ARTICLE IN PRESS

Journal of Econometrics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Econometrics

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

Volatility measurement with pockets of extreme return persistence☆

Torben G. Andersen [a,*], Yingying Li [b], Viktor Todorov [a], Bo Zhou [c]

[a] Kellogg School, Northwestern University, United States of America
[b] Hong Kong University of Science and Technology, Hong Kong
[c] Durham University Business School, United Kingdom

ARTICLE INFO

Article history:
Received 24 February 2020
Received in revised form 30 August 2020
Accepted 26 November 2020
Available online xxxx

JEL classification:
C12
C53
G10
G17

Keywords:
Extreme return persistence
High-frequency data
Integrated volatility estimation
Market microstructure noise
Volatility forecasting

ABSTRACT

Increasing evidence points towards the episodic emergence of pockets with extreme return persistence. This notion refers to intraday periods of non-trivial duration, for which stock returns are highly positively autocorrelated. Such episodes include, but are not limited to, gradual jumps and prolonged bursts in the drift component. In this paper, we develop a family of integrated volatility estimators, labeled differenced-return volatility (DV) estimators, which provide robustness to these types of Itô semimartingale violations. Specifically, we show that, by using differences in consecutive high-frequency returns, our DV estimators can reduce the non-trivial bias that all commonly-used estimators exhibit during such periods of apparent short-term intraday return predictability. A Monte Carlo study demonstrates the reliability of the newly developed volatility estimators in finite samples. In our empirical volatility forecasting application to S&P 500 index futures and individual equities, our DV-based Heterogeneous Autoregressive (HAR) model performs well relative to existing procedures according to standard out-of-sample MSE and QLIKE criteria.

1. Introduction

Following Zhou (1996), Andersen and Bollerslev (1998), Barndorff-Nielsen and Shephard (2002) and Andersen et al. (2003), the literature is ripe with theoretical and empirical evidence demonstrating that the use of high-frequency data enhances the precision in measuring return volatility and, consequently, can be employed to improve volatility forecasting. These findings are consistent with the notion of arbitrage-free markets, implying that the return process must belong to the class of semimartingales. In this setting, the quadratic return variation constitutes the natural measure of return volatility and, absent market frictions, it may be estimated consistently through the realized volatility (RV) estimator as the intraday sampling frequency increases indefinitely. This has inspired a large literature exploring the best approach for constructing volatility measures from intraday return observations. The main contributions revolve around two separate issues. One, there is a trade-off in efficiency, motivating the use of *all* price or quote observations, versus biases induced by market microstructure noise (MMN) effects at the highest frequencies, which arise from frictions inducing noise terms

☆ The research of Andersen and Todorov is partially supported by National Science Foundation grant SES-1530748. Li acknowledges support from the Research Grants Council, Hong Kong SAR GRF16503419 and T31-604/18-N, as well as the National Natural Science Foundation of China grant NSFC19BM03. We thank Serena Ng (the Editor), the Guest CoEditor, anonymous referees as well as participants at the CFE-CMStatistics Conference at the University of London, December 2019 and the 2020 Econometric Society World Congress, for helpful comments and suggestions.

* Corresponding author.
E-mail address: t-andersen@kellogg.northwestern.edu (T.G. Andersen).

https://doi.org/10.1016/j.jeconom.2020.11.005
0304-4076/

that are "large" relative to the instantaneous innovations to the fundamental price. Two, both in terms of economic interpretation and predictive prowess, it is useful to separate the quadratic variation into a continuous term, reflecting the integrated volatility (IV) associated with the diffusive (continuous) volatility component, and the additional contribution stemming from a (cumulative squared) jump component, generated by discontinuities in the asset price.

Subsequently, both the volatility measurement and forecasting literatures have explored procedures involving jump- and noise-robust estimators, either in isolation or in combination. The MMN-robust estimators include deliberate sampling at somewhat sparse frequencies (Andersen and Bollerslev, 1998), first-order autoregressive correction (Zhou, 1996), sparse subsampling, averaging and bias-correcting (Aït-Sahalia et al., 2011), kernel-based methods (Barndorff-Nielsen et al., 2008), a pre-averaging approach (Jacod et al., 2009), and the quasi-maximum likelihood based procedures of Xiu (2010). Meanwhile, the jump extraction procedures include the bipower and multipower estimators (Barndorff-Nielsen and Shephard, 2004; Barndorff-Nielsen et al., 2006), threshold estimators (Mancini, 2009; Andersen et al., 2012) and combinations thereof (Corsi et al., 2010), swap variation based estimators (Jiang and Oomen, 2008), as well as ones based on the empirical characteristic function (Jacod and Todorov, 2014, 2018).

One notable development in noise-robust volatility estimation is the realization that an i.i.d. assumption on the MMN is too strict, as argued originally in Hansen and Lunde (2006). Recent work by Jacod et al. (2017) confirms the point by providing inference tools for the statistical properties of MMN and documenting significant (positive) autocorrelations of the noise term; see also Li et al. (2016, 2018), Jacod et al. (2019), Li and Linton (2019) and Da and Xiu (2019). The heterogeneous nature of the noise term is not surprising, as the microstructure component reflects many distinct types of frictions, some of which likely display substantial persistence and volatility. Even so, this MMN robust volatility literature is concerned strictly with weak noise dependence. The associated infill asymptotic scheme implies that the correlation structure is local and shrinks in terms of calendar time, as the number of intraday observations diverges. As a result, local averaging of the price suffices to annihilate the impact of noise in the volatility measurement despite the presence of (weak) noise dependence.

However, some recent studies stress the potential limitations of this standard Itô semimartingale plus locally dependent noise framework. This is because of episodes of strong trends in the high-frequency return series, ranging from several minutes to multiple hours. Although such occurrences have been identified previously, notably in the form of the "gradual jumps" explored by Barndorff-Nielsen et al. (2009), they were viewed as exceptions to be eliminated through suitably designed prefiltering procedures. The issue was further highlighted by the May 2010 "flash crash" in the S&P 500 e-mini futures market analyzed by Kirilenko et al. (2017), while the more general phenomenon of "drift burst" was studied systematically in the volatility measurement setting by Christensen et al. (2020). Finally, the presence of relatively frequent and randomly occurring instances of persistent return serial correlation is identified in Andersen et al. (2020). Empirically, such episodes induce a prolonged directional price movement, often in excess of half an hour, which is hard to rationalize through the standard framework, as they are at odds with the underlying Itô semimartingale assumption.

As an illustration of a pronounced low-frequency trend, Fig. 1 depicts the S&P 500 futures for Tuesday, September 1, 1998. The strong downward drift over 8:50–9:20, followed by an equally pronounced rally over 9:20–9:35, is striking. This distinct V-shape trajectory is reminiscent of the infamous "flash crash" pattern, debated in the literature on high-frequency trading and potential market fragility, with Christensen et al. (2020) documenting numerous instances of such price trajectories. In the present case, the volatile market conditions follow a large drop in equity prices on the preceding day. A plausible explanation for this market behavior is a disagreement among bulls and bears regarding the future course of the market, generating substantial uncertainty and distinct periods of strong net selling or buying. Ultimately, the bulls carried the day, resulting in a healthy daily gain that partially reversed the losses suffered on Monday.

In Fig. 2, we plot the S&P 500 index futures price for March 22, 2005, and the subsequent trading day. In the afternoon of March 22, the price experienced a rapid descent from a high point of 1195 to a low of 1175, comprised of many small downward moves, but no apparent jumps. This extreme, yet smooth, transition to a new price level over a short period of time seems to involve a degree of short-term predictability that is at odds with the basic no-arbitrage condition. As demonstrated by the second panel, the price level fluctuates around this lower level throughout the next day, March 23. The type of market move observed on March 22 is labeled a "gradual jump" by Barndorff-Nielsen et al. (2009). The impetus was an increase in the U.S. federal funds rate, accompanied by a mixed message about future monetary policy, involving the term "a measured approach" to boosting rates, yet also expressing concern about increasing inflation due to a pickup in pricing power. The statement is not fully transparent regarding the future course of policy, so it is perhaps not surprising that it takes some time for the market to reach a consensus about the new equilibrium level for the equity index. More generally, periods of persistent noise are likely to arise during market environments characterized by ambiguous information arrivals and unusual trading patterns. Such features may induce heightened uncertainty and accelerated learning among agents, resulting in a prolonged price discovery process.[1] From this perspective, our results provide a high-frequency analogue to the lower-frequency episodes of market predictability captured by Farmer et al. (2019).

In both cases, the intraday returns display extreme persistence over non-trivial intervals, possibly indicating a period of dynamic price discovery or temporary disequilibrium. One may suspect that such strong drifts in the price constitute a

[1] For extensive discussion of such heterogeneous and incomplete information rationales behind the observed return persistence, see Andersen et al. (2020).

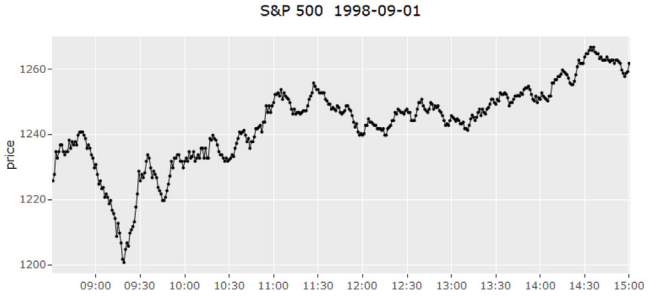

Fig. 1. S&P 500 index futures values on September 1, 1998.

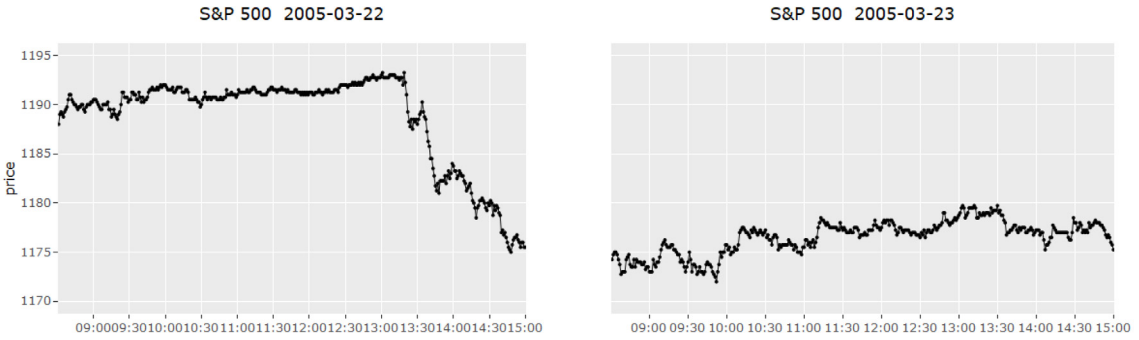

Fig. 2. E-mini S&P 500 index futures price on March 22 and 23, 2005.

Table 1
Daily RV ($\times 10^4$) measured at different frequencies.

Asset	Date	RV_{1m}	RV_{5m}	RV_{10m}	RV_{20m}
S&P 500	1998-09-01	7.14	9.75	10.33	9.72
S&P 500	2005-03-22	0.48	0.58	0.49	0.62

Realized Volatility estimates for the E-mini S&P 500 index futures across the regular trading hours on September 1, 1998, and March 22, 2005.

violation of the basic Itô semimartingale assumption behind all high-frequency volatility estimators—a presumption that we corroborate through a formal statistical test below. Moreover, it is a type of deviation that differs fundamentally from the usual weakly-dependent MMN model. In the latter, noise is typically viewed as uncorrelated with the efficient price, which implies that the unconditional return variation exceeds the return variation of the fundamental return process. The implication is that the realized volatility (RV) measure will tend to decline, if the sampling frequency is lowered substantially. In contrast, given strong price trends, the larger moves will dominate for longer intervals, as high-frequency sampling splits the larger sustained drifts into smaller terms, reducing the cumulative impact of the squared returns. On the other hand, since the price trends appear random and temporary, there are also likely to be quite dramatic shifts in the RV measure for different sampling frequencies. For example, price reversals associated with temporary offsetting trends, e.g., a V-shaped flash crash type trajectory may be missed when the sampling interval lengthens.

A direct way to gauge the impact of irregularities in the price path is through the discrepancy of RV measures obtained at distinct sampling intervals, using the "signature plot" in Andersen et al. (2000). The idea is that RV measures, under the maintained Itô semimartingale condition, should attain similar values across different frequencies. Table 1 provides the daily RV measures for the two trading days above, obtained by sampling at the 1-, 5-, 10-, and 20-min frequencies. For September 1, 1998, the RV increases by almost 45%, as the frequency drops to the 10-min level. For March 22, 2005, the RV measures behave more erratically, as the direction is non-monotonic, but generally increase as the sampling frequency declines.

Of course, these illustrations are merely suggestive. This paper demonstrates how pockets of extremely persistent high-frequency returns may distort the measurement of integrated volatility (IV), and then develops an alternative class of consistent IV estimators, designed to improve the robustness towards this type of episodic Itô semimartingale violations. Finally, the practical impact of the proposed noise persistence-robust estimation is assessed through an out-of-sample volatility forecasting experiment, where predictions based on our new class of robust estimators are compared to forecasts generated from traditional return volatility estimators.

