.496	.844
MED_{60}	.135	.243	.409	.637	.804	.067	.217	.509	.841	.965
MIN_{30}	.426	.612	.782	.914	.964	.646	.850	.950	.983	.992
MED_{30}	.608	.733	.846	.928	.967	.853	.939	.976	.991	.995

local tests										
α	10^{-5}	10^{-4}	10^{-3}	0.01	0.05	10^{-5}	10^{-4}	10^{-3}	0.01	0.05
	SV2FJ					SV1FJ				
	size									
DV^c	.003	.005	.012	.037	.082	.000	.001	.003	.008	.024
$PABV_1$	.003	.007	.015	.037	.078	.000	.000	.001	.004	.017
$PABV_2$	.001	.002	.004	.011	.029	.000	.000	.000	.001	.005
SBV_{300}	.002	.007	.020	.063	.163	.000	.000	.002	.011	.056
SBV_{60}	.012	.028	.070	.191	.371	.000	.001	.007	.041	.134
	extra large jump									
DV^c	.969	.983	.990	.995	.997	.997	.998	.998	.998	.999
$PABV_1$	.439	.554	.684	.797	.868	.518	.665	.797	.893	.944
$PABV_2$	.067	.110	.174	.281	.397	.035	.075	.155	.283	.420
SBV_{300}	.230	.340	.493	.677	.821	.223	.376	.568	.761	.873
SBV_{60}	.984	.991	.995	.998	.999	.999	.999	.999	.999	.999
	large jump									
DV^c	.897	.932	.960	.982	.991	.986	.993	.996	.998	.998
$PABV_1$	.200	.283	.395	.540	.668	.169	.281	.446	.626	.743
$PABV_2$	.021	.039	.072	.129	.204	.007	.018	.043	.104	.189
SBV_{300}	.092	.154	.266	.441	.616	.059	.128	.267	.469	.647
SBV_{60}	.913	.950	.977	.992	.997	.993	.996	.999	.999	.999

Notes: size and power statistics for all daily and local tests when transactions occur every 12 seconds on average, instead of 6 seconds as in tables 11 and 12. “Extra large” means jump variance accounts for 30% of daily integrated variance on average and “large” means it accounts for 20%.

Table 14: Daily jump tests: size and power with volatility bursts, SV1FJJ

α	10^{-5}	10^{-4}	10^{-3}	0.01	0.05
			size		
DV^C	.000	.000	.019	.202	.568
MIN_{60}	.000	.000	.000	.000	.001
MED_{60}	.000	.000	.000	.000	.002
MIN_{30}	.000	.000	.000	.000	.003
MED_{30}	.000	.000	.000	.001	.011
MIN_5	1.000	1.000	1.000	1.000	1.000
MED_5	1.000	1.000	1.000	1.000	1.000
			extra large jump		
DV^C	.628	.779	.909	.970	.992
MIN_{60}	.026	.054	.207	.643	.918
MED_{60}	.191	.429	.749	.951	.993
MIN_{30}	.132	.339	.651	.908	.970
MED_{30}	.715	.882	.958	.988	.995
MIN_5	1.000	1.000	1.000	1.000	1.000
MED_5	1.000	1.000	1.000	1.000	1.000
			large jump		
DV^C	.297	.490	.704	.908	.972
MIN_{60}	.016	.021	.038	.193	.581
MED_{60}	.030	.066	.223	.613	.896
MIN_{30}	.026	.046	.177	.544	.841
MED_{30}	.155	.349	.651	.901	.970
MIN_5	1.000	1.000	1.000	1.000	1.000
MED_5	1.000	1.000	1.000	1.000	1.000
			medium jump		
DV^C	.031	.107	.321	.643	.890
MIN_{60}	.002	.004	.009	.023	.111
MED_{60}	.009	.013	.020	.075	.303
MIN_{30}	.008	.013	.021	.070	.295
MED_{30}	.020	.026	.064	.270	.613
MIN_5	1.000	1.000	1.000	1.000	1.000
MED_5	1.000	1.000	1.000	1.000	1.000
			small jump		
DV^C	.006	.022	.111	.430	.736
MIN_{60}	.000	.000	.001	.004	.021
MED_{60}	.000	.001	.003	.012	.057
MIN_{30}	.000	.001	.004	.012	.070
MED_{30}	.003	.006	.011	.042	.184
MIN_5	1.000	1.000	1.000	1.000	1.000
MED_5	1.000	1.000	1.000	1.000	1.000

Notes: "extra large" means jump variance accounts for 30% of daily integrated variance on average; "large" means for 20%; "medium" for 10% and "small" for 5%. Volatility burst is added by randomly selecting a 10-minute interval and doubling the returns within, resulting in a 7.7% jump in integrated variance.

Table 15: Local jump tests: size and power with volatility bursts, SV1FJJ

α	10^{-5}	10^{-4}	10^{-3}	0.01	0.05
			size		
DV^c	.002	.005	.013	.029	.067
$PABV_1$	.012	.021	.040	.076	.123
$PABV_2$	.004	.008	.017	.030	.052
SBV_{300}	.009	.017	.033	.082	.170
SBV_{60}	.040	.075	.140	.260	.413
SBV_5	.075	.134	.243	.414	.590
			extra large jump		
DV^c	.997	.999	.999	.999	.999
$PABV_1$	.817	.898	.952	.980	.989
$PABV_2$	.137	.221	.358	.538	.670
SBV_{300}	.205	.346	.544	.749	.875
SBV_{60}	.999	.999	.999	.999	.999
SBV_5	.999	.999	1.000	1.000	1.000
			large jump		
DV^c	.990	.994	.997	.999	.999
$PABV_1$	.413	.557	.723	.850	.915
$PABV_2$	.041	.073	.139	.248	.371
SBV_{300}	.069	.135	.266	.476	.673
SBV_{60}	.990	.998	.999	.999	.999
SBV_5	.999	.999	1.000	1.000	1.000
			medium jump		
DV^c	.819	.891	.948	.977	.987
$PABV_1$	.057	.101	.185	.329	.471
$PABV_2$	.010	.018	.034	.067	.115
SBV_{300}	.020	.037	.077	.184	.341
SBV_{60}	.549	.714	.848	.942	.978
SBV_5	.968	.988	.996	.999	1.000
			small jump		
DV^c	.126	.249	.443	.664	.810
$PABV_1$	.022	.037	.062	.117	.192
$PABV_2$	.005	.011	.020	.037	.064
SBV_{300}	.011	.022	.042	.104	.212
SBV_{60}	.110	.202	.353	.576	.744
SBV_5	.442	.609	.768	.899	.958

Notes: "extra large" means jump variance accounts for 30% of daily integrated variance on average; "large" means for 20%; "medium" for 10% and "small" for 5%. Volatility burst is added by randomly selecting a 10-minute interval and doubling the returns within, resulting in a 7.7% jump in integrated variance.

Table 16: Jump and volatility clustering t statistics by runs test

α	10^{-5}	10^{-3}	0.05	10^{-5}	10^{-3}	0.05
	Jump clustering			Volatility clustering		
AA	-0.10	-1.08	-3.71	0.27	-5.65	-12.39
AXP	-1.64	-2.52	-1.29	0.72	-6.77	-18.28
BA	-0.61	-0.26	-3.41	-3.41	-10.64	-22.29
CVX	0.57	1.03	0.82	-7.97	-16.19	-17.41
CAT	-1.93	-2.47	-0.87	-0.77	-4.92	-20.36
DD	1.13	-1.21	-0.77	0.20	-2.73	-9.35
DIS	-1.85	-2.21	-2.70	-3.59	-14.46	-36.46
GE	-3.46	-1.25	-9.71	-6.57	-20.04	-41.08
GS	0.32	1.15	-0.30	0.07	-3.73	-10.36
HD	0.32	-1.08	-0.86	0.30	-3.15	-26.90
IBM	-1.05	-1.97	-0.86	0.51	-5.97	-20.17
JNJ	-0.91	0.31	-0.31	-8.43	-14.78	-30.08
JPM	-2.29	-4.98	-3.56	-19.14	-30.61	-42.37
KO	-4.82	-2.24	-3.45	-2.40	-15.46	-17.53
MCD	-2.12	-0.84	-1.91	-10.22	-9.27	-19.58
MMM	-1.43	-1.25	-2.25	-7.76	-6.43	-21.25
MRK	-0.58	-3.62	-4.78	-6.87	-16.17	-30.49
NKE	-0.15	0.78	-0.24	-1.31	-3.60	-19.51
PFE	-3.74	-4.74	-5.75	-17.83	-20.53	-20.84
PG	-3.75	-4.60	-2.00	-3.32	-10.68	-26.42
TRV	-2.93	-1.13	0.14	0.05	0.32	-4.42
UNH	-0.76	-1.88	-0.72	0.32	-6.39	-11.98
V	-1.06	-1.70	-1.37	-3.15	-11.36	-22.85
VZ	0.62	-0.73	-2.74	-1.05	-5.92	-14.72
UTX	-2.03	-2.50	-3.80	0.52	-4.73	-19.55
WMT	-1.20	-0.69	0.30	-4.14	-8.27	-23.03
XOM	-2.67	-2.22	0.91	-19.63	-21.48	-24.80
CSCO	0.21	-1.62	-4.34	-18.64	-17.95	-25.76
INTC	-3.47	-6.86	-7.11	-19.84	-24.30	-29.25
MSFT	-6.24	-8.20	-4.49	-36.77	-39.33	-38.39
$<$ -2.33	9	9	13	18	29	30

Notes: the time period under examination is 20 years from 2012 to 2021; α values for the left-hand-side panel means the significance levels of the local jump tests and those on the right are significance levels used to remove jumps before applying the daily jump test whose significance level is 5%. The last row summarizes the number of stocks exhibiting significant clustering under 1% significance level.

Table 17: Regression for daily test statistic Z^d, different news impact

	$S+$	$S-$	dv	do	ds	vol	dS	dv	do	ds	vol
	positive and negative news						news difference				
AXP	1.05	2.49	0.96	-1.70	5.36	2.15	-0.29	0.82	-1.75	5.59	2.19
BA	2.63	4.42	0.71	1.30	3.62	2.84	-0.82	0.78	1.34	3.46	2.91
CAT	1.92	2.58	-0.44	14.20	3.60	-1.61	-0.24	-0.50	14.11	3.88	-1.37
CVX	2.24	1.37	3.37	0.17	0.82	-1.03	1.08	3.52	0.36	1.11	-0.17
DIS	1.11	2.83	-2.79	3.05	2.87	0.36	-0.78	-2.90	3.14	2.89	0.49
GE	2.61	0.14	0.32	0.27	0.27	4.09	1.96	0.33	0.28	0.31	4.77
GS	-0.02	0.94	0.09	-0.75	4.71	0.81	-0.80	0.08	-0.75	4.71	0.88
HD	-0.39	-0.60	1.00	-0.04	7.27	1.21	0.13	0.97	0.01	7.34	1.15
IBM	5.16	2.29	0.73	-0.50	7.86	3.70	3.14	0.70	-0.58	7.91	3.92
JNJ	2.05	1.18	7.47	-0.36	3.71	4.30	1.03	7.42	-0.45	3.76	4.24
JPM	4.18	3.07	-3.32	0.57	6.04	12.68	0.42	-3.28	0.42	5.75	13.12
KO	-0.81	-0.57	-1.19	2.76	1.38	2.59	-0.27	-1.15	2.77	1.34	2.58
MCD	3.51	1.79	-1.21	0.27	1.56	1.62	2.20	-1.33	0.31	1.81	2.07
MMM	0.31	0.47	-1.53	0.03	8.16	3.25	-0.11	-1.59	-0.03	8.41	3.30
MRK	3.84	3.09	-0.33	5.27	12.01	5.20	0.98	-0.48	5.20	11.93	5.16
NKE	0.61	2.69	1.19	0.57	0.65	-0.36	-0.88	1.23	0.52	0.63	0.57
PFE	3.22	-0.03	-1.11	-0.71	2.43	4.48	2.90	-1.06	-0.78	2.44	4.42
PG	0.04	-0.05	0.40	1.17	4.05	4.87	0.07	0.40	1.17	4.05	4.87
UNH	-0.58	2.00	0.72	-0.05	2.75	0.72	-1.63	0.71	-0.04	2.85	1.17
V	1.04	0.02	-0.09	4.59	1.83	2.33	0.86	-0.23	4.97	1.84	2.59
VZ	1.02	1.45	-0.28	1.11	2.14	4.28	-0.02	0.24	0.84	2.13	4.38
WMT	4.15	2.50	3.08	-0.30	2.34	3.28	1.78	3.47	-0.37	2.04	3.33
XOM	1.59	3.03	0.53	-0.52	8.76	2.97	-1.07	0.49	-0.55	8.85	2.92
CSCO	4.83	1.36	0.70	-0.97	1.18	4.20	3.11	0.66	-1.07	1.38	5.77
INTC	2.10	3.06	1.45	0.04	3.59	5.14	0.29	1.42	0.02	3.67	5.27
MSFT	5.54	-0.43	1.01	0.06	7.86	3.72	5.60	0.98	0.08	7.86	3.96
avg.	2.04	1.58	0.44	1.13	4.11	2.99	0.72	0.45	1.12	4.15	3.25

Notes: the left-hand-side panel shows t statistics for positive (S^+) and negative (S^-) news variables, and trading variables including volume (dv), order imbalance (do), and bid/ask spread (ds), as well as the control variable, volatility (Vol), following: $Z_t^d = \beta_0 + \beta_1 S_t^+ + \beta_2 S_t^- + \beta_3 dv_t + \beta_4 do_t + \beta_5 ds_t + \beta_6 Vol_t + \epsilon_t$; the right-hand-side panel shows t statistics with the asymmetric news impact variable $S^+ - S^-$, following: $Z_t^d = \beta_0 + \beta_1(S_t^+ - S_t^-) + \beta_2 dv_t + \beta_3 do_t + \beta_4 ds_t + \beta_5 Vol_t + \epsilon_t$.