Table 2
Volatility estimates ($\times 10^4$) for Sep 1, 1998, and Mar 22, 2005.

Asset	Date	TV(3)	$DV_{1-3}(3\sqrt{2})$	T^n_{stat}
S&P 500	1998-09-01	5.448	4.645	5.536
S&P 500	2005-03-22	0.407	0.362	4.057

The underlying data are E-mini S&P 500 index futures. The values 3 and $3\sqrt{2}$ in brackets refer to the thresholds parameters C^{TV}_ζ and C^{DV}_ζ defined in Section 4.1. The t-statistic is given by $T^n_{stat} = \sqrt{n}\big(TV^n_t(3) - DV^n_{1-3,t}(3\sqrt{2})\big)\big/(RQ^n_t/3)^{1/2}$, where RQ^n_t is an estimate defined in Eq. (20).

The features of principal interest are non-trivial periods of excessive diffusive return variation induced by extreme persistence in the drift. Consequently, we initially develop an IV estimator robust to this type of no-arbitrage deviation. Specifically, we propose a new family of volatility measures based on differenced returns, denoted DV estimators. The idea of using first differences of returns in constructing power variations has been employed before in Todorov (2013). The analysis there is restricted to Itô semimartingales of pure-jump type (i.e., without a diffusion term) observed without noise, and the role of differencing the returns in Todorov (2013) is to (i) symmetrize the jumps and (ii) minimize the effect of the drift when jumps are of lower activity. By contrast, the differencing of returns is useful in the current context only if the Itô semimartingale is contaminated with a locally persistent component (drift or noise). Importantly, our DV estimators continue to work, even if the observed price is the usual Itô semimartingale. The only cost incurred by the use of return differencing for volatility measurement is a loss of information, i.e., an efficiency loss, during the periods when the observed price process truly behaves like an Itô semimartingale. We show, however, that this loss of efficiency can be minimized by averaging over DV estimators based on differenced returns across several lags.

Our DV estimators ensure a negligible asymptotic bias for a diverse set of persistent return patterns. In particular, we prove this to be the case for the Christensen et al. (2020) drift-burst model and for other general persistent-noise models that we develop here. These results are complemented by a comprehensive Monte Carlo study, exploring the finite-sample properties of alternative estimators. We confirm that the DV measures attain less finite-sample bias compared to existing volatility estimators under empirically-realistic persistent noise scenarios. We further verify that the cost of deviating from the usual jump-robust RV type of estimators, inevitably, is a (moderate) loss in efficiency under the null hypothesis of no noise contamination.

To foreshadow some of our findings, we provide IV estimates for the return series corresponding to the dates in Fig. 1 and the left panel of Fig. 2, obtained from two jump-robust estimators, the Threshold RV estimator, TV^n, and a version of our DV estimator, DV^n_{1-3}. These estimators are formally introduced in Sections 3.1 and 3.2. Table 2 reports the results, obtained by sampling at the one-minute frequency. Due to the truncation of large squared returns, designed to eliminate the jump contribution to the return variation, the TV estimate is substantially lower than the RV reported in Table 1. More interestingly, the DV estimate is substantially below the TV estimate—by 15% and 11% for the two dates—indicating a strong effect of the correction for a persistent return autocorrelation implemented through the DV estimator. Indeed, the gap is highly statistically significant in both cases, as indicated by the t-statistic reported in the last column of Table 2. This statistic is asymptotically standard normally distributed under the usual Itô semimartingale assumption. As above, the design of the test statistic follows from results derived in Sections 3.1 and 3.2. Thus, our formal analysis provides strong evidence of a significant deviation from the Itô semimartingale hypothesis across the two trading days, suggesting that standard IV estimators may be severely upward biased.

Our empirical application focuses on out-of-sample prediction of future volatility. If our procedure is effective in mitigating IV and RV measurement errors, it should provide better inputs into the forecasting scheme and improve the quality of the resulting predictions. The literature on volatility forecasting has been active over the past decade, following the work of Andersen et al. (2003), documenting that a simple ARFIMA model for daily RV outperforms standard parametric GARCH type models. Corsi (2009) proposes an even simpler approximate long-memory Heterogeneous Autoregressive Realized Volatility (HAR-RV) model which, quite robustly, generates good out-of-sample forecast performance. Due to its ease of implementation, the HAR-RV model is commonly adopted and further modified in subsequent work, including Andersen et al. (2007), Patton and Sheppard (2015), and Bollerslev et al. (2016). We follow this approach and exploit HAR-type models based on our DV estimators, labeled "HAR-DV". We apply these new HAR-DV models on high-frequency data for the S&P 500 index futures and 30 individual constituents of the DJIA, and compare their performance to a set of commonly used benchmark models. We find the HAR-DV model to perform very well when assessed through popular criteria such as the MSE and QLIKE loss functions.[2]

The remainder of the paper is organized as follows. Section 2 describes the setting and Section 3 introduces the persistent drift-robust DV estimator. Section 4 describes our Monte Carlo study, illustrating the finite-sample performance. Section 5 contains the empirical application of our HAR-DV model to forecasting realized variance. Section 6 concludes. All formal proofs are relegated to Appendix A.1.

[2] A detailed discussion about the advantages of these loss functions for realized volatility forecasting is provided by Patton (2011).

2. Price dynamics and the observation scheme

We denote the observed log-price process by Y_t. It is defined on a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$, and we assume it is given by,

$$Y_t = X_t + H_t, \tag{1}$$

where X_t is the underlying efficient price, modeled as an Itô semimartingale, and H_t is a component that accommodates pockets of extreme persistence in returns which cannot be generated by increments in X. This section describes our assumptions regarding the dynamics of the two components for Y_t in Eq. (1), and introduces the observation scheme.[3]

2.1. The efficient price process

The efficient log-price process X follows a one-dimensional Itô semimartingale,

$$X_t = X_0 + \int_0^t b_s\, ds + \int_0^t \sigma_s\, dW_s + J_t\,, \tag{2}$$

where the initial value X_0 is $\mathcal{F}_0$-measurable, the drift coefficient b_t takes value in $\mathbb{R}$, $W = (W_t)_{t\geq 0}$ is a standard Brownian motion, and J_t denotes a jump component of the following form,

$$J_t = \int_0^t \int_{\mathbb{R}} \delta(s,x)\, \mu(ds, dx), \tag{3}$$

where $\delta : \mathbb{R}_+ \times \mathbb{R} \mapsto \mathbb{R}$ is a predictable mapping, and μ is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}$ with compensator $\nu(dt, dx) = dt \otimes F(dx)$. The dynamics of X in Eq. (2) are quite general. It allows for stochastic drift, stochastic volatility, and jumps with stochastic intensity. In line with most existing work, we restrict the jumps in X to be of finite variation, i.e., they are absolutely summable. We impose the following assumption on the efficient price process:

Assumption 1. For the process X in Eq. (2), we have,

(a) The process b_t is locally bounded.
(b) The process σ_t is càdlàg.
(c) There is a localizing sequence (T_n) of stopping times and for each n, a deterministic nonnegative function Γ_n on $\mathbb{R}$ satisfying $\int_{\mathbb{R}} |\Gamma_n(x)|^r F(dx) < \infty$, for some $r \in [0, 1)$, and such that $|\delta(s, x)| \wedge 1 \leq \Gamma_n(x)$ for all (ω, s, x) with $s \leq T_n(\omega)$.

Assumption 1 is adapted from Jacod and Protter (2012). It imposes fairly minimal structure on the dynamics of the processes b and σ as well as the function $\delta(s, x)$. In particular, we allow for general forms of dependence between the return innovations W and J and the evolution of b, σ and $\delta(s, x)$, so that, in particular, the so-called leverage effect is accommodated. Most stochastic volatility models used in applied work are covered by the above assumption, e.g., when volatility is an Itô semimartingale itself, or it is a process driven by fractional Brownian motion.

2.2. The observation scheme

We consider equidistant sampling at $t_i = i\Delta_n$, for $i = 0, 1, \ldots, nT$, where $\Delta_n = 1/n$ is the length of each increment within the fixed time interval $[0, T]$. The high-frequency log-returns are,

$$\Delta_i^n Y = Y_{i\Delta_n} - Y_{(i-1)\Delta_n}. \tag{4}$$

For ease of notation, we normalize the trading day to unity and, in turn, the integer T refers to the number of days, and $n = 1/\Delta_n$ (assumed to be an integer) represents the number of observations per (trading) day. We employ a standard infill asymptotic scheme, where the gap between observations, Δ_n, shrinks towards 0, or, equivalently, $n \to \infty$. The equidistant sampling assumption is readily generalized to non-equidistant cases by requiring $\max_{\{i\in 1,\ldots,nT\}}(t_i - t_{i-1}) \to 0$. The time span of the data, T, is fixed throughout.

2.3. Models for extreme persistence in observed returns

We consider two alternative models capable of generating pockets of extreme persistence in high-frequency returns, so that episodes akin to the flash crash and gradual jump patterns can be produced at a reasonable frequency in line with observations from actual market data.

[3] The above setup and the analysis that follows can be extended by allowing for the observed price to contain, in addition to H, also a standard weakly-dependent microstructure noise component. We do not consider this extension here, as we focus on the locally persistent component, H. Moreover, in our applications, we sample at fairly coarse frequencies, where the impact of the weakly-dependent type of noise is minimal.

2.3.1. The drift burst model

Christensen et al. (2020) provide an initial formalization of episodes characterized by an extreme drift in the observed returns. Their "drift burst" model motivates our own representation in the following section, so it serves as a natural benchmark, and we provide this representation in Assumption 2. Section 2.3.2 generalizes this formulation to allow for a stochastic evolution of persistent mean drift episodes, while strictly nesting the original drift burst specification.

Assumption 2. The bursting drift term in Christensen et al. (2020) is modeled as,

$$H_t = \int_0^t \left(c_l \mathbb{1}_{\{s\in(\tau_l,\tau_{db})\}} (\tau_{db} - s)^{-\alpha} + c_r \mathbb{1}_{\{s\in(\tau_{db},\tau_r)\}} (s - \tau_{db})^{-\alpha} \right) ds, \quad \alpha \in (1/2, 1), \tag{5}$$

for some constants c_l, c_r and $\tau_l \leq \tau_{db} \leq \tau_r$.

The process H in Assumption 2 is a finite variation semimartingale, not an Itô semimartingale, due to the drift coefficient – the integrand in Eq. (5) – not being locally bounded, but exploding at the drift burst time τ_{db}. On one hand, if c_l and c_r are of opposite signs, this specification can mimic the dynamics observed during flash crashes. On the other hand, if $c_r = 0$, it can produce patterns akin to the so-called gradual jumps.

A notable feature of the drift burst representation is that the size of the increment $\Delta_i^n H$ is deterministically linked to the distance of $i\Delta_n$ from the given drift burst time τ_{db}, with small $|i\Delta_n - \tau_{db}|$ values corresponding to a large value for $|\Delta_i^n H|$. We further note that Christensen et al. (2020) treat the H term as a component of the efficient price, implying that the observed return persistence is due to a strong drift in the fundamental price rather than a temporary increase in the degree of autocorrelation of the noise component.[4]

The deterministic specification of the drift burst dynamics is a useful simplification that facilitates modeling, but it should probably not be taken literally to imply that the timing, depth or length of such incidents are known by investors in advance, or even as they unfold.

2.3.2. The persistent noise model

We now introduce an alternative specification of H that can generate episodic flash crashes and gradual jumps along with a variety of other complex price patterns, characterized by spurts of extreme serial correlation in the high-frequency returns. It may be viewed as a stochastic extension of the drift burst model, that can be integrated directly with regular stochastic representations of the asset price dynamics. An important conceptual distinction is that we treat H as a (strongly dependent) noise process and not a component of the efficient price. However, this distinction is immaterial for the inference procedure for volatility that we develop subsequently.

For notational convenience, the specification below describes only the first occurrence of a "persistent noise" episode in a given sample. As time evolves, new episodes arise randomly.

Assumption 3 (*Persistent Noise*)**.** The sequence of stopping times, $0 \leq \tau_1 \leq \tau_2 \leq \cdots \leq \tau_N \leq T$, with N finite almost surely, exists such that,

$$H_t = \left(\sum_{i:\tau_i \leq t} H_t^{(i)} \right) \mathbb{1}_{\{\epsilon_t \geq 0\}}, \tag{6}$$

with $H_t^{(i)}$ given by,

$$H_t^{(i)} = f^{(i)}(\Delta X_{\tau_i}, \eta_{\tau_i}) g^{(i)}(t), \quad t \geq \tau_i, \tag{7}$$

where η_{τ_i} is an $\mathcal{F}_{\tau_i}$-adapted random variable, $f^{(i)}$ is a continuous and bounded function, and $g^{(i)}$ is one of the following two functions,

$$\begin{cases} g_{gj}^{(i)}(s) = \left\{ 1 - \left(\frac{s-\tau_i}{\overline{\tau}_i - \tau_i} \right)^{\alpha_g} \right\} \mathbb{1}_{\{s\in[\tau_i,\overline{\tau}_i]\}}, \text{ for some } \mathcal{F}_{\tau_i}\text{-adapted random } \overline{\tau}_i > \tau_i, \text{ and constant } \alpha_g \in \left(0, \frac{1}{2}\right), \\ g_{fc}^{(i)}(s) = c_l^{(i)} \left\{ 1 - \left(\frac{\breve{\tau}_i - s}{\breve{\tau}_i - \tau_i} \right)^{\alpha_l} \right\} \mathbb{1}_{\{s\in[\tau_i,\breve{\tau}_i]\}} + c_r^{(i)} \left\{ 1 - \left(\frac{s-\breve{\tau}_i}{\overline{\tau}_i - \breve{\tau}_i} \right)^{\alpha_r} \right\} \mathbb{1}_{\{s\in[\breve{\tau}_i,\overline{\tau}_i]\}}, \\ \text{for some } \mathcal{F}_{\tau_i}\text{—adapted random variables } c_l^{(i)} \text{ and } \breve{\tau}_i > \tau_i \text{ and some } \mathcal{F}_{\breve{\tau}_i}\text{- and} \\ \mathcal{F}_{\breve{\tau}_i}\text{-adapted random variables } c_r^{(i)} \text{ and } \overline{\tau}_i > \breve{\tau}_i, \text{ and constants } \alpha_l,\ \alpha_r \in \left(0, \frac{1}{2}\right), \end{cases} \tag{8}$$

[4] This does not necessarily imply arbitrage in the Christensen et al. (2020) representation, as they allow for a corresponding local explosion in the volatility coefficient.

and,

$$\epsilon_t = \sum_{s\in[\tau,t]} \Delta\epsilon_s\,,$$

is a finite activity pure-jump process with negative jumps.

The observation-time noise process, H, in Eq. (6) starts out at zero, but is activated and then terminated randomly depending on the realizations of the stopping times τ_1, and $\overline{\tau}_1$, and the (non-positive) jump process ϵ_t. Hence, the initial phase is governed solely by the $H_t^{(1)}$ component (7), with an impact that persists across the full (random) interval $[\tau_1, \overline{\tau}_1]$, unless killed off (randomly) by the ϵ_t process. Moreover, additional persistent noise terms are added during the episode, if some of the subsequent stopping times, $\tau_2 \leq \cdots \leq \tau_N$, activate in the interim, implying the drift may display abrupt shifts in size and direction. Prior to and following the persistent noise episode, $Y_t = X_t$, i.e., the observed price equals the efficient price.

We let the $\mathcal{F}_{\tau_i}$-adapted function $g^{(i)}(\cdot)$ capture the evolution of the H term over the relevant interval. Obviously, this temporary increase in the stochastic order of the mean drift can take a variety of distinct forms. However, given the identification in the literature of gradual jump and flash crash type patterns as common phenomena during turbulent market conditions featuring strongly correlated price changes, we focus on such events. Hence, our basic noise component takes the form of either a (random) flash crash or a gradual jump, but with possible premature termination through a jump that brings the price back to the efficient level. We allow for the possibility that the event triggering enhanced market uncertainty at time τ_1, and initiating the persistent noise episode, is associated with a jump in either the efficient price or the market price, or both. This feature is controlled by the $f^{(i)}(\cdot)$ function in Eq. (7).

In the gradual jump scenario, $g^{(1)} \equiv g_{gj}^{(1)}$, the efficient price jumps, $\Delta X_{\tau_1} \neq 0$, but market participants are imperfectly informed about this shift in the economic fundamentals. That is, even if the agents infer from incoming news or unusual trading activity, that something may have shifted in the market environment, they are not able to ascertain with certainty whether the efficient price has moved or not. The initial receipt of (incomplete) information may induce an immediate price change, but it will typically not mirror the shift in the efficient price. Specifically, letting $f^{(1)}(\Delta X_{\tau_1}, \eta_{\tau_1}) = -\eta_{\tau_1}\Delta X_{\tau_1}$ allows H to (partially) offset this jump. For example, if $\eta_{\tau_1} = 1$, we have $H_{\tau_1} = f^{(1)}(\Delta X_{\tau_1}, \eta_{\tau_1}) = -\Delta X_{\tau_1}$, implying $\Delta Y_{\tau_1} = 0$, and there is no induced jump in the market price. In contrast, for $\eta_{\tau_1} = \eta \in (0, 1)$, we will observe an initial jump, that partially incorporates the efficient price jump.

In the flash crash scenario, $g^{(1)} \equiv g_{fc}^{(1)}$, the efficient price is unchanged, $\Delta X_{\tau_1} = 0$, but market participants are increasingly worried about a potential shift in fundamentals due to news arrivals or unusual trading patterns. There is not, necessarily, a jump in the market price, but the uncertainty triggers a dynamic with a strong price trend that may further unsettle the agents. Nonetheless, at some point, perhaps due to basic information acquisition, revelations about the source of the initial trading anomalies, or external news releases, the realization that fundamentals may not have shifted starts gaining ground, and arbitrageurs begin to trade in the opposite direction, leading the price back towards the efficient price. Our flash crash dynamic is a stylized representation of the ensuing price path, with a turning point at a (random) time $\check{\tau}_1$.

Examples of price paths resembling a flash crash, a gradual jump, or a gradual jump with an intermittent (small) flash crash, generated by our persistent noise model are provided in Fig. 3. Importantly, in all these scenarios, the agents are aware that the uncertainty may be resolved at any point, leading to a jump, either back to the efficient price (no jump in fundamentals) or to a new efficient price level (the fundamentals did shift abruptly). Thus, there is genuine risk that complicates arbitrage strategies seeking to exploit the serial correlation in returns.

Conceptually, our model for H in Assumption 3 is an extension of Assumption 2. In particular, in both settings, the g function is Hölder continuous with a coefficient taking values in $\left(0, \frac{1}{2}\right)$, which ensures that the increments of H are large asymptotically. The major distinction is that H is random and dependent on the underlying efficient price in Assumption 3, while H in Assumption 2 is deterministic and not related to X. Moreover, the ϵ process generates an extra source of randomness in the persistent noise model, so the duration of the episode is inherently uncertain. Finally, the persistent noise model for a gradual jump, triggered by an efficient price jump, not matched by an identical jump in the observed price, is consistent with the finding that many gradual jumps appear in the aftermath of scheduled macroeconomic announcements, likely reflecting a period of intense price discovery, as in the case depicted in Fig. 2.

Relative to existing market microstructure noise models with (weakly) dependent noise, such as Jacod et al. (2017, 2019), the main distinguishing feature of Assumption 3 is the extent of the strong local persistence and the direct linkage to the efficient price. In particular, in prior models of noise, we have $\text{corr}(H_{i/n}, H_{j/n}) \to 0$ as $|i-j| \to \infty$, while, in contrast, Assumption 3 implies this limiting correlation can equal unity.

Finally, a comment on the potential for arbitrage. As for the drift burst model, the persistent noise model in Assumption 3 implies that H is a finite-variation semimartingale, but not an Itô semimartingale. This implies that the existence of an equivalent martingale measure is not guaranteed given our specification of H, so the observed price could allow for arbitrage opportunities (Delbaen and Schachermayer, 1994, 1995). However, recent work by Strasser (2005), Guasoni (2006), Jarrow et al. (2009) and Bender (2012) extends the results of Delbaen and Schachermayer (1994, 1995) by restricting the set of possible investment strategies either through trading only in discrete time or via the introduction of transaction costs. In particular, with arbitrarily small transaction costs, one can check that the so-called sticky price assumption in Guasoni (2006) is satisfied, ensuring the absence of arbitrage.

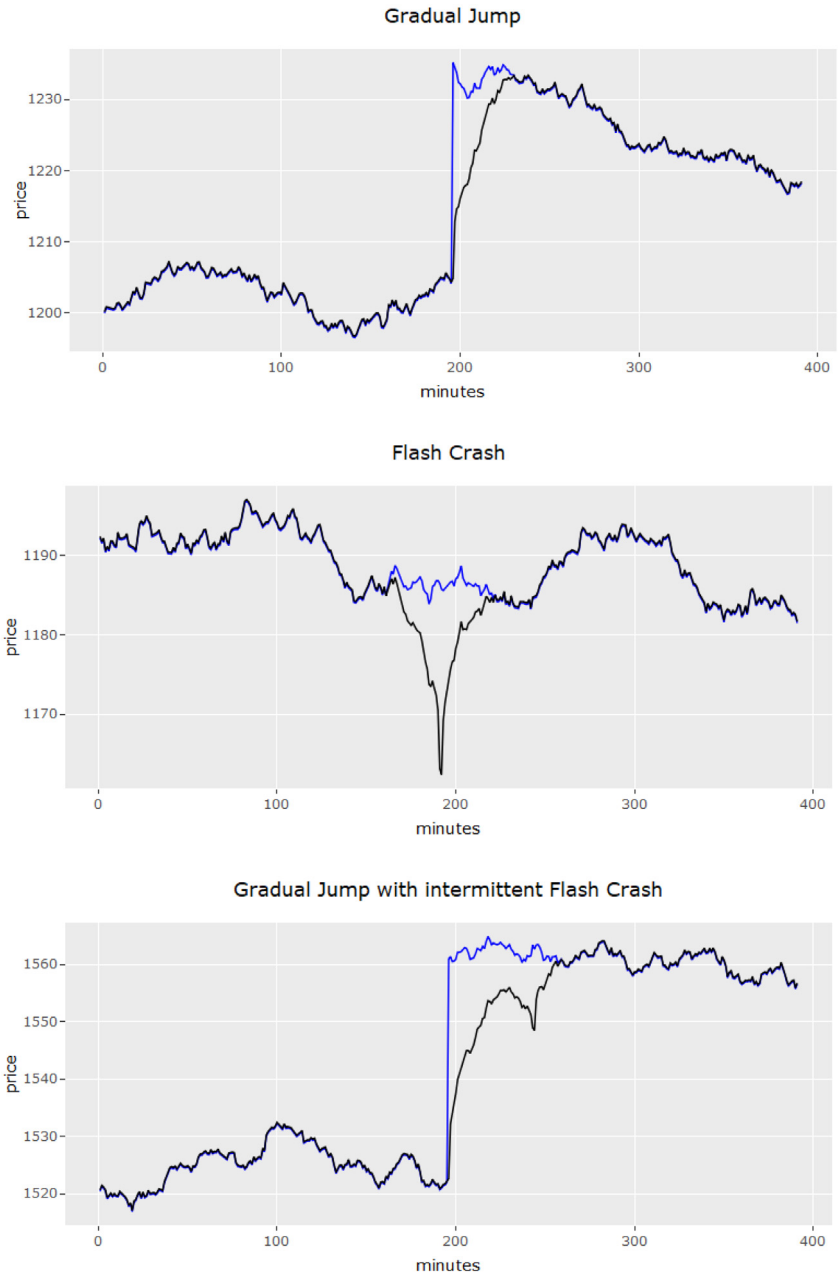

Fig. 3. Simulated price paths of the persistent noise model (6) in Assumption 3 for a gradual jump, flash crash, and gradual jump with an intermittent flash crash. Section 4.1 provides detailed specifications of these cases. The efficient price, $\exp(X)$, is blue and the observed price, $\exp(Y)$, is black. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3. Volatility estimation with extreme return persistence

We now turn to jump-robust volatility estimation. Our goal is to recover the integrated volatility, $\int_{t-1}^{t} \sigma_s^2 \, ds$, from discrete observations of Y, when the observed prices possibly contain a persistent noise component H. From the point of view of volatility estimation, all other components of X and H will be viewed as nuisance parameters. This also means that, as far as volatility estimation is concerned, the exact relation between X and H, e.g., whether H is treated as a component of X or not is immaterial. We finally note that we will not attempt to recover the jumps in X, when the observed price possibly contains H. This problem is more difficult and the answer hinges crucially on the type of H contaminating the observed price.

3.1. The differenced-return volatility (DV) estimator

An efficient estimator of integrated volatility, when $Y \equiv X$, is the truncated volatility (TV) of Mancini (2009). For some $\zeta > 0$ and $\varpi \in \left(0, \frac{1}{2}\right)$, it is given by,

$$TV_t^n = \sum_{i=(t-1)n+1}^{tn} \left(\Delta_i^n Y\right)^2 \mathbb{1}_{\{|\Delta_i^n Y| < \zeta \Delta_n^{\varpi}\}}, \quad t = 1, \ldots, T, \tag{9}$$

Under suitable integrability conditions for X, the Itô semimartingale dynamics imply that,

$$\mathbb{E}_t\left(X_{t+\Delta_n} - X_t\right) = O_p\left(\Delta_n\right) \text{ and } \mathrm{Var}_t\left(X_{t+\Delta_n} - X_t\right) = O_p\left(\Delta_n\right), \quad \text{for } t \text{ fixed, as } \Delta_n \to 0. \tag{10}$$

This result implies that the conditional mean for high-frequency returns of X is of smaller asymptotic order than the conditional standard deviation. Hence, for volatility estimation from high-frequency returns, we do not need to center the latter around their conditional mean.

Things change if the observed price Y contains H, as the latter renders the conditional mean of higher asymptotic order. Although estimation of the mean over short stretches of data is near impossible, we can minimize its impact on volatility estimation, as long as the conditional mean return possesses a certain degree of smoothness in time, by first-order differencing of the high-frequency returns. This is effectively analogous to the removal of fixed effects in panel data.