Table 18: Regression for positive jumps, different news impact

	$S+$	$S-$	dv	do	ds	vol	dS	dv	do	ds	vol	obs.
	positive and negative news						news difference					
AXP	0.53	1.17	1.58	0.03	2.25	0.73	-0.23	1.51	-0.04	2.38	0.74	229
BA	0.61	0.92	2.89	-0.98	0.67	-0.25	-0.09	2.91	-0.98	0.62	-0.25	351
CAT	0.41	1.13	1.75	6.42	1.42	-0.02	-0.56	1.72	6.54	1.59	0.39	225
CVX	0.41	-0.37	5.32	0.11	0.39	0.30	0.61	5.36	0.11	0.39	0.30	140
DIS	0.50	-0.65	-0.61	2.93	0.74	0.70	0.96	-0.61	2.93	0.74	0.68	210
GE	2.06	0.21	1.45	1.31	-1.06	3.13	1.54	1.49	1.32	-1.07	4.30	148
GS	-0.24	-0.69	1.80	-2.00	6.72	0.16	0.39	1.82	-2.00	6.72	0.06	299
HD	-0.53	0.19	1.78	-0.35	3.47	-0.32	-0.70	1.79	-0.35	3.48	-0.38	134
IBM	5.38	0.85	2.48	-0.53	3.28	1.87	3.95	2.53	-0.63	3.22	2.05	424
JNJ	-0.34	0.44	3.49	-0.44	2.87	0.31	-0.63	3.49	-0.45	2.90	0.32	267
JPM	4.02	2.35	0.31	-0.90	2.79	3.43	0.50	0.25	-1.04	2.44	3.55	477
KO	-1.84	1.07	2.17	3.95	-0.85	1.28	-2.47	2.18	3.95	-0.85	1.29	278
MCD	-0.34	2.09	-1.02	0.31	3.50	3.02	-1.71	-1.03	0.33	3.54	3.47	249
MMM	-0.38	-0.84	-1.78	-2.89	5.38	2.65	0.34	-1.77	-2.95	5.46	2.54	190
MRK	6.91	3.11	2.09	-1.78	9.56	2.19	3.46	1.68	-1.82	9.40	2.43	322
NKE	-0.02	0.40	-0.52	2.48	0.20	0.31	-0.27	-0.54	2.51	0.22	0.36	116
PFE	1.34	-0.45	0.17	1.16	1.78	2.77	1.45	0.18	1.19	1.78	2.84	66
PG	0.01	0.43	2.22	-0.83	2.90	1.63	-0.34	2.24	-0.84	2.92	1.63	256
UNH	-0.82	1.43	0.30	1.01	1.12	-0.19	-1.46	0.30	1.00	1.25	0.09	110
V	0.12	0.13	1.43	3.30	2.49	0.26	0.04	1.43	3.32	2.51	0.32	94
VZ	-0.17	1.69	0.59	-0.50	0.71	2.22	-1.04	0.89	-0.66	0.58	2.37	131
WMT	2.48	1.98	3.66	-0.14	1.60	1.96	0.70	4.37	-0.32	1.20	2.08	343
XOM	-0.73	-0.43	-1.03	-1.64	8.78	0.34	-0.24	-1.07	-1.62	8.81	0.28	464
CSCO	0.81	-0.08	0.23	-0.24	1.52	0.83	0.86	0.27	-0.29	1.59	1.34	48
INTC	1.47	1.74	1.02	0.13	2.63	1.25	1.07	1.06	-0.08	2.70	1.46	82
MSFT	0.35	0.08	0.80	-0.20	8.44	-1.93	0.30	0.80	-0.20	8.54	-1.93	227
avg.	0.85	0.69	1.25	0.37	2.82	1.10	0.25	1.28	0.34	2.81	1.24	226

Notes: the left-hand-side panel shows t statistics for positive (S^+) and negative (S^-) news variables, and trading variables including volume (dv), order imbalance (do), and bid/ask spread (ds), as well as the control variable, volatility (Vol) on positive jumps $l+$, following: $Z_t^{l+} = \beta_0 + \beta_1 S_t^+ + \beta_2 S_t^- + \beta_3 dv_t + \beta_4 do_t + \beta_5 ds_t + \beta_6 Vol_t + \epsilon_t$; the right-hand-side panel shows t statistics with the asymmetric news impact variable $S^+ - S^-$, following: $Z_t^{l+} = \beta_0 + \beta_1(S_t^+ - S_t^-) + \beta_2 dv_t + \beta_3 do_t + \beta_4 ds_t + \beta_5 Vol_t + \epsilon_t$. The last column records the numbers of positive-jump days (with news) used for regressions.

Table 19: Regression for negative jumps, different news impact

	$S+$	$S-$	dv	do	ds	vol	dS	dv	do	ds	vol	obs.
	positive and negative news						news difference					
AXP	0.99	3.22	-2.26	-1.18	6.38	0.54	0.87	-2.16	-1.05	6.28	0.62	226
BA	1.40	4.91	0.70	2.49	0.77	-0.21	2.57	0.76	2.48	0.47	-0.11	335
CAT	1.92	2.92	-0.99	12.24	2.87	-1.18	0.33	-0.99	11.82	3.31	-0.32	244
CVX	0.47	1.10	3.08	-0.93	0.26	-0.47	0.29	3.08	-0.75	0.38	-0.14	115
DIS	-0.54	2.22	-3.24	3.36	7.38	0.82	1.84	-3.32	3.33	7.42	0.75	209
GE	3.59	0.81	0.35	2.41	3.99	0.36	-2.19	0.25	2.24	3.91	0.31	119
GS	1.57	0.80	-0.86	0.16	2.98	0.32	-0.50	-0.78	0.20	2.95	0.42	288
HD	-0.86	-1.38	2.39	-1.58	5.72	-0.63	-0.30	2.32	-1.46	5.85	-0.41	110
IBM	-0.23	0.36	2.64	-0.51	2.12	0.79	0.47	2.64	-0.52	2.14	0.79	383
JNJ	1.89	2.36	7.73	-0.14	1.13	2.20	-0.07	7.59	-0.37	1.20	2.03	246
JPM	1.88	1.50	-0.50	0.78	4.69	3.55	-0.20	-0.45	0.70	4.57	3.80	461
KO	-0.92	-0.64	-0.74	0.92	0.51	-1.40	0.22	-0.73	0.99	0.46	-1.55	281
MCD	0.40	1.43	0.41	-0.94	0.89	-0.51	0.29	0.38	-0.94	0.98	-0.28	294
MMM	-0.71	0.52	1.14	3.14	1.08	0.12	0.99	1.20	3.18	1.18	0.12	169
MRK	-0.10	0.03	-1.99	8.12	6.68	1.08	0.12	-2.00	8.14	6.70	1.09	292
NKE	-0.96	0.25	1.70	-0.66	0.94	-0.52	1.08	1.74	-0.67	0.94	-0.68	93
PFE	1.35	0.01	-0.59	-0.23	-0.08	1.65	-1.34	-0.57	-0.28	-0.01	1.60	71
PG	-1.50	-1.29	1.09	4.65	-0.17	1.59	0.38	1.03	4.44	-0.04	1.59	256
UNH	-0.41	1.79	0.93	0.16	1.43	0.99	1.56	0.91	0.16	1.33	1.09	94
V	0.72	0.68	0.04	3.77	0.27	0.27	-0.19	-0.15	4.50	0.30	0.57	94
VZ	0.74	2.54	-1.03	0.88	2.53	-0.75	1.27	-0.44	0.80	2.64	-0.57	100
WMT	2.07	1.48	0.30	-0.68	2.68	1.06	-0.34	0.22	-0.73	2.67	0.98	309
XOM	1.15	2.87	2.94	1.79	2.94	0.01	1.10	2.79	1.62	3.28	-0.06	530
CSCO	3.82	0.97	0.42	-0.31	-0.92	1.68	-2.38	0.20	-0.10	-0.78	2.99	49
INTC	-0.64	1.18	0.43	-0.31	0.93	3.69	1.26	0.44	-0.31	1.01	3.70	94
MSFT	2.52	-0.69	2.89	-1.05	3.56	2.90	-2.71	2.91	-1.08	3.63	2.90	276
avg.	0.75	1.15	0.65	1.40	2.37	0.69	0.17	0.65	1.40	2.41	0.82	221

Notes: the left-hand-side panel shows t statistics for positive (S^+) and negative (S^-) news variables, and trading variables including volume (dv), order imbalance (do), and bid/ask spread (ds), as well as the control variable, volatility (Vol) on negative jumps $l-$, following: $Z_t^{l-} = \beta_0 + \beta_1 S_t^+ + \beta_2 S_t^- + \beta_3 dv_t + \beta_4 do_t + \beta_5 ds_t + \beta_6 Vol_t + \epsilon_t$; the right-hand-side panel shows t statistics with the asymmetric news impact variable $S^+ - S^-$, following: $Z_t^{l-} = \beta_0 + \beta_1(S_t^- - S_t^+) + \beta_2 dv_t + \beta_3 do_t + \beta_4 ds_t + \beta_5 Vol_t + \epsilon_t$. The last column records the numbers of negative-jump days (with news) used for regressions.