To obtain a volatility estimator from differenced returns, we can then make use of the properties of the Brownian motion and the càdlàg assumption for the volatility path to obtain,

$$2\,\mathrm{Var}_t\left(X^c_{t+\Delta_n} - X^c_t\right) \approx \mathrm{Var}_t\left(X^c_{t+2\Delta_n} - 2X^c_{t+\Delta_n} + X^c_t\right), \qquad \text{for } \Delta_n \to 0,$$

where X^c denotes the continuous component of X. This motivates the following DV estimator based on the first-differences of returns,

$$DV_t^n = \frac{1}{2} \sum_{i=(t-1)n+2}^{tn} \left(\Delta_i^n Y - \Delta_{i-1}^n Y\right)^2 \mathbb{1}_{\{|\Delta_i^n Y - \Delta_{i-1}^n Y| < \zeta \Delta_n^{\varpi}\}}, \quad t = 1, \ldots, T. \tag{11}$$

In the following theorems, we provide joint convergence results for TV and DV under the semimartingale model in Assumption 1, the drift burst model in Assumption 2, and the persistent noise model in Assumption 3, respectively. We denote stable convergence in law by $\xrightarrow{\mathcal{L}-s}$ and we let $\mathcal{MN}$ refer to a mixed normal distribution, i.e., a normal distribution conditional on the realization of its $\mathcal{F}$-conditional variance, which is a random variable.

Theorem 1. *Suppose the efficient price X evolves according to Eq. (2) and satisfies Assumption 1. During an ordinary period with $H = 0$ and for $\varpi \in \left[\frac{1}{4-2r}, \frac{1}{2}\right)$, we have,*

$$\sqrt{n}\begin{pmatrix} TV_t^n - IV_t \\ DV_t^n - IV_t \end{pmatrix} \xrightarrow{\mathcal{L}-s} \mathcal{MN}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 2 \\ 2 & 3 \end{pmatrix} IQ_t\right), \tag{12}$$

where $IV_t = \int_t^{t+1} \sigma_s^2\, ds$ *and* $IQ_t = \int_t^{t+1} \sigma_s^4\, ds$.

The result for TV is well known, see, e.g., Theorem 13.24 in Jacod and Protter (2012). We have flexibility in choosing the TV threshold, which is crucial for its finite-sample performance, when jumps are less active, i.e., for low values of r (r controls the rate of explosion of the Lévy density near the origin). In particular, for finite activity jumps, $r = 0$, commonly invoked in parametric models used for applied work, ϖ may take values in the entire interval $\left[\frac{1}{4}, \frac{1}{2}\right)$.

Theorem 1 indicates that, under the Itô semimartingale assumption, the DV estimator is less efficient than TV, with $\sqrt{3/2}$ representing the inflation in the standard error. The efficiency loss is intuitive. Indeed, ignoring the truncation indicator, the main part of DV may be written,

$$\frac{1}{2} \sum_{i=(t-1)n+2}^{tn} \left(\Delta_i^n Y - \Delta_{i-1}^n Y\right)^2 \approx \sum_{i=(t-1)n+1}^{tn} \Delta_i^n Y^2 - \sum_{i=(t-1)n+2}^{tn} \Delta_i^n Y \Delta_{i-1}^n Y,$$

where the first term equals the main component of the TV estimator, and the second may be viewed as a term trading off efficiency and robustness. Specifically, by sacrificing some efficiency in IV estimation, this term assists in removing the estimation bias due to the possible (first-order) autocorrelation in high-frequency returns. In the following two theorems, we explicitly show to what extent, relative to TV, the DV estimator can reduce the estimation bias in volatility estimation due to the presence of the H component in the observed price process.

Theorem 2. *Suppose there is a drift-burst period of the form (5) with $\tau_{db} = 0$ for day t and Assumption 2 holds with $J \equiv 0$. Then, for $\varpi \in [1/4, 1/2)$ with $\alpha > 1 - \varpi$, we have,*

$$\begin{aligned} TV_t^n &= IV_t + O_p\left(\Delta_n^{2\varpi - \frac{\varpi+\alpha-1}{\alpha}} \bigvee \sqrt{\Delta_n}\right), \\ DV_t^n &= IV_t + O_p\left(\Delta_n^{2\varpi - \frac{\varpi+\alpha-1}{\alpha+1}} \bigvee \sqrt{\Delta_n}\right). \end{aligned} \tag{13}$$

For simplicity, Theorem 2 is stated only for the case when X does not contain jumps, as this suffices to demonstrate the effect from a drift burst in the observed price.

The drift burst induces a bias term of order $O_p\big(\Delta_n^{2\varpi - \frac{\varpi+\alpha-1}{\alpha}}\big)$ in TV. The source of bias is twofold. One is due to the truncation of price increments very close to the drift burst time τ_{db}. In this neighborhood, H contributes to large returns and, due to truncation, this precludes estimation of volatility over this region. The second source of bias stems from the presence of H in price increments not in the vicinity of τ_{db}, which are unlikely to trigger truncation.

Theorem 2 implies that the DV estimator suffers much less estimation bias under the drift burst model, simply because $\Delta_i^n H - \Delta_{i-1}^n H$ is of lower asymptotic order than $\Delta_i^n H$ near τ_{db}. Hence, less truncation is induced by H and, even when truncation is not triggered, the role of H is smaller for the differenced increments.

The bias due to drift burst is asymptotically negligible and has no effect on the CLT for TV, if $\alpha < \frac{1-\varpi}{3/2-2\varpi}$. The associated condition for DV is weaker: $\alpha < \frac{1/2+\varpi}{3/2-2\varpi}$. This implies that, whenever $\varpi < 1/2$, there will be a sufficiently high value of $\alpha < 1$ rendering the CLT for TV in Theorem 1 invalid while, for this to happen for DV, we require $\varpi < 1/3$. Recall that for Theorem 1 to apply, we need $\varpi < 1/2$. Thus, by picking $\varpi > 1/3$ and close to $1/2$, DV is robust to drift burst of any kind, while TV is not robust to drift bursts with parameter $\alpha \in [\frac{1-\varpi}{3/2-2\varpi}, 1)$.

We next state the result for volatility estimation when H signifies persistent noise.

Theorem 3. *Suppose there is a persistent noise period of the form (6) for day t and Assumption 3 applies with $r = 0$ (finite activity jumps). Then, restricting attention to the set,*

$$\Omega^n = \{\omega : \exists \tau \in [t-1, t] \neq 0\},$$

where τ denotes the stopping time in Assumption 3, we have for $\varpi \in (1/4, 1/2)$ with $\alpha_l \vee \alpha_r < \varpi$,

$$\begin{aligned} TV_t^n &= IV_t + O_p\left(\Delta_n^{2\varpi - \frac{\varpi-\alpha_l\wedge\alpha_r}{1-\alpha_l\wedge\alpha_r}} \bigvee \sqrt{\Delta_n}\right), \\ DV_t^n &= IV_t + O_p\left(\Delta_n^{2\varpi - \frac{\varpi-\alpha_l\wedge\alpha_r}{2-\alpha_l\wedge\alpha_r}} \bigvee \sqrt{\Delta_n}\right). \end{aligned} \tag{14}$$

This result is analogous to the one in Theorem 2. Obviously, the asymptotic effect of the noise term H depends on the lower of the two parameters α_l and α_r, governing the roughness of the path of H (lower values imply rougher trajectories).

3.2. A general family of DV estimators

We now extend the DV estimator in Eq. (11) to a more general set of estimators by differencing the returns at higher order lags. This generalized version is more efficient under the Itô semimartingale assumption and achieves the identical asymptotic order of bias reduction under the drift-burst and persistent-noise models for H. Specifically, the DV estimator may be nested in the following general class of estimators,

$$DV_{m,t}^n = \frac{1}{2} \sum_{i=(t-1)n+m+1}^{tn} \left(\Delta_i^n Y - \Delta_{i-m}^n Y\right)^2 \mathbb{1}_{\{|\Delta_i^n Y - \Delta_{i-m}^n Y| \leq \zeta \Delta_n^{\varpi}\}}, \quad m = 1, 2, \ldots \tag{15}$$

We introduce the following new volatility estimator by simply averaging DV_1, DV_2, …, DV_m,

$$DV_{1-m,t}^n = \frac{1}{m}\left(DV_t^n + DV_{2,t}^n + \cdots + DV_{m,t}^n\right). \tag{16}$$

The DV_{1-m} estimator can be viewed as an approximation of the TV estimator, extended with a set of correction terms accounting for 1st to mth order autocorrelation. Theorems 4–6 constitute the counterparts to Theorems 1–3 for the general DV_{1-m} estimator.

Theorem 4. *Under the identical conditions as in Theorem 1, we have,*

$$\sqrt{n}\begin{pmatrix} TV_t^n - IV_t \\ DV_{1-m,t}^n - IV_t \end{pmatrix} \xrightarrow{\mathcal{L}-s} \mathcal{MN}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}2 & 2\\ 2 & 2+1/m\end{pmatrix} IQ_t\right). \tag{17}$$

Theorem 5. *Under the same conditions as in Theorem 2, we have*

$$DV^n_{1-m,t} \;=\; IV_t + O_p\left(\Delta_n^{2\varpi-\frac{\varpi+\alpha-1}{\alpha+1}} \bigvee \sqrt{\Delta_n}\right). \tag{18}$$

Theorem 6. *Under the same conditions as in Theorem 3, we have*

$$DV^n_{1-m,t} \;=\; IV_t + O_p\left(\Delta_n^{2\varpi-\frac{\varpi-\alpha_l\wedge\alpha_r}{2-\alpha_l\wedge\alpha_r}} \bigvee \sqrt{\Delta_n}\right). \tag{19}$$

Theorems 5 and 6 provide identical results to Theorems 2 and 3, implying that the DV_{1-m} estimator is as effective in reducing the bias in IV estimation as the DV estimator in terms of asymptotic order. Theorem 4, on the other hand, shows that DV_{1-m} suffers a much smaller efficiency loss as a consequence of the "efficiency-robustness trade-off" with respect to TV. The efficiency gains for DV_{1-m}, compared to DV, may come across as a free lunch, yet it does have a simple rationale: DV_{1-m} exploits the fact that, during flash crash or gradual jump episodes (equivalently, under the drift-burst or persistent-noise model for H), the mean distortion in the observed returns is approximately identical across consecutive small increments.

Section 4 demonstrates through simulations that the DV estimators can be very effective in reducing the bias for IV estimation stemming from various phenomena generating extreme local persistence in the observed high-frequency returns.

3.3. Feasible inference

For feasible implementation of the CLT results, we need consistent estimates of the conditional asymptotic variances of the corresponding limiting variables which, in turn, all depend on $\int_{t-1}^t \sigma_s^4 ds$. We may form such estimators using the differenced increments in an analogy to the DV estimators for the integrated volatility above. More specifically, we define,

$$RQ_t^n = \frac{1}{12}\frac{1}{\Delta_n}\sum_{i=(t-1)n+2}^{tn}\left(\Delta_i^n Y - \Delta_{i-1}^n Y\right)^4 \mathbb{1}_{\{|\Delta_i^n Y-\Delta_{i-1}^n Y|\le\zeta\,\Delta_n^{\varpi}\}}. \tag{20}$$

We can further generalize this estimator by using the higher order differences $\Delta_i^n Y - \Delta_{i-m}^n Y$, for some integer m. The next theorem shows that RQ_t^n is a consistent estimate of the integrated quarticity, even in the presence of drift bursts or persistent noise.

Theorem 7. *Under the same conditions as in Theorem 3 and, in addition, $\varpi > \frac{2}{7}$, we have,*

$$RQ_t^n \;\xrightarrow{\mathbb{P}}\; \int_{t-1}^t \sigma_s^4 ds. \tag{21}$$

To establish the consistency result for RQ_t^n, we need to strengthen the requirement on the truncation parameter ϖ slightly relative to Theorem 3. Since RQ_t^n is an estimate for a higher order high-frequency return moment, it is naturally harder to estimate in the presence of pockets of extreme return persistence. Even so, in practice, one typically will choose ϖ close to 1/2 for added robustness against jumps, so the sharper restriction on ϖ is not binding in practice.

3.4. Testing for presence of extreme return persistence

Our primary focus is inference for volatility that is robust to pockets of extreme return persistence. Nonetheless, in this section, we develop an explicit test for the presence of such episodes of elevated persistence over a given fixed time interval. We note, however, that the nature and duration of these events are unknown and therefore might induce biases, that are inherently difficult to detect. As such, we only expect our test to identify some of the more significant deviations from the standard semimartingale setting. One option is to construct our test from the statistical significance of TV-DV, which equates to the autocovariance of the truncated returns. Since there is an element of arbitrariness in picking the threshold level (we require thresholds both for TV and DV) and, importantly, to gain statistical power, we instead base our test on the raw return autocovariances, absent truncation. This means that the jump component will affect the limiting distribution of the statistic. Formally, our statistic is defined by,

$$T_t^n(k) = \frac{\sum_{i=(t-1)n+k+1}^n \Delta_i^n Y \Delta_{i-k}^n Y}{\sqrt{\Delta_n \widehat{\mathrm{Avar}}(T_t^n(k))}}, \qquad k = 1, 2, \dots, \tag{22}$$

where the estimate for the asymptotic variance is given by,

$$\widehat{\mathrm{Avar}}(T_t^n(k)) = RQ_t^n + \left[\sum_{i=(t-1)n+2}^{n}(\Delta_i^n Y - \Delta_{i-1}^n Y)^2 \mathbb{1}_{\{|\Delta_i^n Y-\Delta_{i-1}^n Y|>\zeta\,\Delta_n^{\varpi}\}}(\widehat{c}_{i,-}^n + \widehat{c}_{i,+}^n)\right] \bigwedge RQ_t^n \log(n), \tag{23}$$

and

$$\widehat{c}^n_{i,-} = \frac{1}{\Delta_n k_n}\frac{1}{2}\sum_{j=1}^{k_n}\left(\Delta^n_{i-j-1}Y - \Delta^n_{i-j-2}Y\right)^2 \mathbb{1}_{\{|\Delta^n_{i-j-1}Y - \Delta^n_{i-j-2}Y| \le \zeta\Delta_n^{\varpi}\}}, \tag{24}$$

$$\widehat{c}^n_{i,+} = \frac{1}{\Delta_n k_n}\frac{1}{2}\sum_{j=1}^{k_n}\left(\Delta^n_{i+j+1}Y - \Delta^n_{i+j}Y\right)^2 \mathbb{1}_{\{|\Delta^n_{i+j+1}Y - \Delta^n_{i+j}Y| \le \zeta\Delta_n^{\varpi}\}}, \tag{25}$$

for some sequence $k_n \to \infty$ with $k_n\Delta_n \to 0$. The second term in $\widehat{\text{Avar}}(T^n_t(k))$ is an estimate of the asymptotic variance of the summands in the return autocovariance that contain jumps. The next theorem shows that under the null hypothesis of $Y \equiv X$, the test statistic has a standard normal distribution.

Theorem 8. *Under the conditions of Theorem 1, we have,*

$$T^n_t(k) \xrightarrow{\mathcal{L}-s} Z_k, \quad \text{for } k = 1, 2, \ldots, \tag{26}$$

where $\{Z_k\}_{k=1,2,\ldots}$ is a sequence of independent standard normal random variables.

We note that the asymptotic behavior of autocovariances of Itô semimartingales, that do not contain jumps, has been derived already in Kinnebrock and Podolskij (2008). The above theorem considers the case with jumps. It is noteworthy that the $\mathcal{F}$-conditional behavior of the return autocovariance remains mixed-Gaussian in spite of the jumps in X. This is unlike many other high-frequency statistics in which jumps appear in the limiting distribution, see, e.g., Jacod and Protter (2012). The difference stems from the fact that the autocovariances involve products of neighboring returns. This causes the uncertainty about the location of the jump in the discrete sampling interval to be of higher asymptotic order.

From the proof of Theorems 2–3 as well as Theorem 7, we infer that a test based on whether $T^n_t(k)$ lies above a critical value will have power against the presence of the term H in the observed process Y, as long as $\alpha_l \wedge \alpha_r < \frac{1}{2}$. Critically, this implies that we should avoid the use of truncation, when constructing the autocovariance in $T^n_t(k)$, because, when H is present, many of the increments of X will exceed the threshold due to the trend induced by H. Thus, their removal will lead to a loss of power.

The above result for $T^n_t(k)$ can be generalized slightly by using the pre-averaged increments instead of the raw increments, i.e., we can form,

$$\overline{U}_j = \sum_{i=(j-1)l+1}^{jl} g\left(\frac{i-(j-1)l}{l}\right)\Delta^n_i Y, \quad j = 1, \ldots, \lfloor n/l \rfloor, \tag{27}$$

for some integer l and $g(s) = s \wedge (1-s)$. Then, we can replace $\{\Delta^n_i Y\}_{i=1,\ldots,n}$ with $\{\overline{U}_j\}_{j=1,2,\ldots,\lfloor n/l \rfloor}$ in constructing $T^n_t(k)$. We may correspondingly construct a properly rescaled threshold, i.e., replace it with $\sqrt{\sum_{i=1}^{l} g^2(i/l)} \times \zeta\Delta_n^{\varpi}$. Given these modifications, it is trivial to show that Theorem 8 continues to apply, provided l is fixed. This is because the autocovariance of the terms $\overline{U}_j$ is a weighted average of the autocovariances at different lags of the raw increments $\Delta^n_i Y$, and the latter are asymptotically independent. The reason for turning to the pre-averaged returns is to guard against the presence of regular market microstructure noise in the observed prices, which will generate a downward bias in $T^n_t(k)$.

Finally, another modification of $T^n_t(k)$ without an asymptotic effect is to winsorize the increments, i.e., consider $(\Delta^n_i Y \wedge c) \vee -c$ or $(\overline{U}_j \wedge c) \vee -c$, for some fixed $c > 0$. Since c is fixed (and not shrinking asymptotically), this will not impact the asymptotic properties of $T^n_t(k)$ in Theorem 8. On the other hand, it will render the statistic less noisy on days with large jumps in the asset price.

We finish this section with a comment on the use of the test statistic $T^n_t(k)$ for improved volatility inference. While DV is robust to the presence of the noise term H in the observed price, this robustness involves some loss of efficiency, as is evident from Theorem 4. This motivates us to consider a hybrid estimator, that is a randomly weighted average of TV and DV, with weights chosen to not sacrifice efficiency (in theory at least), when H is absent. More specifically, we consider the following estimator,

$$HV^n_t = \omega^n_t \times TV^n_t + (1-\omega^n_t) \times DV^n_{1-m,t}, \tag{28}$$

for some weight function of the data $\omega^n_t \in [0, 1]$. Then, if we choose the weight function ω^n_t suitably, we can make this hybrid estimator mimic the behavior of TV^n_t, when H is not present in the observed price, and mimic $DV^n_{1-m,t}$ when H is present. Recall that when there are pockets of extreme return persistence, the bias in volatility estimation can be of a larger asymptotic order than $\sqrt{\Delta_n}$, and hence affect the CLT for TV^n_t.

We can use the test statistic $T^n_t(k)$ to construct an ω^n_t that delivers HV^n_t with the properties detailed above. This is accomplished, if ω^n_t converges to 1, when H is not present in the observed price, and it converges at an exponential rate to 0, if H is present. For this choice of ω^n_t, and due to the stable convergence, we may ensure that HV^n_t has the same CLT as TV^n_t with $Y \equiv X$, and the CLT for $DV^n_{1-m,t}$, when Y and X are not equivalent. That is, this hybrid estimator will achieve

asymptotic efficiency, when persistent noise is absent or its impact on volatility estimation is asymptotically negligible, and it will enjoy the robustness features of $DV^n_{1-m,t}$, when the noise can induce a nontrivial bias in the volatility estimation.

Since the loss of efficiency from using the robust $DV^n_{1-m,t}$ is relatively small, we opt for simplicity and do not analyze this hybrid volatility estimation approach any further here.

4. Monte Carlo study

This section reports on a simulation study designed to explore the finite-sample properties of the DV estimator corresponding to the new asymptotic results developed in Section 3. In the Supplementary Appendix we also report simulation results for the performance of the test for the presence of extreme return persistence derived in the previous section.

4.1. The simulation setting

For the efficient log-price X, we rely on the following model,

$$dX_t = b_t\,dt + \sigma_t\,dW_{1,t} + dJ_{X,t}, \tag{29}$$

$$d\sigma_t^2 = \kappa\,(\gamma - \sigma_t^2)\,dt + \xi\,\sigma_t\,dW_{2,t}, \tag{30}$$

where W_1 and W_2 are correlated standard Brownian motions with $\mathbb{E}\left(dW_{1,t}, dW_{2,t}\right) = \rho\,dt$. The jump term $J_{X,t}$ is a compound Poisson process with intensity p_X and Gaussian jump size governed by the distribution $N(0, \lambda_X^2)$.

In the experiments below, the log-price process X is simulated for $n = 390$ minutes per day and $T = 5000$ days, corresponding to one-minute intraday sampling across the trading day for about 20 years. The initial log-price X_0 equals log(1200), the unit of time is one trading day, and the drift term b_t is set to zero. The volatility process σ_t^2 is initiated at its unconditional mean value of γ on day one while, for other days, it is initiated at the ending value of the previous day. The annualized parameter vector for the Heston model is fixed at $(\kappa, \gamma, \xi, \rho) = (5, 0.0225, 0.4, -\sqrt{0.5})$, following Christensen et al. (2020). For the jump components, we let $p_X = 1/5$ corresponding to one jump per week on average with $\lambda_X = 0.9\%$. These parameter settings imply that the mean value of JV is about 20% of the mean for IV.

We simulate the persistent noise model (6) to generate a flash crash, a gradual jump, and a gradual jump with an intermittent flash crash. Specifically,

(i) for the gradual jump case: we add a jump in X at $\tau = 0.5$ of magnitude 2.5% for each day. We let $i \in \{1\}$, $\tau_1 = \tau$, and employ $f^{(1)}(\Delta X_\tau, \eta_\tau) = \eta_\tau\,\Delta X_\tau$ with $\eta_\tau = -1$ and $g^{(1)} = g^{(1)}_{gj}$. For parameter specifications, we set $(\underline{\tau}_1, \overline{\tau}_1) = (0.50, 0.59)$ and explore $\alpha_g \in \{0.45, 0.35, 0.25\}$ for three scenarios;

(ii) for the flash crash case: we let $i \in \{1\}$ and employ $f^{(1)}(\Delta X_{\tau_1}, \eta_{\tau_1}) = \eta_{\tau_1}$, $g^{(1)} = g^{(1)}_{fc}$. We set $\eta_{\tau_1} = -2\%$, $(\underline{\tau}_1, \check{\tau}_1, \overline{\tau}_1) = (0.41, 0.49, 0.57)$, $c_l^{(1)} = c_r^{(1)} = 1$, and explore $\alpha_l = \alpha_r = \alpha_f \in \{0.45, 0.35, 0.25\}$ for three scenarios;

(iii) for the case of a gradual jump with an intermittent flash crash: we add a jump in X at $\tau = 0.5$ of magnitude 2.5% for each day. We let $i \in \{1, 2\}$. We assume $f^{(1)}(\Delta X_\tau, \eta_\tau) = -\Delta X_\tau$ and $g^{(1)} = g^{(1)}_{gj}$ and set $(\underline{\tau}_1, \overline{\tau}_1) = (0.50, 0.65)$ for the gradual jump pattern. For the embedded (small) flash crash episode, we employ $f^{(2)}(\Delta X_{\tau_2}, \eta_{\tau_2}) = \eta_{\tau_2}$ and $g^{(2)} = g^{(2)}_{fc}$, and set $\eta_{\tau_2} = -0.75\%$, $c_l^{(2)} = c_r^{(2)} = 1$, $(\underline{\tau}_2, \check{\tau}_2, \overline{\tau}_2) = (\tau_{fc}, \tau_{fc} + 0.04, \tau_{fc} + 0.08)$ where τ_{fc} follows an exponential distribution with rate parameter $\lambda = 15$. We again explore $\alpha_g = \alpha_l = \alpha_r = \alpha_{gf} \in \{0.45, 0.35, 0.25\}$ for three scenarios.

The illustrative paths presented previously in Fig. 3 are simulated under the above three modeling choices and parameter specifications.

We next provide details about the truncation level for the TV and DV estimators. We set the threshold parameters in the following data-driven way,[5]

$$\zeta = C_\zeta\,\hat{\sigma}^{med}_{t-1} \quad \text{and} \quad \varpi = \frac{1}{2}, \tag{31}$$

where C_ζ is a positive constant and $\hat{\sigma}^{med}_{t-1}$ is the MedRV jump-robust volatility estimator of Andersen et al. (2012), given by,

$$(\hat{\sigma}^{med}_t)^2 = \frac{\pi}{6 - 4\sqrt{3} + \pi} \sum_{i=t(n-1)+3}^{tn} \text{median}(\,|\,\Delta^n_{i-2}Y|, |\Delta^n_{i-1}Y|, |\Delta^n_i Y|\,)^2. \tag{32}$$

We experiment with two pairs of thresholds for C_ζ: 4 for TV and $4\sqrt{2}$ for DV versus 3 for TV and $3\sqrt{2}$ for DV. The extra factor of $\sqrt{2}$ in C_ζ for the DV estimators is motivated by the fact that the standard deviation of differenced (identical and independent) returns exceeds the one for individual returns by a factor of $\sqrt{2}$.

[5] Theoretically, we should have $\varpi < \frac{1}{2}$, but setting $\varpi = \frac{1}{2}$ for a fixed sampling frequency obviously makes no difference.

Table 3
Monte Carlo RMSE results.