Table 20: Hit rates and jump sizes by DV^c: individual stocks

	hit	size (δ)	hit	size (δ)	hit	size (δ)	hit	size (δ)	hit	size (δ)
α	10^{-5}		10^{-4}		10^{-3}		0.01		0.05	
AA	1.00	8.70	1.00	8.12	0.99	7.18	0.99	6.29	0.99	5.62
AXP	1.00	9.53	0.99	8.80	0.99	7.84	0.99	6.77	0.99	6.04
BA	0.99	10.01	1.00	9.25	1.00	8.27	0.98	7.10	0.99	6.36
CVX	1.00	10.22	0.99	9.45	1.00	8.60	0.99	7.61	0.99	6.75
CAT	0.99	9.23	1.00	8.44	0.99	7.64	0.99	6.84	0.98	6.18
DD	1.00	8.77	0.99	8.01	0.99	7.26	0.99	6.35	1.00	5.71
DIS	1.00	9.76	0.99	8.86	0.99	7.86	0.99	6.77	0.98	6.07
GE	1.00	8.70	0.98	7.88	1.00	7.10	1.00	6.28	0.98	5.76
GS	1.00	9.83	1.00	8.87	1.00	8.23	0.99	7.11	0.98	6.42
HD	1.00	8.62	1.00	7.93	1.00	7.31	1.00	6.48	0.98	5.84
IBM	1.00	9.64	1.00	8.95	0.99	8.09	0.99	7.06	0.98	6.22
JNJ	1.00	9.68	1.00	8.79	1.00	7.97	0.99	6.99	0.98	6.21
JPM	0.99	9.67	1.00	8.92	0.98	8.00	0.99	7.10	0.98	6.30
KO	0.99	8.69	1.00	8.12	1.00	7.21	1.00	6.24	0.99	5.58
MCD	0.99	9.00	1.00	8.36	1.00	7.53	0.98	6.60	0.98	5.92
MMM	1.00	9.28	1.00	8.57	0.99	7.85	0.99	6.90	0.98	6.20
MRK	1.00	9.59	0.99	8.71	0.99	7.78	0.99	6.84	0.99	6.07
NKE	1.00	9.46	1.00	8.76	0.99	7.82	0.99	6.99	0.98	6.26
PFE	1.00	9.12	1.00	8.29	1.00	7.60	0.98	6.85	0.99	6.03
PG	1.00	9.02	1.00	8.28	0.99	7.42	0.99	6.55	0.99	5.98
TRV	1.00	10.10	1.00	8.85	1.00	7.82	0.99	6.72	0.99	5.89
UNH	1.00	9.77	1.00	8.98	1.00	8.07	1.00	7.17	0.98	6.37
V	1.00	9.90	0.98	9.12	1.00	8.15	0.98	7.09	0.99	6.41
VZ	1.00	10.05	1.00	9.28	0.99	8.25	1.00	7.07	0.99	6.24
UTX	0.99	9.25	1.00	8.54	1.00	7.79	0.99	6.79	0.98	6.11
WMT	0.99	9.02	0.99	8.25	1.00	7.53	0.99	6.55	0.98	5.90
XOM	1.00	9.60	1.00	8.81	0.99	7.97	0.99	7.03	0.98	6.43
CSCO	1.00	9.58	1.00	9.05	1.00	8.06	1.00	7.16	0.97	6.27
INTC	1.00	10.27	1.00	9.10	1.00	8.34	0.99	7.41	0.98	6.71
MSFT	1.00	11.55	0.98	10.68	0.98	9.47	0.98	8.37	0.97	7.55
avg.	1.00	9.52	1.00	8.73	0.99	7.87	0.99	6.90	0.98	6.18

Table 21: Hit rates and jump sizes by other local methods: individual stocks ($\alpha = 10^{-5}$)

	hit	size (δ)	hit	size (δ)	hit	size (δ)	hit	size (δ)	hit	size (δ)
	SBV_{300}		SBV_{60}		SBV_{30}		$PABV_1$		$PABV_2$	
AA	0.48	3.16	0.34	2.17	0.26	1.80	0.49	4.63	0.38	5.65
AXP	0.48	3.91	0.34	3.20	0.30	2.74	0.61	5.68	0.56	5.74
BA	0.35	3.71	0.30	3.20	0.28	2.84	0.51	5.78	0.58	6.72
CVX	0.41	4.97	0.35	4.04	0.32	3.74	0.70	7.74	0.63	6.18
CAT	0.39	3.90	0.30	3.32	0.27	2.96	0.50	5.38	0.56	5.72
DD	0.41	3.12	0.38	2.80	0.31	2.41	0.51	5.27	0.54	4.70
DIS	0.39	3.81	0.31	2.84	0.28	2.49	0.55	5.46	0.48	6.32
GE	0.48	1.86	0.31	1.43	0.26	1.26	0.50	3.36	0.50	3.78
GS	0.27	3.38	0.32	3.28	0.24	2.90	0.57	4.94	0.33	1.84
HD	0.47	3.16	0.34	2.89	0.31	2.63	0.51	3.96	0.43	3.63
IBM	0.38	3.91	0.33	3.34	0.29	3.02	0.48	5.34	0.38	5.39
JNJ	0.43	4.00	0.36	3.34	0.31	2.86	0.49	4.94	0.52	5.53
JPM	0.43	4.01	0.35	3.41	0.33	2.97	0.66	6.24	0.57	6.86
KO	0.40	2.44	0.32	2.25	0.28	1.92	0.61	4.95	0.52	4.96
MCD	0.39	3.66	0.35	3.02	0.30	2.65	0.46	4.43	0.35	4.26
MMM	0.46	3.94	0.35	3.40	0.29	2.95	0.55	6.54	0.59	9.42
MRK	0.41	4.11	0.33	2.99	0.30	2.51	0.56	7.11	0.51	9.58
NKE	0.37	3.40	0.25	3.05	0.23	2.76	0.55	7.12	0.50	7.64
PFE	0.41	2.29	0.26	1.75	0.23	1.43	0.37	4.06	0.50	4.88
PG	0.36	3.51	0.35	3.18	0.30	2.76	0.54	6.35	0.53	7.50
TRV	0.50	3.55	0.37	3.03	0.24	2.53	0.36	3.57	0.00	2.12
UNH	0.34	3.56	0.31	3.32	0.26	3.00	0.42	4.61	0.21	3.64
V	0.38	3.88	0.26	3.05	0.25	2.87	0.42	5.90	0.47	7.44
VZ	0.34	3.00	0.27	2.26	0.23	1.84	0.50	7.43	0.83	13.17
UTX	0.39	3.78	0.36	3.35	0.31	2.92	0.47	5.19	0.52	5.32
WMT	0.41	3.75	0.30	3.03	0.29	2.63	0.63	5.49	0.55	6.32
XOM	0.46	4.87	0.41	4.25	0.38	3.63	0.65	7.67	0.62	11.00
CSCO	0.40	2.42	0.29	1.85	0.24	1.60	0.63	5.81	0.40	5.35
INTC	0.38	3.38	0.25	2.43	0.23	2.10	0.70	11.57	0.78	11.28
MSFT	0.35	5.00	0.28	4.10	0.25	3.46	0.42	12.42	0.38	10.51
avg.	0.40	3.58	0.32	2.98	0.28	2.60	0.53	5.96	0.49	6.42

Table 22: Powers of DV^c and DV^C with single jumps, jump size = $10s$

	DV^c					DV^C		
α	10^{-5}	10^{-4}	10^{-3}	0.01	0.05	10^{-3}	0.01	0.05
AA	.33	.51	.81	.97	1.00	.00	.01	.27
AXP	.54	.70	.76	.91	.98	.00	.09	.68
BA	.81	.95	.98	1.00	1.00	.00	.34	.98
CVX	.78	.91	.98	1.00	1.00	.01	.23	.94
CAT	.70	.89	1.00	1.00	1.00	.00	.36	.95
DD	.52	.69	.84	.94	.99	.00	.06	.70
DIS	.19	.37	.71	.94	.99	.00	.02	.21
GE	.18	.42	.78	.97	1.00	.00	.02	.12
GS	1.00	1.00	1.00	1.00	1.00	.00	1.00	1.00
HD	.39	.50	.74	.93	.98	.00	.06	.43
IBM	.87	.96	.99	.99	.99	.01	.50	.98
JNJ	.36	.48	.72	.91	.96	.00	.04	.46
JPM	.18	.31	.57	.85	.96	.00	.01	.29
KO	.19	.35	.64	.89	.96	.00	.01	.24
MCD	.41	.57	.78	.96	.99	.00	.07	.54
MMM	.72	.91	.96	.99	1.00	.00	.20	.93
MRK	.19	.33	.60	.87	.97	.00	.01	.28
NKE	.32	.55	.74	.92	.95	.00	.16	.58
PFE	.06	.24	.57	.87	.97	.00	.00	.04
PG	.37	.47	.64	.83	.95	.00	.07	.52
TRV	.57	.86	1.00	1.00	1.00	.00	.29	1.00
UNH	.71	1.00	1.00	1.00	1.00	.00	.64	1.00
V	.55	.71	.84	.92	.96	.00	.43	.78
VZ	.03	.13	.43	.79	.93	.00	.00	.16
UTX	.69	.90	.96	.99	1.00	.00	.13	.90
WMT	.27	.44	.64	.86	.96	.00	.04	.45
XOM	.28	.39	.59	.83	.93	.00	.03	.47
CSCO	.00	.03	.23	.65	.97	.00	.00	.25
INTC	.02	.05	.34	.63	.89	.00	.00	.29
MSFT	.06	.09	.24	.67	.85	.00	.00	.33
avg.	.41	.56	.74	.90	.97	.00	.16	.56

Table 23: Powers of DV^c and DV^C with gradual reversals, jump size = 10s

	DV^c					DV^C		
α	10^{-5}	10^{-4}	10^{-3}	0.01	0.05	10^{-3}	0.01	0.05
AA	.30	.50	.80	.97	1.00	.00	.01	.27
AXP	.54	.69	.76	.91	.98	.00	.08	.68
BA	.79	.95	.98	1.00	1.00	.00	.35	.98
CVX	.77	.91	.98	1.00	1.00	.01	.23	.96
CAT	.70	.87	1.00	1.00	1.00	.00	.34	.95
DD	.51	.68	.82	.94	.99	.00	.05	.71
DIS	.18	.36	.70	.94	.99	.00	.02	.21
GE	.16	.38	.76	.96	1.00	.00	.02	.11
GS	1.00	1.00	1.00	1.00	1.00	.00	1.00	1.00
HD	.38	.49	.73	.92	.98	.00	.06	.43
IBM	.85	.96	.99	.99	.99	.01	.49	.98
JNJ	.36	.48	.72	.90	.96	.00	.04	.47
JPM	.18	.31	.56	.84	.96	.00	.01	.29
KO	.19	.35	.63	.88	.96	.00	.01	.24
MCD	.40	.56	.78	.96	.99	.00	.07	.54
MMM	.72	.91	.96	.99	1.00	.00	.21	.93
MRK	.18	.33	.59	.86	.97	.00	.01	.28
NKE	.32	.55	.74	.92	.95	.00	.13	.58
PFE	.05	.23	.56	.87	.97	.00	.00	.04
PG	.36	.46	.63	.83	.95	.00	.06	.52
TRV	.57	.86	1.00	1.00	1.00	.00	.29	1.00
UNH	.71	1.00	1.00	1.00	1.00	.00	.64	1.00
V	.55	.71	.84	.92	.96	.00	.43	.80
VZ	.02	.12	.41	.78	.93	.00	.00	.15
UTX	.67	.90	.96	.99	1.00	.00	.12	.91
WMT	.27	.43	.64	.86	.96	.00	.04	.45
XOM	.27	.38	.58	.82	.93	.00	.03	.48
CSCO	.00	.03	.23	.63	.95	.00	.00	.25
INTC	.02	.05	.34	.63	.89	.00	.00	.27
MSFT	.06	.09	.24	.67	.85	.00	.00	.36
avg.	.40	.55	.73	.90	.97	.00	.16	.56

Table 24: Powers of competing tests with single jumps, jump size = 10s

	MED_{30}			MED_{60}			SBV_{300}		SBV_{60}		$PABV_1$		$PABV_2$	
α	10^{-3}	.01	.05	10^{-3}	.01	.05	10^{-5}	10^{-4}	10^{-5}	10^{-4}	10^{-5}	10^{-4}	10^{-5}	10^{-4}
AA	.32	.60	.82	.15	.31	.59	.13	.18	.81	.85	.23	.33	.04	.07
AXP	.20	.41	.67	.09	.19	.37	.09	.14	.54	.65	.18	.20	.04	.07
BA	.22	.55	.85	.02	.18	.40	.07	.15	.71	.81	.13	.19	.02	.02
CVX	.04	.38	.79	.00	.04	.33	.02	.02	.56	.71	.03	.08	.00	.00
CAT	.15	.54	.89	.00	.15	.38	.00	.02	.67	.77	.03	.07	.00	.00
DD	.21	.45	.74	.10	.24	.43	.06	.10	.65	.71	.21	.25	.03	.06
DIS	.22	.47	.75	.12	.29	.52	.14	.19	.72	.80	.30	.38	.09	.13
GE	.29	.55	.76	.19	.41	.66	.13	.20	.82	.87	.39	.50	.08	.13
GS	.00	1.00	1.00	.00	.50	1.00	.00	.00	1.00	1.00	.00	.00	.00	.00
HD	.23	.39	.67	.11	.22	.39	.08	.12	.62	.72	.20	.24	.04	.06
IBM	.29	.70	.91	.07	.25	.55	.05	.11	.81	.88	.25	.32	.04	.06
JNJ	.20	.39	.67	.11	.20	.43	.10	.15	.62	.70	.23	.29	.09	.13
JPM	.11	.25	.50	.05	.13	.26	.04	.10	.44	.50	.12	.17	.03	.05
KO	.19	.43	.70	.08	.22	.47	.09	.13	.65	.72	.18	.25	.04	.07
MCD	.31	.57	.80	.18	.31	.53	.18	.22	.70	.78	.27	.33	.12	.14
MMM	.14	.51	.83	.01	.14	.38	.03	.07	.63	.73	.07	.12	.01	.02
MRK	.13	.33	.66	.06	.16	.36	.06	.10	.57	.63	.14	.20	.04	.06
NKE	.05	.32	.71	.03	.03	.21	.03	.08	.42	.47	.00	.05	.00	.00
PFE	.21	.36	.62	.07	.18	.31	.05	.10	.49	.60	.11	.16	.02	.04
PG	.15	.30	.65	.07	.15	.33	.07	.11	.49	.57	.13	.19	.05	.07
TRV	.29	.71	.86	.00	.00	.29	.00	.00	.71	.86	.00	.00	.00	.00
UNH	.36	.71	1.00	.07	.07	.43	.00	.07	.71	.93	.00	.00	.00	.00
V	.18	.47	.78	.00	.18	.41	.04	.06	.61	.69	.02	.06	.00	.00
VZ	.05	.23	.55	.00	.05	.20	.02	.05	.40	.49	.03	.04	.00	.01
UTX	.23	.50	.83	.05	.22	.55	.06	.10	.68	.78	.18	.22	.02	.05
WMT	.12	.27	.60	.04	.11	.26	.05	.07	.46	.57	.09	.13	.02	.03
XOM	.09	.22	.47	.04	.08	.21	.03	.06	.34	.45	.09	.12	.03	.03
CSCO	.00	.08	.47	.00	.02	.12	.00	.05	.20	.33	.00	.00	.00	.00
INTC	.02	.04	.50	.00	.02	.02	.00	.07	.13	.25	.00	.04	.00	.00
MSFT	.00	.15	.61	.00	.03	.12	.00	.03	.36	.45	.00	.00	.00	.00
avg.	.17	.43	.72	.06	.17	.38	.05	.09	.58	.68	.12	.16	.03	.04

Table 25: Powers of competing tests with gradual reversals, jump size = 10s

	MED_{30}			MED_{60}			SBV_{300}		SBV_{60}		$PABV_1$		$PABV_2$	
α	10^{-3}	.01	.05	10^{-3}	.01	.05	10^{-5}	10^{-4}	10^{-5}	10^{-4}	10^{-5}	10^{-4}	10^{-5}	10^{-4}
AA	.05	.08	.15	.00	.02	.03	.02	.03	.35	.42	.00	.00	.00	.00
AXP	.02	.03	.09	.00	.00	.00	.01	.02	.21	.29	.00	.01	.00	.00
BA	.01	.06	.13	.00	.00	.02	.01	.06	.33	.39	.00	.01	.00	.00
CVX	.00	.03	.14	.00	.00	.01	.00	.00	.22	.34	.00	.00	.00	.00
CAT	.00	.03	.15	.00	.00	.00	.00	.02	.21	.25	.00	.00	.00	.00
DD	.03	.06	.15	.00	.00	.01	.02	.04	.27	.35	.00	.01	.00	.00
DIS	.01	.04	.09	.00	.00	.01	.01	.03	.32	.40	.00	.01	.00	.00
GE	.02	.04	.07	.00	.01	.01	.02	.03	.31	.37	.00	.01	.00	.00
GS	.50	.50	1.00	.00	.00	.00	.00	.00	1.00	1.00	.00	.00	.00	.00
HD	.01	.03	.07	.00	.00	.01	.01	.03	.22	.31	.00	.01	.00	.00
IBM	.02	.04	.11	.00	.00	.01	.02	.03	.27	.40	.00	.00	.00	.01
JNJ	.00	.02	.06	.00	.00	.00	.01	.03	.21	.29	.00	.01	.00	.00
JPM	.00	.01	.03	.00	.00	.00	.00	.02	.15	.20	.00	.00	.00	.00
KO	.02	.05	.13	.00	.00	.01	.01	.03	.28	.36	.00	.00	.00	.00
MCD	.02	.05	.14	.00	.00	.01	.01	.03	.29	.34	.00	.00	.00	.00
MMM	.01	.05	.15	.00	.00	.00	.01	.05	.24	.36	.00	.00	.00	.00
MRK	.01	.03	.09	.00	.00	.01	.01	.02	.22	.29	.00	.01	.00	.00
NKE	.00	.00	.11	.00	.00	.00	.00	.03	.16	.26	.00	.00	.00	.00
PFE	.04	.08	.17	.00	.01	.01	.01	.03	.24	.31	.00	.00	.00	.00
PG	.01	.02	.07	.00	.00	.00	.00	.03	.16	.22	.00	.01	.00	.00
TRV	.00	.00	.14	.00	.00	.00	.00	.00	.29	.43	.00	.00	.00	.00
UNH	.07	.07	.14	.00	.00	.00	.00	.07	.21	.29	.00	.00	.00	.00
V	.00	.06	.16	.00	.00	.00	.00	.00	.20	.33	.00	.00	.00	.00
VZ	.01	.06	.18	.00	.01	.02	.00	.02	.18	.26	.00	.00	.00	.00
UTX	.01	.05	.13	.00	.00	.01	.02	.04	.30	.38	.00	.00	.00	.00
WMT	.01	.01	.09	.00	.00	.00	.00	.01	.15	.25	.00	.01	.00	.00
XOM	.00	.00	.03	.00	.00	.00	.00	.01	.10	.16	.00	.01	.00	.00
CSCO	.00	.00	.08	.00	.00	.02	.00	.03	.05	.15	.00	.00	.00	.02
INTC	.00	.04	.09	.00	.00	.00	.00	.05	.05	.14	.00	.02	.00	.00
MSFT	.03	.03	.33	.00	.00	.00	.00	.00	.21	.24	.00	.00	.00	.00
avg.	.03	.05	.15	.00	.00	.01	.01	.03	.25	.33	.00	.00	.00	.00

Table 26: Common proportions by our daily and local tests

α^c	10^{-5}			10^{-4}			10^{-3}			0.01		
α^C	.01	.05	.10	.01	.05	.10	.01	.05	.10	.01	.05	.10
AA	0.78	0.98	1.00	0.66	0.92	0.98	0.51	0.81	0.93	0.34	0.65	0.81
AXP	0.86	1.00	1.00	0.78	0.99	1.00	0.64	0.93	0.99	0.51	0.82	0.97
BA	0.92	1.00	1.00	0.88	1.00	1.00	0.77	0.97	1.00	0.60	0.92	0.98
CAT	0.87	1.00	1.00	0.78	1.00	1.00	0.71	0.98	1.00	0.60	0.93	0.99
CVX	0.97	1.00	1.00	0.92	0.99	1.00	0.87	0.99	1.00	0.76	0.96	1.00
DD	0.81	1.00	1.00	0.64	0.94	1.00	0.52	0.83	0.98	0.36	0.69	0.92
DIS	0.90	0.99	1.00	0.81	0.95	0.99	0.74	0.92	0.98	0.59	0.84	0.94
GE	0.80	0.95	1.00	0.65	0.87	0.97	0.53	0.77	0.86	0.43	0.65	0.80
GS	0.87	1.00	1.00	0.78	1.00	1.00	0.68	0.98	1.00	0.43	0.92	0.99
HD	0.78	0.96	1.00	0.68	0.90	0.99	0.61	0.86	0.97	0.46	0.80	0.94
IBM	0.91	0.99	1.00	0.84	0.99	1.00	0.76	0.97	1.00	0.62	0.94	1.00
JNJ	0.89	0.99	1.00	0.83	0.96	0.99	0.75	0.92	0.99	0.64	0.85	0.96
JPM	0.90	0.99	1.00	0.86	0.98	1.00	0.77	0.93	0.98	0.70	0.87	0.95
KO	0.67	0.95	1.00	0.63	0.90	0.99	0.54	0.79	0.94	0.37	0.63	0.83
MCD	0.82	1.00	1.00	0.75	0.98	1.00	0.64	0.91	0.98	0.50	0.82	0.95
MMM	0.90	1.00	1.00	0.82	1.00	1.00	0.74	0.98	1.00	0.60	0.92	0.99
MRK	0.84	0.97	1.00	0.78	0.94	0.99	0.69	0.90	0.98	0.59	0.80	0.93
NKE	0.94	1.00	1.00	0.87	0.99	1.00	0.78	0.98	1.00	0.67	0.95	0.99
PFE	0.71	0.90	1.00	0.68	0.82	0.98	0.65	0.79	0.95	0.61	0.77	0.91
PG	0.81	1.00	1.00	0.72	0.97	1.00	0.63	0.93	0.98	0.52	0.87	0.96
TRV	0.80	1.00	1.00	0.70	1.00	1.00	0.50	0.87	1.00	0.29	0.78	0.97
UNH	0.92	1.00	1.00	0.86	1.00	1.00	0.77	0.99	1.00	0.60	0.91	0.99
V	0.96	1.00	1.00	0.92	1.00	1.00	0.84	0.99	1.00	0.74	0.92	0.99
VZ	0.82	1.00	1.00	0.72	0.97	0.98	0.66	0.91	0.99	0.52	0.77	0.86
UTX	0.89	1.00	1.00	0.82	0.99	1.00	0.72	0.97	1.00	0.60	0.92	0.98
WMT	0.85	0.99	1.00	0.79	0.97	1.00	0.70	0.92	0.98	0.55	0.82	0.94
XOM	0.87	1.00	1.00	0.80	0.98	1.00	0.71	0.94	0.99	0.63	0.89	0.98
CSCO	0.90	1.00	1.00	0.85	1.00	1.00	0.72	0.92	1.00	0.68	0.88	0.99
INTC	0.98	1.00	1.00	0.93	0.99	1.00	0.88	0.99	1.00	0.83	0.93	0.98
MSFT	1.00	1.00	1.00	0.99	1.00	1.00	0.97	0.99	1.00	0.94	0.98	0.99
avg.	0.86	0.99	1.00	0.79	0.97	1.00	0.70	0.92	0.98	0.58	0.85	0.95

Table 27: JV differences between our daily and local tests on common jump days

α^c	10^{-5}			10^{-4}			10^{-3}			0.01		
α^C	.01	.05	.10	.01	.05	.10	.01	.05	.10	.01	.05	.10
AA	.047	.037	.036	.046	.035	.032	.054	.043	.037	.048	.043	.036
AXP	.064	.058	.058	.066	.058	.057	.068	.059	.056	.068	.061	.056
BA	.072	.069	.069	.074	.069	.069	.078	.070	.069	.081	.072	.069
CAT	.081	.075	.075	.083	.074	.073	.082	.071	.071	.081	.072	.070
CVX	.082	.081	.081	.081	.079	.078	.083	.079	.079	.084	.078	.077
DD	.059	.052	.052	.060	.051	.047	.060	.050	.045	.053	.049	.043
DIS	.093	.085	.084	.091	.081	.078	.089	.079	.075	.085	.074	.069
GE	.058	.047	.045	.059	.047	.043	.057	.045	.041	.069	.054	.047
GS	.071	.067	.067	.084	.074	.074	.084	.074	.072	.085	.075	.072
HD	.070	.060	.058	.075	.064	.057	.080	.068	.062	.078	.067	.061
IBM	.077	.073	.073	.077	.072	.071	.079	.072	.071	.081	.074	.071
JNJ	.072	.068	.067	.071	.067	.064	.073	.067	.064	.073	.068	.063
JPM	.090	.085	.084	.092	.085	.083	.089	.081	.078	.089	.080	.075
KO	.057	.045	.043	.063	.049	.045	.062	.052	.045	.062	.055	.048
MCD	.075	.063	.063	.075	.063	.061	.076	.064	.059	.076	.064	.059
MMM	.084	.079	.079	.086	.078	.078	.087	.077	.076	.086	.076	.073
MRK	.076	.068	.066	.078	.070	.066	.077	.068	.063	.075	.067	.061
NKE	.075	.071	.071	.077	.073	.072	.083	.076	.075	.085	.077	.075
PFE	.090	.070	.061	.093	.076	.060	.093	.078	.063	.084	.070	.057
PG	.069	.059	.059	.070	.059	.058	.074	.062	.060	.074	.063	.060
TRV	.048	.040	.040	.061	.048	.048	.061	.048	.041	.064	.055	.047
UNH	.090	.086	.086	.086	.080	.080	.087	.079	.079	.086	.077	.073
V	.095	.093	.093	.094	.089	.089	.089	.083	.082	.088	.081	.078
VZ	.061	.054	.054	.057	.049	.049	.063	.054	.052	.057	.051	.048
UTX	.087	.079	.079	.091	.080	.080	.091	.079	.077	.088	.077	.074
WMT	.087	.081	.081	.089	.080	.080	.089	.079	.078	.086	.077	.074
XOM	.071	.066	.065	.071	.065	.063	.072	.065	.062	.070	.064	.060
CSCO	.090	.084	.084	.088	.080	.080	.087	.077	.073	.087	.078	.072
INTC	.107	.105	.105	.108	.104	.103	.107	.101	.100	.104	.098	.095
MSFT	.155	.155	.155	.154	.153	.152	.150	.148	.147	.143	.139	.138
avg.	.078	.072	.071	.080	.072	.070	.081	.072	.068	.080	.071	.067