Model	RV	TV	DV	DV_{1-3}	TV	DV	DV_{1-3}
		$(C_\zeta^{TV}, C_\zeta^{DV}) = (4, 4\sqrt{2})$			$(C_\zeta^{TV}, C_\zeta^{DV}) = (3, 3\sqrt{2})$		
Null, $H = 0$	65.71	7.29	8.99	7.80	7.79	9.19	8.13
PN-GJ, $\alpha_g = 0.45$	85.74	14.98	9.43	8.42	10.69	8.84	7.58
PN-GJ, $\alpha_g = 0.35$	104.05	13.12	9.15	8.02	9.39	8.92	7.61
PN-GJ, $\alpha_g = 0.25$	147.59	10.97	9.02	7.91	8.20	8.93	7.62
PN-FC, $\alpha_f = 0.45$	98.03	22.06	12.16	10.66	17.95	9.11	7.87
PN-FC, $\alpha_f = 0.35$	125.48	19.77	10.50	8.98	16.61	9.06	7.84
PN-FC, $\alpha_f = 0.25$	186.55	16.21	9.40	8.18	14.84	9.07	7.89
PN-GJFC, $\alpha_{gf} = 0.45$	83.33	16.97	11.87	11.07	11.64	9.04	7.81
PN-GJFC, $\alpha_{gf} = 0.35$	101.14	14.60	11.64	10.37	10.37	9.03	7.71
PN-GJFC, $\alpha_{gf} = 0.25$	146.29	11.86	11.17	9.59	9.15	8.96	7.64

RMSE (multiplied by 10^6) based on simulated 1-min data for 5000 days, for the null model, $H = 0$, and three Persistent Noise alternatives from Section 4.1: a gradual jump (i) labeled "PN-GJ", a flash crash (ii) labeled "PN-FC", a gradual jump with an intermittent flash crash (iii) labeled "PN-GJFC". The threshold values for the TV and DV estimators are indicated by the respective $(C_\zeta^{TV}, C_\zeta^{DV})$ values.

Finally, as an additional benchmark, we also employ the realized volatility (RV) estimator defined by,

$$RV_t^n = \sum_{i=(t-1)n+1}^{tn} \left(\Delta_i^n Y\right)^2. \tag{33}$$

Overall, below we report the results for RV, TV, DV, and DV_{1-3} defined through Eqs. (33), (9), (11), and (16), respectively.

4.2. Simulation results

The estimation performance is assessed through the root-mean-square error (RMSE),

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{IV}_t - IV_t\right)^2},$$

where $\widehat{IV}_t$ denotes a given estimate of the integrated volatility IV_t for day t.

Table 3 presents the RMSEs for the selected estimators under the setting in Section 4.1. First, we note that all jump-robust estimators dramatically reduce the bias in estimating IV, stemming from the jump term and, not surprisingly, improve hugely on the RMSE of RV, confirming that the continuous variation can only be estimated with precision via jump-robust estimators.

Second, we compare the standard jump-robust threshold estimator TV to the corresponding jump- and persistent drift-robust DV estimator. Under the null model with $H = 0$, the RMSE of DV exceeds the one for TV by 23.3% and 18.0% under $(C_\zeta^{TV}, C_\zeta^{DV}) = (4, 4\sqrt{2})$ and $(C_\zeta^{TV}, C_\zeta^{DV}) = (3, 3\sqrt{2})$, respectively, roughly matching the result in Theorem 1 of $\sqrt{3/2} - 1 \approx 22.5\%$.[6] In contrast, under the persistent noise model, the RMSEs of TV increase quite steeply, while the performance of the DV estimator is less sensitive, resulting in a slower deterioration in the RMSE measure. For instance, under the persistent noise model for the flash crash pattern (PN-FC) in Table 3 with $\alpha = 0.35$, the RMSE of DV($3\sqrt{2}$) is 45.5% below that of TV(3).

Third, we inspect the performance of the DV_{1-3} estimator, designed to display robustness towards nontrivial values of the first three return autocorrelations. Under the null model, $H = 0$, the RMSE of DV_{1-3} exceeds the one of TV by 7.0% and 4.4% for the larger and smaller threshold values, respectively. The asymptotic result in Theorem 4 predicts a gap of $\sqrt{(2 + 1/3)/2} - 1 \approx 8.0\%$. Under the persistent drift alternatives, we find, in analogy to the results for DV above, that the DV_{1-3} estimator has a lower RMSE than TV, and the gap widens for more severe drift distortions. For instance, under the persistent noise model for the flash crash scenario (PN-FC) with $\alpha = 0.35$, the RMSE of TV is 19.77 compared to 8.98 for DV_{1-3}, and 16.61 versus 7.84, for the large and small threshold choice, respectively. In general, we find DV_{1-3} to also dominate DV across all relevant scenarios. Consequently, we rely primarily on the persistent drift-robust DV_{1-3} estimator for the volatility forecast performance evaluation below.

To further illustrate the relative performance of the TV and DV_{1-3} estimators during persistent-noise episodes, we depict estimates across 500 days for large and small threshold choices, respectively, in Figs. 4 and 5. The TV (blue) and DV_{1-3} (red) estimates are plotted against the true IV (black) value each day exclusively for the [0.4, 0.65] fraction of the daily unit time interval, featuring the extreme drift persistence, to highlight the consequences of the Itô semimartingale

[6] This is further corroborated through additional simulation results, available upon request.

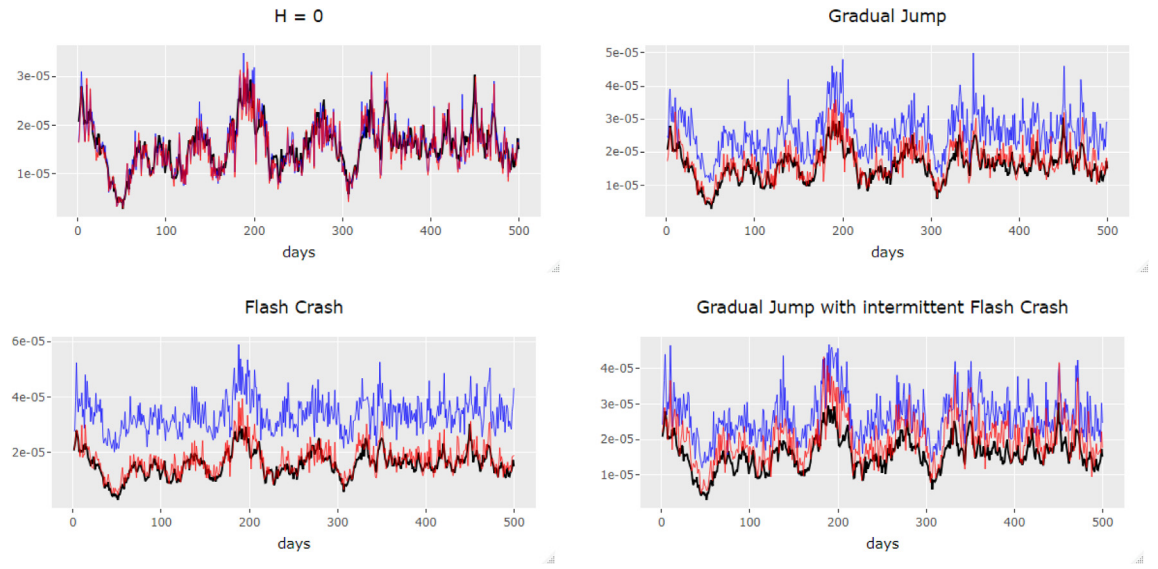

Fig. 4. IV estimation over the $[0.4, 0.65]$ fraction of daily unit time interval with simulated 1-min data for 500 days under null ($H = 0$), the PN-GJ case (with $\alpha_g = 0.35$), the PN-FC case (with $\alpha_f = 0.35$), and the PN-GJFC case (with $\alpha_{gf} = 0.35$) specified in Section 4.1. We plot the TV(4) in blue, $\mathrm{DV}_{1-3}(4\sqrt{2})$ in red, and the true IV in black. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

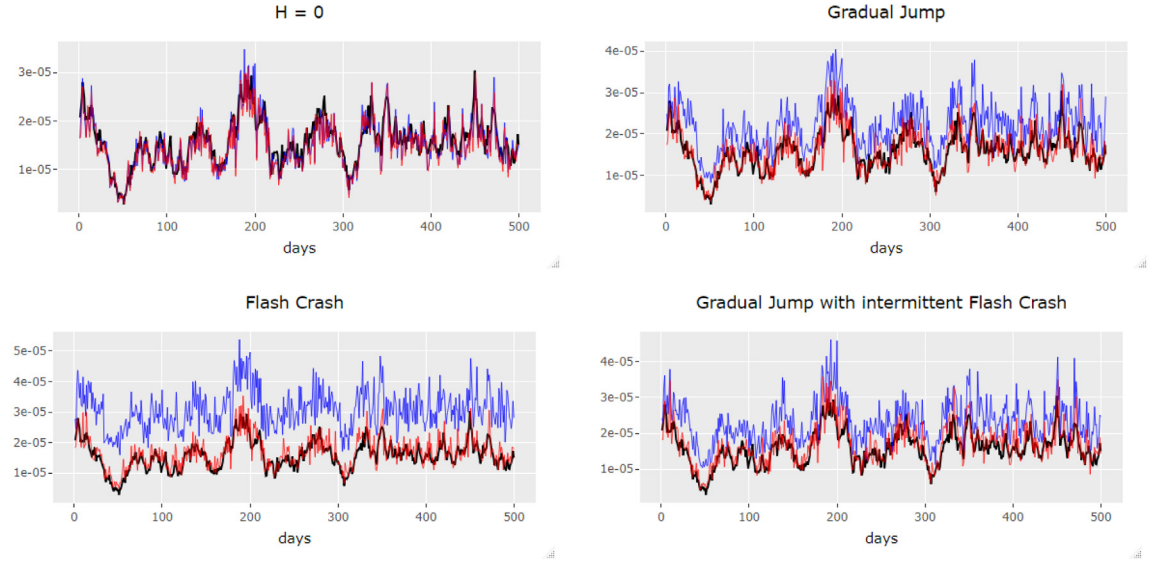

Fig. 5. IV estimation over the $[0.4, 0.65]$ fraction of daily unit time interval with simulated 1-min data for 500 days under null ($H = 0$), the PN-GJ case (with $\alpha_g = 0.35$), the PN-FC case (with $\alpha_f = 0.35$), and the PN-GJFC case (with $\alpha_{gf} = 0.35$) specified in Section 4.1. We plot TV(3) in blue, $\mathrm{DV}_{1-3}(3\sqrt{2})$ in red, and the true IV in black. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

violation. It is evident that, in both cases, the TV estimates are significantly upward biased relative to the true IV, while the DV_{1-3} estimator displays notably less bias under all the persistent noise scenarios investigated, including flash crash, gradual jump, and mixed gradual jump and flash crash episodes. We further note that the more aggressive threshold choice – $(C_\zeta^{TV}, C_\zeta^{DV})$ equal to $(3, 3\sqrt{2})$ versus $(4, 4\sqrt{2})$ – helps reduce the estimation bias for both estimators under the persistent noise model, rendering them the preferred choice under the alternative hypotheses. Additional insights on this point may be gauged from the empirical results in Section 5.4.

5. Empirical application

We first describe our high-frequency asset price data. We then characterize the type of events our formal test identifies as containing "persistent noise" episodes. Finally, we explore how our DV estimators deviate from, in particular, the related TV estimator, but also other realized return variation measures. Volatility is inherently latent, so we address the latter question indirectly, by exploring whether the DV estimator provides a competitive – perhaps even superior – input for standard volatility forecasting procedure based on the popular HAR framework. We do, in fact, find that noteworthy gains are obtained by adopting the DV-based forecasts.

5.1. Data

Our application focuses on the S&P 500 equity (SPX) index futures and the thirty constituents of the Dow Jones Industrial Average (DJIA) index. Our S&P futures sample covers tick-by-tick transactions from January 4, 1995, through July 31, 2018, for a total of 5878 days.[7] The high-frequency transaction records for individual stocks are from the NYSE/TAQ database, covering 3870 days from January 2, 2003 to July 2, 2018. To avoid idiosyncratic overnight and weekend effects, we only use data for the regular trading hours and eliminate days with reduced trading hours. Finally, we use an alternative S&P 500 volatility measure, obtained from the www.tailindex.com website, which is based on the theoretical work of Todorov (2019). This latter measure, labeled SV, provides nonparametric estimates of spot diffusive volatility, based on option prices at market close, but ranges only from January 2008 until the end of July 2018.

To mitigate market microstructure effects, we sample only at one-minute or lower frequencies, and we restrict our sample to transactions between 9:31:00 and 16:00:00 Eastern Time (ET) for the S&P 500, and 9:41:00–16:00:00 ET for individual stocks.

5.2. Persistent noise episodes

Our test for excessive return drift is based on the $T_t^n(k)$ statistic, Eqs. (22)–(23), with one-minute pre-averaged and winsorized returns and $k = 2$. The pre-averaging exploits five-second returns for non-overlapping one-minute intervals with (absolute) returns subsequently winsorized at 0.1%. The sampling frequency and length of the pre-averaging window are chosen with the aim of gaining power (thus sampling at the relatively high frequency of five seconds), while simultaneously guarding against standard types of microstructure noise (thus pre-averaging the raw returns to one minute). Finally, our choice of $k = 2$ aims at further guarding against regular microstructure noise, as the latter may introduce first-order return autocorrelation.