Table 28: JF and JV by DV^C and MED_{30}

	DV^C								MED_{30}			
α	10^{-3}		0.01		0.05		0.1		10^{-3}		0.01	
AA	.012	.215	.027	.185	.072	.153	.139	.130	.282	.303	.428	.253
AXP	.024	.211	.073	.157	.235	.122	.414	.108	.346	.263	.545	.218
BA	.039	.178	.112	.148	.321	.121	.541	.108	.359	.258	.579	.220
CAT	.039	.182	.135	.147	.419	.120	.653	.109	.342	.267	.527	.226
CVX	.159	.148	.335	.131	.614	.115	.790	.107	.113	.188	.312	.153
DD	.011	.215	.035	.166	.123	.123	.254	.104	.224	.252	.444	.198
DIS	.047	.174	.118	.140	.272	.116	.383	.106	.127	.211	.312	.169
GE	.012	.190	.021	.173	.034	.152	.048	.133	.228	.234	.394	.197
GS	.008	.209	.048	.171	.274	.139	.565	.125	.810	.339	.921	.318
HD	.020	.185	.075	.145	.250	.120	.412	.108	.246	.248	.441	.202
IBM	.033	.179	.116	.143	.377	.115	.631	.103	.281	.235	.505	.199
JNJ	.055	.182	.139	.146	.323	.119	.470	.107	.169	.217	.357	.178
JPM	.119	.146	.219	.129	.355	.114	.464	.105	.093	.254	.222	.193
KO	.013	.180	.031	.154	.077	.128	.148	.109	.271	.240	.469	.200
MCD	.025	.200	.079	.158	.242	.127	.427	.112	.321	.264	.526	.218
MMM	.034	.215	.102	.173	.287	.139	.510	.121	.504	.296	.698	.255
MRK	.053	.200	.116	.161	.227	.133	.338	.117	.206	.253	.402	.203
NKE	.059	.161	.180	.137	.468	.118	.633	.110	.317	.214	.573	.185
PFE	.023	.169	.035	.149	.054	.129	.073	.116	.256	.218	.478	.184
PG	.041	.183	.112	.145	.283	.120	.442	.106	.194	.226	.405	.184
TRV	.004	.298	.016	.212	.118	.159	.295	.137	.838	.401	.928	.378
UNH	.031	.191	.104	.159	.345	.130	.603	.117	.553	.285	.750	.250
V	.079	.157	.210	.134	.447	.116	.611	.108	.307	.291	.483	.236
VZ	.027	.201	.047	.164	.089	.130	.131	.112	.227	.215	.445	.181
UTX	.045	.216	.120	.176	.289	.140	.472	.120	.319	.263	.536	.217
WMT	.034	.183	.101	.145	.252	.118	.388	.105	.214	.232	.403	.192
XOM	.094	.149	.200	.123	.403	.102	.567	.093	.073	.222	.195	.175
CSCO	.034	.168	.061	.143	.109	.123	.161	.110	.575	.285	.747	.255
INTC	.104	.170	.162	.149	.254	.130	.322	.120	.417	.272	.600	.239
MSFT	.411	.171	.456	.165	.525	.156	.595	.148	.228	.241	.410	.202
avg.	.056	.188	.120	.154	.271	.128	.416	.114	.315	.256	.501	.216

Table 29: JF and JV by DV^c

	JF	JV	JF	JV	JF	JV	JF	JV	JF	JV	JF	JV	JF	JV
α	10^{-5}		UT		10^{-4}		10^{-3}		0.01		0.05		0.1	
AA	.018	.177	.021	.177	.022	.173	.036	.145	.058	.126	.093	.110	.114	.104
AXP	.026	.152	.030	.145	.034	.137	.051	.119	.083	.098	.127	.085	.160	.077
BA	.026	.129	.030	.124	.035	.117	.050	.103	.089	.084	.139	.075	.170	.072
CAT	.028	.112	.034	.106	.041	.102	.062	.093	.102	.081	.160	.072	.205	.068
CVX	.057	.092	.068	.087	.074	.084	.101	.075	.149	.065	.218	.056	.260	.053
DD	.021	.145	.026	.135	.029	.128	.040	.116	.066	.101	.099	.089	.124	.083
DIS	.025	.118	.031	.110	.035	.106	.052	.092	.086	.080	.127	.071	.157	.066
GE	.009	.150	.010	.143	.012	.131	.017	.122	.029	.104	.040	.095	.047	.092
GS	.007	.155	.011	.133	.014	.125	.023	.116	.056	.097	.103	.088	.145	.082
HD	.016	.125	.020	.117	.024	.112	.037	.097	.065	.083	.107	.072	.138	.068
IBM	.027	.111	.031	.105	.035	.102	.050	.092	.084	.077	.141	.066	.177	.062
JNJ	.040	.130	.047	.123	.052	.119	.072	.102	.111	.087	.165	.075	.199	.070
JPM	.044	.090	.051	.084	.057	.081	.080	.074	.118	.066	.171	.060	.205	.058
KO	.016	.126	.018	.122	.020	.117	.031	.106	.055	.089	.082	.082	.102	.078
MCD	.024	.139	.028	.132	.032	.127	.049	.114	.085	.095	.133	.084	.169	.079
MMM	.029	.133	.034	.126	.042	.120	.062	.106	.106	.094	.167	.084	.211	.081
MRK	.040	.144	.047	.135	.053	.129	.076	.115	.117	.099	.162	.089	.197	.084
NKE	.028	.112	.033	.105	.038	.097	.062	.081	.101	.070	.156	.062	.189	.061
PFE	.019	.107	.022	.105	.025	.105	.034	.096	.042	.091	.057	.084	.069	.080
PG	.035	.128	.041	.121	.047	.113	.072	.097	.117	.082	.164	.075	.195	.071
TRV	.007	.239	.011	.200	.012	.195	.020	.170	.045	.140	.095	.116	.136	.108
UNH	.023	.120	.027	.117	.032	.114	.050	.099	.085	.088	.143	.078	.179	.075
V	.030	.096	.035	.091	.041	.087	.062	.077	.105	.067	.163	.059	.202	.056
VZ	.023	.160	.025	.152	.027	.148	.035	.129	.053	.110	.073	.094	.087	.089
UTX	.035	.137	.043	.126	.051	.124	.074	.111	.125	.095	.182	.089	.217	.084
WMT	.030	.114	.036	.109	.041	.105	.058	.093	.096	.080	.142	.071	.178	.067
XOM	.072	.095	.083	.089	.093	.084	.125	.075	.178	.066	.230	.060	.267	.057
CSCO	.017	.117	.019	.111	.021	.109	.029	.099	.042	.085	.063	.074	.073	.070
INTC	.037	.119	.048	.103	.053	.098	.072	.084	.101	.075	.128	.071	.147	.070
MSFT	.113	.059	.130	.057	.142	.056	.200	.051	.265	.050	.327	.050	.372	.050
avg.	.031	.128	.036	.120	.041	.115	.059	.102	.094	.088	.139	.078	.170	.074

Table 30: JF and JV by SBV and PABV

	SBV_{300}		SBV_{60}		$PABV_1$			$PABV_2$		
	JF	JV	JF	JV	Δ	JF	JV	Δ	JF	JV
AA	.043	.359	.350	.133	7.4	.024	.181	14.8	.010	.248
AXP	.048	.369	.249	.128	7.8	.027	.173	15.6	.010	.224
BA	.058	.369	.291	.134	7.9	.021	.150	15.7	.009	.210
CAT	.050	.368	.265	.127	7.7	.020	.149	15.4	.008	.221
CVX	.044	.368	.187	.133	7.1	.012	.163	14.3	.004	.228
DD	.043	.359	.249	.125	6.9	.020	.178	13.8	.009	.206
DIS	.046	.371	.247	.130	6.7	.023	.179	13.5	.010	.245
GE	.043	.359	.266	.121	6.8	.027	.143	13.6	.008	.223
GS	.058	.359	.384	.133	11.9	.003	.212	23.8	.001	.219
HD	.051	.367	.252	.127	7.4	.023	.170	14.8	.011	.250
IBM	.043	.369	.240	.129	7.5	.020	.133	15.0	.007	.180
JNJ	.057	.368	.257	.132	6.7	.036	.160	13.4	.013	.215
JPM	.047	.365	.198	.128	5.8	.024	.155	11.6	.009	.208
KO	.041	.366	.260	.122	7.2	.016	.150	14.5	.004	.203
MCD	.047	.367	.298	.130	7.9	.024	.170	15.7	.009	.225
MMM	.049	.365	.311	.132	8.8	.023	.157	17.6	.007	.292
MRK	.053	.364	.288	.131	6.8	.030	.164	13.5	.011	.253
NKE	.056	.352	.256	.126	8.3	.010	.145	16.5	.005	.138
PFE	.049	.361	.252	.123	7.3	.008	.102	14.6	.003	.165
PG	.039	.354	.233	.127	6.7	.018	.152	13.4	.006	.221
TRV	.056	.353	.491	.155	13.1	.011	.176	26.2	.003	.142
UNH	.059	.372	.336	.135	9.9	.015	.153	19.8	.006	.171
V	.053	.379	.271	.133	8.2	.014	.169	16.5	.008	.213
VZ	.061	.368	.255	.133	7.6	.007	.231	15.1	.003	.339
UTX	.041	.382	.281	.133	7.1	.027	.153	14.1	.008	.227
WMT	.046	.376	.233	.130	6.6	.028	.167	13.3	.011	.216
XOM	.039	.361	.175	.131	5.5	.025	.137	11.0	.007	.212
CSCO	.046	.345	.324	.126	9.6	.004	.141	19.3	.002	.137
INTC	.063	.380	.305	.134	8.8	.012	.158	17.6	.004	.279
MSFT	.043	.367	.233	.131	7.1	.011	.130	14.2	.004	.131
avg.	.049	.365	.275	.130	7.8	.019	.160	15.6	.007	.215

Internet Appendix B. Proofs of Main Results

Throughout, C is taken to mean some generic positive constant that may take different values in different places, unless defined otherwise.

Proof of the validity of Assumptions A and C in Example 1.