Fig. 6 displays smoothed daily values of the $T_t^n(2)$ statistics for the S&P 500 and quantiles for the cross-section of DJIA stocks. The values are averaged over a week-long moving window to facilitate clarity, although this renders a few daily peaks invisible.[8] The figure does not convey any clear association between the statistics for S&P 500 and individual stocks, apart from the pronounced outliers during the financial crisis. Comparing the set of largest positive daily outliers for the S&P and DJIA quantile statistics, covering well over 100 extreme observations, we find only four common extreme realizations outside of the 2008–2009 period. These are associated with the flash crash on May 6, 2010, the market turmoil following the downgrade of the U.S. public debt, August 9–10, 2011, and February 15, 2018, which saw very strong returns, following dramatic losses in the prior week, when the U.S. bond markets were rocked by worries over a revival of inflation and prospects for a tightening in monetary policy.

In addition, for Figs. 1 and 2, depicted in Section 1, we find that both of those dates produce a highly significant $T_t^n(2)$ statistic for S&P 500, with values of, respectively, 2.81 and 3.35. However, the corresponding observation from the 75th% quantile for the DJIA stocks is not an outlier for the latter date, while the former falls outside our DJIA sample period.[9]

The results demonstrate that indications of evolving frictions or uncertainty in the market, as captured by our test for an elevated return drift, do not necessarily – or even typically – coincide for the market index and underlying individual assets. To help shed light on this finding, we now compare the S&P 500 index and DJIA stocks along a few critical dimensions. First, we focus on the frequency of significant results for the $T_t^n(2)$-statistic over our sample. At the 1% critical level, we reject the absence of persistent noise on 2.45% of the trading days for S&P 500, while the corresponding rejection frequencies for individual stocks range from 1.72% for Pfizer Inc to 4.20% for JP Morgan-Chase.[10] That is, we observe a uniform, but relatively small excess proportion of rejections relative to test size. This may, indeed, signify that the number of incidents featuring an unusually strong drift is small, but it may also reflect a lack of power.

[7] We switch from the S&P 500 Futures to E-mini S&P 500 Futures from January 2004 on, since the liquidity of the latter dominates that of the former over that period.

[8] Under the null of no semimartingale violations, individual daily series of the statistic should mimic white noise, and quantiles should fluctuate randomly around fairly stable values. Despite deviations from this benchmark, the realizations are sufficiently close to random noise, rendering the display for daily values very "noisy".

[9] Note that the T-statistics in Table 2 refer to the significance of the gap between the TV and DV estimators, based on the result in Eq. (17). In contrast, we now test for "persistent noise", via the statistic defined in Eqs. (22)–(23).

[10] The analogous rejection frequencies for the 2.5% and 5% levels are 4.45% and 6.72% for the S&P 500 futures and ranges of 3.1%–7.1% and 5.5%–10.9%, respectively, for the 30 individual stocks.

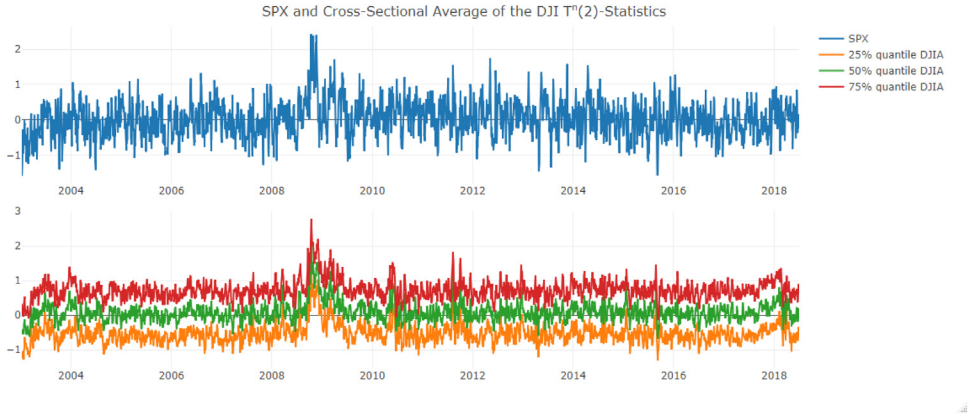

Fig. 6. Weekly moving averages of the $T_t^n(2)$-statistic for SPX (upper panel) and the cross-sectional 25%- (orange), 50%- (green), and 75%- (red) quantiles of the $T^n(2)$-statistics for the DJIA stocks (lower panel), from January 2, 2003 to July 2, 2018. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
First-order autocorrelation for daily $T_t^n(2)$ statistic.

Stock	Pre-crisis	Crisis-period	Post-crisis	Stock	Pre-crisis	Crisis-period	Post-crisis
SPX	0.1801	0.2154	0.0309				
AXP	0.0888	0.1970	0.0803	BA	0.0901	0.1198	0.0528
BAC	0.0869	0.3524	0.0445	CAT	0.0250	0.3369	0.0766
CSCO	0.0267	0.1632	0.0364	CVX	0.0888	0.1970	0.0803
DD	0.0737	0.1692	0.0349	DIS	0.0056	0.1918	0.0901
GE	0.0485	0.2664	0.0708	GS	0.0859	0.3087	0.0255
HD	0.0913	0.1621	0.0329	IBM	0.0654	0.1363	0.0081
INTC	0.0772	0.1759	0.0798	JNJ	0.0470	−0.0089	0.0112
JPM	0.0471	0.2997	0.0236	KO	0.0114	0.1269	0.0326
MCD	0.0464	0.0686	0.0530	MMM	0.0527	0.1128	0.0373
MRK	0.1050	0.1984	0.0465	MSFT	0.0632	0.1352	0.0553
NKE	0.0288	0.1610	0.0738	PFE	0.1200	0.0737	0.0245
PG	0.0696	0.1590	0.0003	T	0.1085	0.1244	0.0387
TRV	0.1023	0.2594	0.0541	UNH	0.0846	0.1675	0.0522
UTX	0.0253	0.0860	0.0242	VZ	0.0223	0.1684	0.0249
WMT	0.0345	0.0880	0.0598	XOM	0.0707	0.1746	0.0363

The $T_t^n(2)$ statistic is computed from one-minute pre-averaged and winsorized returns, as described in the main text. The sample period is from January 3, 1995 to July 31, 2018 for SPX and from January 2, 2003 to July 2, 2018 for DJIA. The crisis period covers January 2, 2008 to December 31, 2009.

On the one hand, we would expect the index futures to experience less problematic events than individual stocks, as the market is extremely liquid and viewed as quite efficient. On the other hand, the index has lower volatility than individual stocks, rendering an unusual drift easier to identify. These effects may partially cancel out, leaving the index with a similar average number of significant events as the DJIA stocks. Nonetheless, we would still expect a volatile market episode – whether it induces a significant excess drift episode or not – to also manifest itself among individual stocks. This prediction is consistent with the S&P index only having a limited number of extreme drift days during the financial crisis, while the individual stocks signal a more consistent elevation across the entire period in Fig. 6. Market jumps and turmoil may be associated with lagged responses across the underlying assets, leading to periods of persistent drift instead. Finally, of course, firm- or sector-specific news may induce disorderly trading conditions for individual stocks, that are not discernible at the overall market level.

Another feature of interest is the degree of serial correlation for days involving significant "persistent noise" episodes. Table 4 reports the first-order autocorrelation coefficient for the S&P 500 index and individual DJIA stocks, separately for the pre-crisis, financial crisis, and post-crisis periods. The correlations are almost uniformly positive, indicative of positive dependence in the strong drift episodes. Moreover, it is evident that it is, again almost uniformly, highest during the financial crisis. Finally, there is a tendency for a lower degree of persistence post-crisis, which is particularly striking for the market index.

In summary, persistent noise episodes for DJIA stocks are mildly correlated over time, display substantial cross-sectional correlation, as indicated by the quantiles in Fig. 6, and are weakly related to corresponding events for the market index. However, they are evident during extreme events, like the financial crisis, Treasury downgrade, and 2010 flash

crash. Our tests do not distinguish flash crashes from gradual jumps, but informal inspection indicates that flash crashes are rare, so most events seem related to gradual jumps or prolonged periods of price discovery.

5.3. Volatility forecasting setup and models

We next focus on the explanatory prowess of the DV estimators relative to existing popular forecast procedures within a standard setup. A commonly adopted approach is the HAR model of Corsi (2009). It captures the impact of recent volatility innovations as well as long-memory style features through a heterogeneous autoregressive specification, which has proven quite effective in out-of-sample forecast comparisons. It takes the form,

$$V_{t,t+h} = \beta_0 + \beta_D\, RV_t + \beta_W\, RV_{t-5,t} + \beta_M\, RV_{t-22,t} + \varepsilon_{t,t+h}\,, \tag{34}$$

where $RV_{t,t+h}$ denotes the multi-period normalized realized return variation,

$$RV_{t,t+h} = h^{-1}\,(\,RV_{t+1} + RV_{t+2} + \cdots + RV_{t+h}\,)\,, \quad h = 1, 2, \ldots\,, \tag{35}$$

while $V_{t,t+h}$ is the ex-post realized value of the volatility proxy we seek to forecast over $[t+1, t+h]$. Often, $V_{t,t+h} = RV_{t,t+h}$, but we explore alternative measures as well below. Mainly, we also forecast future DV and the option-based SV. These two volatility measures, unlike RV, are jump robust and should be less sensitive to the presence of persistent noise episodes. For DV, this follows from our theoretical results above, while for SV it follows, if the persistent noise contaminating the underlying asset price is not present in the option market at the end of the trading day, when the option prices used to construct SV are recorded.

The choice of lagged RV regressors covering the previous trading day, week (5 days) and month (22 days), is standard. Given these regressors, we explore forecasts for $V_{t,t+h}$ over the coming trading day ($h = 1$), week ($h = 5$), and month ($h = 22$) in our predictive analysis.

The basic HAR-RV model has subsequently been modified in numerous ways. In particular, Andersen et al. (2007) split RV into continuous and jump components. They found only the continuous component to provide significant predictive power, motivating the HAR-C variant,

$$V_{t,t+h} = \beta_0 + \beta_D\, C_t + \beta_W\, C_{t-5,t} + \beta_M\, C_{t-22,t} + \varepsilon_{t,t+h}\,, \tag{36}$$

where the continuous return variation measures are normalized as above for RV, $C_{t,t+h} = h^{-1}\,(C_{t+1} + C_{t+2} + \cdots + C_{t+h})$.

A common estimator for the continuous component C is the TV estimator, defined in Eq. (9), which truncates for jumps using the threshold (31). Henceforth, we denote this model HAR-TV(C_ζ^{TV}). Similarly, if our DV_{1-m} estimator, as defined in Eq. (16), serves as the basis for the predictor variables, then the associated model is labeled HAR-DV(C_ζ^{DV}).

Another commonly adopted candidate for estimating the continuous part of the quadratic return variation is the Bipower Variation (BV) of Barndorff-Nielsen and Shephard (2006),

$$BV_t = \frac{\pi}{2}\left(\frac{n}{n-1}\right)\sum_{i=(t-1)n+1}^{tn} |\Delta Y_i||\Delta Y_{i+1}|\,. \tag{37}$$

Exploiting the corresponding BV estimators as regressor variables leads to the HAR-BV model,

$$V_{t,t+h} = \beta_0 + \beta_D\, BV_t + \beta_W\, BV_{t-5,t} + \beta_M\, BV_{t-22,t} + \varepsilon_{t,t+h}\,. \tag{38}$$

Finally, we consider the HAR modification of Bollerslev et al. (2016), designed to reduce the impact of measurement errors—the gap between the latent integrated volatility, IV, and its proxy, RV—by allowing model parameters to vary with estimates for the size of the measurement error for RV. They achieve the best HARQ performance from the following specification,

$$V_{t,t+h} = \beta_0 + \left(\beta_D + \beta_Q\sqrt{RQ_t}\right) RV_t + \beta_W RV_{t-5,t} + \beta_M RV_{t-22,t} + \varepsilon_{t,t+h}\,, \tag{39}$$

where the Realized Quarticity (RQ), defined as $RQ_t = \frac{n}{3}\sum_{i=1}^{n}\Delta Y_i^4$, provides the requisite estimate for the magnitude of the measurement error. For this HARQ model, we follow their lead in also applying the "insanity filter" (IF), introduced by Swanson and White (1997), as the model occasionally produces implausibly large or small forecasts. Specifically, the IF algorithm replaces any forecast falling outside the range of values of the target variable observed during the estimation period by the unconditional mean of the variable over that period.

In summary, our benchmark models include the HAR-RV, HAR-C, HAR-BV, and HARQ models, specified in Eqs. (34), (36), (38), and (39), respectively. The limited choice of alternatives reflects our focus on the information content of the DV estimator relative to commonly adopted measures, rather than the development of a new model for optimal volatility prediction. As such, we avoid extensive data exploration (and mining), yet still illustrate the potential gains from our new volatility measure. We consider HARQ due to its recent popularity, but relegate many of the associated results to the Supplementary Appendix, as HARQ introduces a number of new features, that we purposely exclude from the analysis of the other models. Nonetheless, the inclusion of HARQ demonstrates that our robust volatility proxy can facilitate forecast procedures that compare favorably, on many dimensions, to even very recently developed techniques.

Table 5
Estimated MSE $\times 10^8$ of RV forecasts.

Asset	Frequency			
	1-min	2-min	3-min	5-min
S&P 500	3.42	3.34	3.31	3.67
DJIA (average)	23.8	23.6	22.8	24.3

The sample MSE for S&P 500 futures and DJIA stocks exploiting the procedure in Theorem 3 of Bandi and Russell (2008). The realized measures are obtained from jump-robust returns exploiting the same truncation threshold as TV in Eq. (31), $3\,\hat{\sigma}_{t-1}^{med}\,\Delta_n^{1/2}$.