We first verify Assumption C. Under the additive noise model $Y_{t_j} = X_{t_j} + \delta_n^{\alpha} v_{t_j}$ in (9), we write $\Delta v_j = v_{t_{j+1}} - v_{t_j}$ and $\varepsilon_{t_j} := \delta_n^{\alpha}\Delta v_j$. Then, for all $j \le N_t$ we have

$$\mathbb{E}|\varepsilon_j|^2 = \delta_n^{2\alpha}\mathbb{E}|\Delta v_j|^2 \le C\delta_n^{2\alpha}\mathbb{E}|t_{j+1}-t_j|^{2\beta} \le C\delta_n^{2\alpha}\cdot\delta_n^{4\beta} = C\delta_n^{2(\alpha+2\beta)} \tag{48}$$

by Assumption B and the Hölder continuity. Now, since $\delta_n \le |Y_{t_{j+1}} - Y_{t_j}| = |\Delta X_{t_j} + \varepsilon_{t_j}|$, on the set $\{|\Delta X_{t_j}| \le \delta_n\}$ we obtain $0 \le \delta_n - |\Delta X_j| \le |\varepsilon_{t_j}|$ by the reverse triangle inequality. Therefore, it follows that $0 \le \delta_n^2 - (\Delta X_{t_j})^2 = (\delta_n - |\Delta X_{t_j}|)(\delta_n + |\Delta X_{t_j}|) \le 2\delta_n|\varepsilon_{t_j}| + |\varepsilon_{t_j}|^2$, and hence

$$\begin{aligned}\sum_{j=1}^{N_t}\left\{\delta_n^2 - (\Delta X_j)^2 1_{\{|\Delta X_{t_j}|\le\delta_n\}}\right\} &\le 2\delta_n\sum_{j\le N_t}|\varepsilon_{t_j}| + \sum_{j\le N_t}|\varepsilon_{t_j}|^2\\ &\le \sqrt{N_t}\left(\sum_{j=1}^{N_t}|\varepsilon_j|^2\right)^{1/2} + N_t\cdot C\delta_n^{2(\alpha+2\beta)}\end{aligned} \tag{49}$$

by the Cauchy-Schwarz inequality and the linearity of expectations. Now, since $N_t\delta_n^2$ is bounded in probability (see Proof of Proposition 1 below for formal derivation), the RHS of (49) is $O_p(\delta_n^{\alpha+2\beta-1} + \delta_n^{2(\alpha+2\beta)-2})$, which is $o_p(\delta_n)$ as long as $\alpha + 2\beta > 2$, which yields Assumption C, as desired.

Now we check Assumption A(i). Since X follows (1), for $1 \le j \le N_t$, using the Itô isometry and Markov's inequality we have $\mathbb{E}(|\Delta X_{t_j}|) = O(|t_{j+1}-t_j|^{1/2}) + \lambda\mathbb{E}|U|\mathbb{E}(t_{j+1}-t_j) + o(\mathbb{E}[|t_{j+1}-t_j|])$ where λ is the Poisson intensity, so $|\Delta X_{t_j}| = o_p(\delta_n)$. Now, because $|\Delta X_{t_j}| \le \delta_n$ with probability tending to one, using the triangle inequality we have

$$|a_{n,j} - 1| = \left|\frac{|\Delta X_{t_j} + \varepsilon_{t_j}|}{\delta_n} - 1\right| \le \frac{||\Delta X_{t_j}| - \delta_n|}{\delta_n} + \frac{|\varepsilon_{t_j}|}{\delta_n} \le \frac{2|\varepsilon_{t_j}|}{\delta_n}. \tag{50}$$

Consequently, the Cauchy-Schwarz inequality and (49) give

$$\mathbb{E}\left|\frac{1}{N_t}\sum_{j=1}^{N_t}a_{n,j} - 1\right| \le \frac{2}{\delta_n}\cdot\frac{1}{N_t}\sum_{j=1}^{N_t}\mathbb{E}|\varepsilon_{t_j}| \le \frac{2}{\delta_n}\left(\mathbb{E}|\varepsilon_{t_j}|^2\right)^{1/2} \le C\delta_n^{\alpha+2\beta-1} = o(\delta_n), \tag{51}$$

because $\alpha + 2\beta > 2$. Hence, $N_t^{-1}a_{n,j} = 1 + o_p(\delta_n)$ and $\xi_{1,t} = 1$. Also, by the same argument,

$$\left|\frac{1}{N_t}\sum_{j=1}^{N_t}a_{n,j}^2 - 1\right| \le \frac{C}{N_t}\sum_{j=1}^{N_t}\left(\frac{|\varepsilon_{t_j}|}{\delta_n} + \frac{|\varepsilon_{t_j}|^2}{\delta_n^2}\right) = O_p(\delta_n^{\alpha+2\beta-1} + \delta_n^{2(\alpha+2\beta)-2}) = o(\delta_n), \tag{52}$$

implying $N_t^{-1}a_{n,j}^2 = 1 + o_p(\delta_n)$ and $\xi_{2,t} = 1$. Assumption A(i) is therefore satisfied.

Lastly, we verify Assumption A(ii). From (18) and the above, we have

$$\frac{1}{N_t}\sum_{i=1}^{N_t-1}\sum_{j=i+1}^{N_t}\left(a_{n,i}-a_{n,j}\right)^2 \;=\; \frac{1}{N_t}\left(N_t\sum_{j=1}^{N_t}a_{n,j}^2-\left(\sum_{j=1}^{N_t}a_{n,j}\right)^2\right) \;=\; O_p(\delta_n^{2(\alpha+2\beta)-2}) \tag{53}$$

and since $N_t = O_p(\delta_n^{-2})$ and $\alpha + 2\beta > 2$, we have (53) $= o_p(1)$, and Assumption A(ii) holds. The proof is now complete. □

Proof of the validity of Assumptions A and C in Examples 2 and 3.

We write $Y_{t_j} = X_{t_j} + r_{t_j}$, where the microstructure component r_{t_j} satisfies $|\Delta r_j| := |r_{t_{j+1}} - r_{t_j}| \le \Delta_n$, with $\Delta_n = o(\delta_n^2)$. This setting covers the Roll model (Example 2) and the rounding model (Example 3). Note that $|\Delta X_{t_j}| \ge \delta_n - |\Delta r_{t_j}| \ge \delta_n - \Delta_n$ and $|\Delta X_{t_j}| \le \delta_n + \Delta_n$ by the triangle inequality. Therefore, uniformly over $j \le N_t$, it follows that

$$a_{n,j} = \frac{|\Delta X_j + \Delta r_j|}{\delta_n} = 1 + O_p\left(\frac{\Delta_n}{\delta_n}\right). \tag{54}$$

Averaging this yields $\frac{1}{N_t}\sum a_{n,j}^k \to 1$ for $k = 1, 2$, and we have Assumption A(i) with $\xi_{k,t} = 1$ with the rate $o_p(\delta_n)$ since $\Delta_n/\delta_n = o(\delta_n)$. In the absence of jumps, $N_t^{-1}\sum_{i<j}(a_{n,i} - a_{n,j})^2 \le C(\Delta_n/\delta_n)^2 \to 0$, so A(ii) holds. Lastly, for Assumption C, we note that $|\Delta X_{t_j}| \le \delta_n + \Delta_n$ and hence

$$\sum_{j=1}^{N_t}\left\{\delta_n^2 - (\Delta X_{t_{j+1}})^2 1_{\{|\Delta X_{t_{j+1}}| \le \delta_n\}}\right\} \le 2\delta_n\Delta_n \cdot N_t. \tag{55}$$

Since $N_t\delta_n^2$ is bounded in probability and $\Delta_n = o(\delta_n^2)$, it follows that $2\delta_n\Delta_n \cdot N_t = O_p(\delta_n\Delta_n \cdot \delta_n^{-2}) = O_p(\Delta_n/\delta_n) = o_p(\delta_n)$, which completes the proof. □

Proof of Proposition 1.

Writing $\Delta X_{t_{j+1}} := X_{t_{j+1}} - X_{t_j}$, we have

$$\begin{aligned} DV_t(\delta_n) \;&=\; N_t\delta_n^2 \;=\; \sum_{j=1}^{N_t}\delta_n^2 - \left(\Delta X_{t_{j+1}}\right)^2 1_{\{|\Delta X_{t_{j+1}}| \le \delta_n\}} + \left(\Delta X_{t_{j+1}}\right)^2 1_{\{|\Delta X_{t_{j+1}}| \le \delta_n\}} \\ &=\; \sum_{j=1}^{N_t}\left(\Delta X_{t_{j+1}}\right)^2 1_{\{|\Delta X_{t_{j+1}}| \le \delta_n\}} + \sum_{j=1}^{N_t}\left\{\delta_n^2 - \left(\Delta X_{t_{j+1}}\right)^2 1_{\{|\Delta X_{t_{j+1}}| \le \delta_n\}}\right\} \\ &=\; \sum_{j=1}^{N_t}\left(X_{t_{j+1}} - X_{t_j}\right)^2 1_{\{|X_{t_{j+1}} - X_{t_j}| \le \delta_n\}} + R_{n,t}. \end{aligned} \tag{56}$$

By Assumption C, $R_{n,t} = o_p(\delta_n)$. Therefore, for the first result, it suffices to show that the first term in (56) converges in probability to $\int_0^t \sigma_s^2 ds$. Fix $m \ge 1$ and define the stopping time

$$\tau_m := \Big(\inf\{u \in [0,t] : |\mu_u| > m \text{ or } |\sigma_u| > m\}\Big) \wedge t. \tag{57}$$

Since μ is locally bounded and σ is càdlàg, and hence locally bounded, we have $\tau_m \uparrow t$ a.s. as $m \to \infty$ and thus $\mathbb{P}(\tau_m < t) \to 0$. Therefore, we first show the consistency on the event $\{\tau_m = t\}$ for each fixed m.

Let X' denote the continuous Itô semimartingale part of X, i.e. $X'_u := X_u - \sum_{s \le u} \Delta X_s$. Let K_t be the set of indices $j \le N_t$ for which a jump occurs over $(t_j, t_{j+1}]$. Since there are at most finitely many jumps on $[0,t]$, we have $|K_t| < \infty$ a.s, where $|\mathcal{C}|$ is taken to mean the cardinality of the set $\mathcal{C}$. Thus, we have

$$\begin{aligned}\sum_{j=1}^{N_t} \left(\Delta X_{t_{j+1}}\right)^2 1_{\{|\Delta X_{t_{j+1}}| \le \delta_n\}} &= \sum_{j \notin K_t} \left(\Delta X'_{t_{j+1}}\right)^2 1_{\{|\Delta X'_{t_{j+1}}| \le \delta_n\}} + \sum_{j \in K_t} \left(\Delta X_{t_{j+1}}\right)^2 1_{\{|\Delta X_{t_{j+1}}| \le \delta_n\}} \\ &= \sum_{j=1}^{N_t} \left(\Delta X'_{t_{j+1}}\right)^2 - A_{n,t} - B_{n,t} + O_p(\delta_n^2), \end{aligned} \tag{58}$$

where

$$A_{n,t} := \sum_{j \in K_t} (\Delta X'_{t_{j+1}})^2, \qquad B_{n,t} := \sum_{j \notin K_t} (\Delta X'_{t_{j+1}})^2 1_{\{|\Delta X'_{t_{j+1}}| > \delta_n\}}.$$

We first prove that $A_{n,t} = o_p(1)$. By Assumption B, the mesh $\max_{j \le N_t}(t_{j+1} - t_j) = O_p(\delta_n^2) = o_p(1)$. Since X' has continuous paths, it follows that $\max_{j \le N_t} |\Delta X'_{t_{j+1}}| \xrightarrow{p} 0$. Because $|K_t| < \infty$ a.s., we obtain

$$0 \le A_{n,t} \le |K_t| \cdot \max_{j \le N_t} (\Delta X'_{t_{j+1}})^2 \xrightarrow{p} 0. \tag{59}$$

Next, we show that $B_{n,t} = o_p(1)$. Define $J_{n,t} := \{j \le N_t : |\Delta X_{t_{j+1}}| > \delta_n\}$. Note that each summand in $R_{n,t}$ is nonnegative, and if $j \in J_{n,t}$ then $\delta_n^2 - (\Delta X_{t_{j+1}})^2 1_{\{|\Delta X_{t_{j+1}}| \le \delta_n\}} = \delta_n^2$. Therefore,

$$\delta_n^2 |J_{n,t}| = \sum_{j=1}^{N_t} \delta_n^2 1_{\{|\Delta X_{t_{j+1}}| > \delta_n\}} \le R_{n,t}, \tag{60}$$

and therefore, $|J_{n,t}| = o_p(\delta_n^{-1})$, by Assumption C.

Now, fix any $\eta \in (0, 1/2)$ and choose an even integer q such that $q\eta > 2$. On $\{\tau_m = t\}$ we have $\sup_{s \le t}(|\mu_s| \vee |\sigma_s|) \le m$. Fix $M > 0$ and define the event $\mathcal{B}_{n,M} := \{N_t \delta_n^2 \le M, \max_{j \le N_t}(t_{j+1} - t_j) \le M\delta_n^2\}$. By Minkowski's inequality and the Burkholder-Davis-Gundy inequality, on $\mathcal{B}_{n,M} \cap \{\tau_m = t\}$, there exists a constant $C_{q,m}$ such that, uniformly in $j \le N_t$,

$$\mathbb{E}|\Delta X'_{t_{j+1}}|^q \le C_{q,m}\Big(|t_{j+1} - t_j|^q + |t_{j+1} - t_j|^{q/2}\Big). \tag{61}$$

Meanwhile, write $X = B + M + J$ with $B_u = \int_0^u \mu_s ds$ and $M_u = \int_0^u \sigma_s dW_s$. Since $M^2 - \langle M \rangle$ is a martingale, the optional sampling theorem yields

$$\mathbb{E}\left[\sum_{j=1}^{N_t} (\Delta M_{t_{j+1} \wedge t})^2 1_{\{\tau_m = t\}}\right] = \mathbb{E}[\langle M \rangle_t 1_{\{\tau_m = t\}}] \le m^2 t,$$

while the drift contribution is $O_p(\delta_n^2)$ by Assumption B and J has finite activity. Therefore, $\sum_{j=1}^{N_t}(\Delta X_{t_{j+1}})^2 = O_p(1)$ on $\{\tau_m = t\}$, and $N_t\delta_n^2 = DV_t(\delta_n) \leq \sum_{j=1}^{N_t}(\Delta X_{t_{j+1}})^2 + R_{n,t} = O_p(1)$ since $R_{n,t} = o_p(1)$. Now, since $\max_{j\leq N_t}(t_{j+1} - t_j) = O_p(\delta_n^2)$ by Assumption B and $N_t\delta_n^2 = O_p(1)$ on $\{\tau_m = t\}$, for any $\varepsilon > 0$ we can choose M large enough such that $\limsup_{n\to\infty} \mathbb{P}(\mathcal{B}_{n,M}^c \cap \{\tau_m = t\}) \leq \varepsilon$.

Consequently, by (61) and Markov's inequality, and because $N_t \leq M\delta_n^{-2}$ on $\mathcal{B}_{n,M}$, we have

$$\begin{aligned}\mathbb{P}\Big(\max_{j\leq N_t}|\Delta X'_{t_{j+1}}| > \delta_n^{1-\eta}, \mathcal{B}_{n,M}, \tau_m = t\Big) &\leq M\delta_n^{-2}\delta_n^{-q(1-\eta)}\, C_{q,m}\Big((M\delta_n^2)^q + (M\delta_n^2)^{q/2}\Big) \\ &\leq M\delta_n^{-2}\cdot C'_{q,m}M^{q/2}\delta_n^{q\eta} = C'_{q,m}M^{1+q/2}\delta_n^{q\eta-2}, \qquad (62)\end{aligned}$$

for some constant $C'_{q,m}$, and the bound in (62) approaches zero because $q\eta > 2$, yielding that on $\{\tau_m = t\}$, we have the following:

$$\max_{j\leq N_t}|\Delta X'_{t_{j+1}}| = O_p(\delta_n^{1-\eta}). \qquad (63)$$

Finally, on $\{j \notin K_t\}$ we have $\Delta X_{t_{j+1}} = \Delta X'_{t_{j+1}}$, so $\{|\Delta X'_{t_{j+1}}| > \delta_n\} \subseteq J_{n,t}$ on K_t^c. Using (60) and (63), we have $0 \leq B_{n,t} \leq (\max_{j\leq N_t}|\Delta X'_{t_{j+1}}|^2)\cdot|J_{n,t}| = O_p(\delta_n^{2-2\eta})\cdot o_p(\delta_n^{-1}) = o_p(1)$, since $\eta < 1/2$.

Plugging $A_{n,t} = o_p(1)$ and $B_{n,t} = o_p(1)$ into (58) gives

$$\sum_{j=1}^{N_t}\Big(\Delta X_{t_{j+1}}\Big)^2 1_{\{|\Delta X_{t_{j+1}}|\leq\delta_n\}} = \sum_{j=1}^{N_t}\Big(\Delta X'_{t_{j+1}}\Big)^2 + o_p(1). \qquad (64)$$

Since the mesh of the partition $\{t_j\}$ converges to 0 in probability by Assumption B, Jacod and Shiryaev (2003, Theorem 1.4.47) yield

$$\sum_{j=1}^{N_t}\Big(\Delta X'_{t_{j+1}}\Big)^2 \xrightarrow{p} \langle X', X'\rangle_t = \int_0^t \sigma_s^2\, ds. \qquad (65)$$

Since $R_{n,t} = o_p(1)$, letting $m \to \infty$ and using $\mathbb{P}(\tau_m < t) \to 0$ yield the desired result (17). Further, taking the difference between DV^C and DV, we have

$$\begin{aligned}DV_t^C(\delta_n) - DV_t(\delta_n) &= \frac{1}{N_t}\left(\sum_{j=1}^{N_t}\frac{|Y_{t_{j+1}} - Y_{t_j}|}{\delta_n}\right)^2\delta_n^2 - N_t\delta_n^2 \\ &= \left[\left(\frac{1}{N_t}\sum_{j=1}^{N_t}\frac{|Y_{t_{j+1}} - Y_{t_j}|}{\delta_n}\right)^2 - 1\right]\cdot DV_t(\delta_n) \xrightarrow{p} (\xi_{1,t}^2 - 1)\int_0^t \sigma_s^2 ds, \quad (66)\end{aligned}$$

in view of Assumption A(i) and (17), and hence (18) follows. Equation (19) is an immediate consequence of Theorem 1 below. □

Proof of Theorem 1.

For the first part of the theorem concerning DV_t^C, we note that

$$DV_t^C(\delta_n) = \left[\frac{1}{N_t}\sum_{j=1}^{N_t}\left(\frac{|Y_{t_{j+1}} - Y_{t_j}|}{\delta_n}\right)\right]^2 N_t\delta_n^2 = A_{n,t}^2 \cdot DV_t(\delta_n). \tag{67}$$

Therefore, since $\delta_n^{-1}(A_{n,t}^2 - \xi_{1,t}^2) = o_p(1)$ by Assumption A(i), we have, by Hong, Nolte, Taylor, and Zhao (2023) that

$$\delta_n^{-1}\Big(DV_t^C(\delta_n) - \xi_{1,t}^2\int_0^t \sigma_s^2 ds\Big) = \delta_n^{-1}\xi_{1,t}^2\Big(DV_t(\delta_n) - \int_0^t \sigma_s^2 ds\Big) + o_p(1) \longrightarrow \mathcal{W}_{1,t}, \tag{68}$$

$\mathcal{F}$-stably in law, where $\mathcal{W}_{1,t}$ is a continuous centered Gaussian martingale with $\mathbb{E}(\mathcal{W}_{1,t}^2|\mathcal{F}) = \frac{2}{3}\cdot\xi_{1,t}^4\int_0^t \sigma_s^2 ds$.

We now turn to the second component involving $RV_t(\delta_n) = \sum_{j=1}^{N_t}(Y_{t_{j+1}} - Y_{t_j})^2$. We also introduce the "noise-free" realized variance on the same endogenous grid,

$$RV_t^X(\delta_n) := \sum_{j=1}^{N_t}\left(X_{t_{j+1}} - X_{t_j}\right)^2. \tag{69}$$

Consider additionally the exit-time scheme from a regular grid studied by Vetter and Zwingmann (2017), which can be thought of as a jump-adjusted version of the regular grid sampling of size δ_n in Fukasawa and Rosenbaum (2012):

$$t_{n,j+1}^* := \inf\{u > t_{n,j}^* :\ X_u \notin \mathcal{G}_{X_{t_{n,j}^*}}\}, \qquad j \geq 0, \tag{70}$$

where

$$\mathcal{G}_x = \begin{cases} [x-\delta_n, x+\delta_n] & \text{if } \ x\delta_n^{-1} \in \mathbb{N} \\ \big[\lfloor\delta_n^{-1}x\rfloor\delta_n, (\lfloor\delta_n^{-1}x\rfloor + 1)\delta_n\big] & \text{if } \ x\delta_n^{-1} \notin \mathbb{N} \end{cases} \tag{71}$$

with the associated realized variance

$$RV_t^{X,*}(\delta_n) := \sum_{j=1}^{N_t^*}\left(X_{t_{j+1}^*} - X_{t_j^*}\right)^2. \tag{72}$$

By Theorem 3.1 of Vetter and Zwingmann (2017) we have, $\mathcal{F}$-stably in law, $\delta_n^{-1}(RV_t^{X,*}(\delta_n) - [X,X]_t) \to \widetilde{\mathcal{W}}_{2,t} + \widetilde{\mathcal{Y}}_t$, where $\widetilde{\mathcal{W}}_{2,t}$ is a continuous centered Gaussian martingale with $\mathbb{E}(\widetilde{\mathcal{W}}_{2,t}^2|\mathcal{F}) = \frac{2}{3}\int_0^t \sigma_s^2 ds$, and $\widetilde{\mathcal{Y}}_t = \sum_{T_p\leq t} 2\eta_p U_{T_p}$, with $(\eta_p)_{p\geq 1}$ i.i.d. having the density given in Eq. (3.2) of Vetter and Zwingmann (2017).

Since J has finite activity, $\Omega_n := \{\inf_{T_p\leq t}|\Delta X_{T_p}| > 2\delta_n\}$ satisfies $\mathbb{P}(\Omega_n) \to 1$. On Ω_n, $\{t_{n,j}\}$ and $\{t_{n,j}^*\}$ can differ only on finitely many intervals, and with Assumption B, the discrepancy in realized variance is $O_p(\delta_n^2)$. Consequently, we have $\delta_n^{-1}\Big(RV_t^X(\delta_n) - RV_t^{X,*}(\delta_n)\Big) = o_p(1)$, and hence the limit law continues to hold with $RV_t^X(\delta_n)$ in place of $RV_t^{X,*}(\delta_n)$.

Finally, by Assumptions A and C, the microstructure component perturbs the realized variance only through the scale factor $\xi_{2,t}$ at the δ_n-rate, so that $\delta_n^{-1}(RV_t(\delta_n) - \xi_{2,t}RV_t^X(\delta_n)) = o_p(1)$. Combining this with the stable limit for $RV_t^X(\delta_n)$ yields

$$\delta_n^{-1}\Big(RV_t(\delta_n) - \xi_{2,t}[X,X]_t\Big) \implies \mathcal{W}_{2,t} + \mathcal{Y}_t, \tag{73}$$

$\mathcal{F}$-stably in law, where $\mathcal{W}_{2,t} := \xi_{2,t}\widetilde{\mathcal{W}}_{2,t}$ is a continuous centered Gaussian martingale satisfying $\mathbb{E}(\mathcal{W}_{2,t}^2|\mathcal{F}) = \frac{2}{3}\xi_{2,t}^2\int_0^t \sigma_s^2 ds$, and $\mathcal{Y}_t := \xi_{2,t}\widetilde{\mathcal{Y}}_t = \xi_{2,t}\sum_{T_p\leq t} 2\eta_p U_{T_p}$. The continuous Gaussian martingales $\mathcal{W}$ and the jump-induced term $\mathcal{Y}$ are pairwise uncorrelated as we show. $\mathcal{W}_{1,t}$ and $\mathcal{W}_{2,t}$ are orthogonal, conditional on $\mathcal{F}$, and this follows from the same argument as in the proof of Theorem 2 below; the finitely many intervals containing jumps do not contribute to the covariance of the continuous Gaussian limit. What remains therefore is to clarify the dependence between $\mathcal{W}_{i,t}$ and $\mathcal{Y}_t$ for $i = 1, 2$.

Let $\mathcal{H}_\eta := \sigma((\eta_p)_{p\geq 1})$. We work on the stable extension, writing $\widetilde{\mathbb{E}}$ and $\widetilde{\mathrm{Cov}}$ for expectation and covariance there. After localization, assume square integrability. The auxiliary Brownian motion(s) driving $(\mathcal{W}_1, \mathcal{W}_2)$ and the variables $(\eta_p)_{p\geq 1}$ are chosen on independent auxiliary coordinates, independent of $\mathcal{F}$. Now, for $i = 1, 2$, we have

$$\widetilde{\mathbb{E}}(\mathcal{W}_{i,t}\,|\,\mathcal{F}\vee\mathcal{H}_\eta) \;=\; \widetilde{\mathbb{E}}(\mathcal{W}_{i,t}|\mathcal{F}) \;=\; 0 \tag{74}$$

since $\mathcal{W}_i$ is conditionally centered Gaussian. Moreover, given that $\xi_{2,t}$, U_{T_p} and $1\{T_p \leq t\}$ are $\mathcal{F}$-measurable, the sum $\mathcal{Y}_t = \xi_{2,t}\sum_{T_p\leq t} 2\eta_p U_{T_p}$ is finite and $\mathcal{F}\vee\mathcal{H}_\eta$-measurable. It now follows that

$$\begin{aligned}\widetilde{\mathbb{E}}(\mathcal{W}_{i,t}\mathcal{Y}_t|\mathcal{F}) =&\; \widetilde{\mathbb{E}}[\widetilde{\mathbb{E}}(\mathcal{W}_{i,t}\mathcal{Y}_t|\mathcal{F}\vee\mathcal{H}_\eta)|\mathcal{F}]\\ =&\; \widetilde{\mathbb{E}}[\mathcal{Y}_t\widetilde{\mathbb{E}}(\mathcal{W}_{i,t}|\mathcal{F}\vee\mathcal{H}_\eta)|\mathcal{F}] \;=\; 0\end{aligned} \tag{75}$$

by the tower property, and $\widetilde{\mathrm{Cov}}(\mathcal{W}_{i,t}, \mathcal{Y}_t|\mathcal{F}) = 0$ for $i = 1, 2$ by (74) and (75). Hence, taking expectations gives, whenever the unconditional covariance is well-defined, $\widetilde{\mathrm{Cov}}(\mathcal{W}_{i,t}, \mathcal{Y}_t) = 0$.

The joint stable convergence of the two-dimensional vector follows by applying the stable central limit theorem to arbitrary linear combinations and then using the Cramér-Wold device for stable convergence, as in the proof of Theorem 2 below. The proof of Theorem 1 is now complete. □

Proof of Theorem 2.

Under the null of no jump, we have

$$
\begin{aligned}
RV_t(\delta_n) &= \sum_{j=1}^{N_t}\Big(Y_{t_{j+1}}-Y_{t_j}\Big)^2\\
&= \left[\frac{1}{N_t}\sum_{j=1}^{N_t}\left(\frac{|Y_{t_{j+1}}-Y_{t_j}|}{\delta_n}\right)^2\right]\cdot\sum_{j=1}^{N_t}\Big(X_{t_{j+1}}-X_{t_j}\Big)^2 \;+\; o_p(\delta_n)\\
&= B_{n,t}\cdot\sum_{j=1}^{N_t}\Big(X_{t_{j+1}}-X_{t_j}\Big)^2 \;+\; o_p(\delta_n), \qquad (76)
\end{aligned}
$$

since $\delta_n^{-1}B_{n,t}(DV_t - S^X_{n,t}) = o_p(1)$. In the meantime, we recall that $A_{n,t} = \widehat{\xi}_{1,t}$ and $B_{n,t} = \widehat{\xi}_{2,t}$ consistently estimate ξ_1 and ξ_2, respectively, by Assumption A(i). Now write

$$
M^{(1)}_{n,t} := \delta_n^{-1}\left(DV_t - IV_t\right) \quad \text{and} \quad M^{(2)}_{n,t} := \delta_n^{-1}\sum_{j=1}^{N_t}\Big((\Delta X_j)^2 - V_j\Big), \qquad (77)
$$

where $IV_t = \int_0^t \sigma_s^2 ds$, $\Delta X_j = X_{t_{j+1}} - X_{t_j}$, and $V_j := \int_{t_j}^{t_{j+1}}\sigma_s^2 ds$. Then, it follows that

$$
M^{*,(1)}_{n,t} = \delta_n^{-1}\Big(DV_t^C - \xi^2_{1,t}\int_0^t \sigma_s^2 ds\Big) = \xi^2_{1,t}M^{(1)}_{n,t} + r^{(1)}_{n,t} \qquad (78)
$$

$$
M^{*,(2)}_{n,t} = \delta_n^{-1}\Big(RV_t - \xi_{2,t}\int_0^t \sigma_s^2 ds\Big) = \xi_{2,t}M^{(2)}_{n,t} + r^{(2)}_{n,t} \qquad (79)
$$

where the remainders $r^{(1)}_{n,t}$ and $r^{(2)}_{n,t}$ are $o_p(1)$ under Assumption A(i), cf. (67) and (76). In fact, with Assumption C we can easily show that they both converge to zero also in u.c.p. sense, Aït-Sahalia and Jacod (2014). Now, noting that X is continuous on Ω_t^C and that $(X_u^2 - \langle X\rangle_u)$ is a martingale, by the tower property and the optional sampling theorem at the stopping times $t_j < t_{j+1}$, we have

$$
\begin{aligned}
\Big\langle M^{(1)}_n, M^{(2)}_n\Big\rangle_t &= \sum_{j=1}^{N_t}\mathbb{E}\left[\Big(\delta_n^{-1}(\delta_n^2 - V_j)\Big)\Big(\delta_n^{-1}\Big((\Delta X_j)^2 - V_j\Big)\Big)|\mathcal{F}_{t_j}\right].\\
&= \delta_n^{-2}\sum_{j=1}^{N_t}\Big\{\mathbb{E}\left[\delta_n^2\Big((\Delta X_j)^2 - V_j\Big)|\mathcal{F}_{t_j}\right] - \mathbb{E}\left[V_j\Big((\Delta X_j)^2 - V_j\Big)|\mathcal{F}_{t_j}\right]\Big\}\\
&= \delta_n^{-2}\sum_{j=1}^{N_t}\Big\{\mathbb{E}\left[\delta_n^2\mathbb{E}\left[(\Delta X_j)^2 - V_j|\mathcal{F}_{t_j}\vee\sigma(V_j)\right]\Big|\mathcal{F}_{t_j}\right] - \mathbb{E}\left[V_j\cdot\Big(\mathbb{E}[(\Delta X_j)^2|\mathcal{F}_{t_j}\vee\sigma(V_j)]\Big)\Big|\mathcal{F}_{t_j}\right]\Big\}\\
&= \delta_n^{-2}\sum_{j=1}^{N_t}\mathbb{E}\Big[\delta_n^2\Big\{\mathbb{E}\Big[\Big(\int_{t_j}^{t_{j+1}}\sigma_s dW_s\Big)^2|\mathcal{F}_{t_j}\vee\sigma(V_j)\Big] - V_j\Big\}\Big|\mathcal{F}_{t_j}\Big]\\
&\quad - \delta_n^{-2}\sum_{j=1}^{N_t}\mathbb{E}\Big[V_j\Big\{\mathbb{E}\Big[\Big(\int_{t_j}^{t_{j+1}}\sigma_s dW_s\Big)^2|\mathcal{F}_{t_j}\vee\sigma(V_j)\Big] - V_j\Big\}\Big|\mathcal{F}_{t_j}\Big] \;=\; 0+0=0, \qquad (80)
\end{aligned}
$$

where the last line is due to the Itô isometry.

Furthermore, since $r_{n,t}^{(1)} = o_p(1) = r_{n,t}^{(2)}$ and $M_{n,t}^{(1)}$ and $M_{n,t}^{(2)}$ are bounded in probability, by the Cauchy-Schwartz inequality, $\langle r_n^{(1)}, M_n^{(2)}\rangle_t$, $\langle M_n^{(1)}, r_n^{(2)}\rangle_t$, and $\langle r_n^{(1)}, r_n^{(2)}\rangle_t$ all converge in probability to zero. Since $\langle M_{n,t}^{*,(1)}, M_{n,t}^{*,(2)}\rangle_t \to^p 0$, for any $a, b \in \mathbb{R}$ we have:

$$\langle aM_n^{*,(1)} + bM_n^{*,(2)}\rangle_t \;=\; a^2\langle M_n^{*,(1)}\rangle_t + b^2\langle M_n^{*,(2)}\rangle_t + o_p(1). \tag{81}$$

Therefore, applying the same stable central limit theorem as in Theorem 1 to the martingale $aM_{n,t}^{*,(1)} + bM_{n,t}^{*,(2)}$ yields its stable convergence to $a\mathcal{W}_{1,t} + b\mathcal{W}_{2,t}$ under the null of no jump. Consequently, by the Cramér-Wold device for stable convergence, see Corollary 3.8 of Häusler and Luschgy (2015), $(M_{n,t}^{*,(1)}, M_{n,t}^{*,(2)})$ jointly converges to $(\mathcal{W}_{1,t}, \mathcal{W}_{2,t})$.

Therefore, in restriction to the set $\Omega_t^C = \{\omega : t \mapsto X_t(\omega) \text{ is continuous over } [0,t]\}$, we have

$$\begin{aligned}\delta_n^{-1}\left(1 - \frac{DV_t^C(\delta_n)}{RV_t(\delta_n)}\right) \;&=\; \delta_n^{-1}\left(\frac{RV_t(\delta_n) - DV_t^C(\delta_n)}{RV_t(\delta_n)}\right)\\ &\to\; \frac{\mathcal{W}_{2,t} - \mathcal{W}_{1,t}}{\xi_{2,t}\int_0^t \sigma_s^2 ds} \;=\; \frac{\widetilde{\mathcal{W}}_t}{\xi_{2,t}\int_0^t \sigma_s^2 ds},\end{aligned} \tag{82}$$

where

$$\mathbb{E}(\widetilde{\mathcal{W}}_t^2|\mathcal{F}) \equiv \; \frac{2}{3}\left(\xi_{1,t}^4 + \xi_{2,t}^2\right)\int_0^t \sigma_s^2 ds. \tag{83}$$

To obtain a feasible statistic, we establish the consistency theory for the integrated variance (83) with respect to the endogenous sampling scheme we adopt. Theorem 3.1 of Fukasawa (2010), Theorem 2.3 of Fukasawa and Rosenbaum (2012) along with Assumption C suggest that

$$\mathcal{D}_n = \delta_n^{2-4}\sum_{j=1}^{N_t}\delta_{n,j}^4 \;\xrightarrow{p}\; \int_0^t |\varphi'(\varphi^{-1}(X_t))|^{4-2}\sigma_s^2 ds, \tag{84}$$

where $\varphi(\cdot)$ is the identity function in our sampling scheme.

Therefore, we can construct a feasible CLT for the standardized statistic, which is asymptotically standard normal distributed conditional on the set A, where $A \in \mathcal{F}, A \subset \Omega_t^C, \mathbb{P}(A) > 0$. Noting that RV consistently estimates the denominator of (82) we have

$$Z_{n,t} \;=\; \frac{\delta_n^{-1}\left(1 - \frac{DV_t^C}{RV_t}\right)}{\sqrt{\frac{\frac{2}{3}(\widehat{\xi}_{2,t}^2 + \widehat{\xi}_{1,t}^4)\mathcal{D}_n(\delta_n)}{RV_t^2}}} \;=\; \frac{1 - \frac{DV_t^C}{RV_t}}{\sqrt{\frac{\frac{2}{3}(\widehat{\xi}_{2,t}^2 + \widehat{\xi}_{1,t}^4)\sum_{j=1}^{N_t}\delta_{n,j}^4}{RV_t^2}}} \;\longrightarrow\; N(0,1) \tag{85}$$

$\mathcal{F}$-stably, as required. The proof of the result on power follows the same argument as the proof of Theorem 10.2 in Aït-Sahalia and Jacod (2014), and is therefore omitted for brevity. □

References

Aït-Sahalia, Y. and J. Jacod (2014). *High-Frequency Financial Econometrics*. Princeton University Press.

Barndorff-Nielsen, O. E., P. Hansen, A. Lunde, and N. Shephard (2009). Realized kernels in practice: Trades and quotes. *Econometrics Journal 12*, 1–32.

Fukasawa, M. (2010). Central limit theorem for the realized volatility based on tick time sampling. *Finance and Stochastics 14*(2), 209–233.

Fukasawa, M. and M. Rosenbaum (2012). Central limit theorems for realized volatility under hitting times of an irregular grid. *Stochastic Processes and Their Applications 122*(12), 3901–3920.

Häusler, E. and H. Luschgy (2015). *Stable Convergence and Stable Limit theorems*. Springer.

Hong, S. Y., I. Nolte, S. Taylor, and X. Zhao (2023). Volatility estimation and forecasts based on price durations. *Journal of Financial Econometrics 21*(1), 106–144.

Nolte, I. (2008). Modeling a multivariate transaction process. *Journal of Financial Econometrics 6*, 143–170.

Vetter, M. and T. Zwingmann (2017). A note on central limit theorems for quadratic variation in case of endogenous observation times. *Electronic Journal of Statistics 11*(1), 963–980.