5.4. Empirical forecasting results

The out-of-sample forecast performance is assessed through two widely used loss functions, namely the mean square error (MSE),

$$\mathrm{MSE}(V_{t,t+h}, \widehat{V}_{t,t+h}) = \left(V_{t,t+h} - \widehat{V}_{t,t+h}\right)^2 , \tag{40}$$

and the quasi-likelihood (QLIKE) loss function,

$$\mathrm{QLIKE}(V_{t,t+h}, \widehat{V}_{t,t+h}) = \frac{V_{t,t+h}}{\widehat{V}_{t,t+h}} - \log \frac{V_{t,t+h}}{\widehat{V}_{t,t+h}} - 1 , \tag{41}$$

where $\widehat{V}_t$ denotes the forecast of V for trading day t, and $\widehat{V}_{t,t+h} = h^{-1}\left(\widehat{V}_{t+1} + \cdots + \widehat{V}_{t+h}\right)$. The MSE loss penalizes outliers heavily, whether due to large (unexpected) volatility innovations or excessively large forecasts, while the QLIKE loss reflects the relative forecast error.[11]

To determine a suitable sampling frequency for the high-frequency returns, when constructing realized variation measures—balancing the trade-off between estimation efficiency from an increasing number of observations and the bias induced by microstructure noise—we initially obtain the MSE of intraday realized variance forecasts for different frequencies based on jump-robust returns, using the same truncation as for TV previously, i.e., the threshold is $C_\zeta^{TV} = 3$.[12] In particular, MSE results for the S&P 500 futures and DJIA stocks for 1-, 2-, 3-, and 5-min sampling are reported in Table 5. In both cases, the 3-min returns provide the minimal MSE, so we base the predictive analysis below on variation measures obtained from 3-min sampling.

5.4.1. S&P 500 index

We first consider volatility forecasts for the S&P 500 futures, for which we have the longest sample and access to high-quality option-based spot variance measures over the recent years. Estimation and forecasting are carried out recursively. For in-sample estimation, we rely on OLS to generate both a rolling window forecast using the prior 1000 trading days and an increasing window forecast using all prior observations, starting from an initial set of 1000 trading days.

We start with one-day-ahead intraday RV forecasts. Table 6 reports the MSE and QLIKE values, using rolling (RW) or increasing windows (IW), for HAR models exploiting RV, BV, TV, DV, or HARQ regressors. For TV and DV, we report results for the truncation levels 3 and $3\sqrt{2}$, respectively,[13] while we provide results for HARQ with and without implementation of the insanity filter (IF).[14] The first striking result is the uniformly poor performance of the HAR-RV and HAR-BV forecasts. For every metric and scenario, they fare much worse than the corresponding TV- and DV-forecasts. Evidently, the robust truncation implemented through the latter estimators are important to curtail the impact of outliers in the regressors, while BV clearly fails in this regard. Second, the HARQ forecasts are competitive on the MSE criterion. For the full sample results in the top panel, HARQ with associated IF generates the second-lowest MSE for the RW-based forecasts and the overall lowest value for the IW-based forecasts. In contrast, the overall ranking of HARQ according to the QLIKE criterion is less stellar, although still competitive with the TV- and DV-forecasts. Third, the middle panel of Table 6 refers to the performance only for trading days, when our test for the presence of persistent noise is significant. These days represent highly turbulent conditions, as the MSEs of the forecasts are orders of magnitude larger than those for the remaining (non-persistent noise) days, reported in the bottom panel. We note that the overall good MSE performance of HARQ stems from these fairly rare episodes, where it improves notably on the other models. This advantage is lessened for the QLIKE criterion, where the TV- and DV-models generally perform better. Finally, we also note the importance of the IF in this regard. The HARQ (no IF) generates huge outliers for the QLIKE, implying an erratic behavior in the absence of the filter.

11 Patton (2011) demonstrates that the MSE and QLIKE loss functions have desirable properties in this context.

12 This follows the original approach of Bandi and Russell (2008).

13 The supplementary Appendix provides evidence for alternative choices of thresholds. The qualitative results are identical, but the alternative choices usually produce marginally worse outcomes.

14 Recall, HARQ has a time-varying coefficient on the one-day lagged RV regressor to account for the corresponding measurement error, and it invokes a filter, if the basic HARQ forecast otherwise is unusually large or small.

Table 6
Daily out-of-sample RV forecasts for S&P 500.

Sample	Procedure	RW		IW	
		MSE × 10^6	QLIKE	MSE × 10^6	QLIKE
Full sample	HAR-RV	0.6561	0.1557	0.5034	0.1555
	HAR-BV	0.7041	0.1563	0.5666	0.1552
	HAR-TV (3)	0.4904	0.1390	0.4361	0.1318
	HAR-DV_{1-3} ($3\sqrt{2}$)	0.4702	0.1387	0.4266	0.1304
	HARQ (IF)	0.4891	0.1404	0.4209	0.1338
	HARQ (no IF)	0.4888	0.1573	0.4210	0.1578
Persistent Noise days	HAR-RV	20.4287	0.2153	16.5482	0.2139
	HAR-BV	22.8546	0.2129	20.2140	0.2072
	HAR-TV (3)	16.3814	0.1784	14.1436	0.1769
	HAR-DV_{1-3} ($3\sqrt{2}$)	15.4031	0.1868	13.4505	0.1867
	HARQ (IF)	13.1762	0.1852	12.7023	0.1992
	HARQ (no IF)	13.1809	0.3352	12.7066	1.2313
Remaining days	HAR-RV	0.3333	0.1547	0.2414	0.1546
	HAR-BV	0.3425	0.1554	0.2458	0.1543
	HAR-TV (3)	0.2309	0.1383	0.2123	0.1311
	HAR-DV_{1-3} ($3\sqrt{2}$)	0.2264	0.1379	0.2140	0.1295
	HARQ (IF)	0.2819	0.1397	0.2204	0.1327
	HARQ (no IF)	0.2816	0.1544	0.2204	0.1402

The forecast results for January 3, 1995 to July 31, 2018 (upper panel) are decomposed by Persistent-Noise days, whenever $T_t^n(2) > 2.325$ (middle panel), and remaining days, when $T_t^n(2) < 2.325$ (bottom panel). The jump-truncation thresholds, C_ζ^{TV} and C_ζ^{DV}, are provided in the parentheses for the TV and DV statistics, while imposition of the insanity filter (or not) is indicated by IF for HARQ.

Table 7
Diebold–Mariano p-values for DV-based S&P 500 volatility forecasts.

	HAR-DV_{1-3} against TV				HAR-DV_{1-3} against HARQ			
	RW		IW		RW		IW	
	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
Forecasting RV								
Daily	0.1976	0.3933	0.2083	0.0477	0.2559	0.2490	0.7239	0.0584
Weekly	0.3644	0.0513	0.4803	0.0156	0.6040	0.0740	0.3443	0.0763
Monthly	0.2195	0.0007	0.2034	0.0002	0.5553	0.0618	0.3844	0.0642
Forecasting DV_{1-3}								
Daily	0.1597	0.0196	0.1694	0.0000	0.9004	0.1503	0.8916	0.7684
Weekly	0.4436	0.0005	0.4801	0.0000	0.1331	0.0654	0.2139	0.0499
Monthly	0.2136	0.0000	0.1806	0.0000	0.2909	0.0689	0.2442	0.0676
Forecasting SV								
Daily	0.0004	0.0000	0.0000	0.0000	0.0131	0.0001	0.0000	0.0000
Weekly	0.0000	0.0000	0.0000	0.0000	0.0046	0.0004	0.0000	0.0000
Monthly	0.0000	0.0008	0.0000	0.0000	0.0503	0.0000	0.0001	0.0000

Null hypothesis, H_0: HAR-DV_{1-3} is less accurate than its competitor. The predictions cover January 3, 1995 to July 31, 2018 for RV and DV_{1-3}, and January 2, 2008 to July 31, 2018 for SV. The threshold values for the TV and DV equal $(C_\zeta^{TV}, C_\zeta^{DV}) = (3, 3\sqrt{2})$.

Of course, we do not know to whether the discrepancies in Table 6 are statistically significant. However, it turns out that this setting is close to a best-case scenario for HARQ, as the HARQ procedure is less effective for longer forecast horizons and alternative ex-post volatility proxies. For illustration, we review some Diebold–Mariano test results for equal predictive performance across our full S&P 500 sample, using RV-, DV-, and SV-proxies for ex-post diffusive volatility.[15] The SV series is of particular interest, as it is constructed exclusively from option prices, and thus is void of the specific form for noise structure present in the high-frequency asset prices.[16]

Table 7 reports p-values for the hypothesis that forecasts based on the DV-regressors perform worse than those based on an alternative procedure. That is, large p-values favor the alternative forecast procedure, while low p-values favor the DV-based forecasts. The first line of the table corresponds to the evidence for RV forecasts in Table 6. Most of the entries in the first row are below 50%, with none larger than 72.5%, so there are no indications of subpar DV performance. The

[15] We follow Diebold (2015, Section 2.1) and adjust the DM statistic with Newey–West Heteroskedasticity and Autocorrelation Corrected (HAC) standard errors.

[16] To render the series compatible with the other intraday volatility measures, it is scaled so that its average value coincides with that of the ex-post TV measure.

Table 8
Monthly out-of-sample DV and SV forecasts for S&P 500.

	DV_{1-3}				SV			
	RW		IW		RW		IW	
	MSE $\times 10^6$	QLIKE	MSE $\times 10^6$	QLIKE	MSE $\times 10^7$	QLIKE	MSE $\times 10^7$	QLIKE
Full sample								
HAR-RV	0.1320	0.1665	0.1010	0.1570	0.1033	0.1708	0.1953	0.3030
HAR-BV	0.1467	0.1661	0.1144	0.1562	0.1099	0.1742	0.2138	0.3177
HAR-TV	0.1045	0.1575	0.0966	0.1466	0.0917	0.1591	0.1745	0.2828
HAR-DV_{1-3}	0.1018	0.1557	0.0949	0.1430	0.0884	0.1573	0.1610	0.2693
Persistent Noise days								
HAR-RV	2.3201	0.0978	1.1896	0.1170	0.0736	0.1203	0.1843	0.2566
HAR-BV	2.8773	0.0955	1.9189	0.1131	0.0779	0.1231	0.2080	0.2721
HAR-TV	1.3165	0.0917	0.8782	0.1171	0.0692	0.1124	0.1612	0.2359
HAR-DV_{1-3}	1.2428	0.0901	0.8069	0.1140	0.0672	0.1114	0.1482	0.2236
Remaining days								
HAR-RV	0.0962	0.1676	0.0831	0.1577	0.1036	0.1714	0.1954	0.3036
HAR-BV	0.1019	0.1672	0.0848	0.1569	0.1103	0.1748	0.2139	0.3182
HAR-TV	0.0846	0.1586	0.0838	0.1470	0.0919	0.1596	0.1746	0.2834
HAR-DV_{1-3}	0.0831	0.1568	0.0832	0.1435	0.0886	0.1579	0.1611	0.2698

The forecasting results for all days—from January 2, 2003 to July 2, 2018 for DV_{1-3} and from January 2, 2008 to July 31, 2018 for SV—(upper panel) are decomposed by Persistent-Noise days whenever $T_t^n(2) > 2.325$ (middle panel) and remaining days when $T_t^n(2) < 2.325$ (bottom panel), where $T_t^n(k)$ is defined in Eq. (26). The thresholds for the TV and DV estimators are $(C_\zeta^{TV}, C_\zeta^{DV}) = (3, 3\sqrt{2})$. We scale SV to have the same average value of TV.

evidence in favor of DV is also limited, but stronger: compared to the QLIKE values obtained from the TV- and HARQ-IW procedures, the test statistic indicates rejection at the 10% level (0.0477 and 0.0584, respectively). This is noteworthy, as it is consistent with the stronger evidence emerging for the longer forecast horizons. Across the eight combinations of RW/IW- and TV/HARQ-based forecasts, they all attain p-values below 10% for QLIKE.

Moreover, as we consider alternative ex-post volatility proxies, these results are not only confirmed, but strengthened. Specifically, for HARQ, there is no indication of underperformance for the daily DV-forecasts in the middle panel, but issues arise at the weekly and monthly horizons where, across the eight RW/IW and MSE/QLIKE DV-forecast scenarios based on HARQ, we observe p-values from below 5% to only 30%. Finally, for the SV forecasts in the bottom panel, the evidence is uniformly, and strongly, in favor of DV relative to HARQ.

We conclude that HARQ is unlikely to improve on our DV-based forecasts in general, even if it may point towards opportunities for improving short-term predictions. As noted, this is due to the introduction of auxiliary features, that we cannot explore in earnest within the confines of the current study. Hence, we defer most of the HARQ evidence to the Supplementary Appendix, but note that the findings generally are qualitatively similar, even for the individual DJIA stocks.

Consequently, in the sequel we focus on forecasts based on DV- and corresponding RV-, BV-, and TV-based procedures. RV and BV measures signify the dominant choices in the extant literature, while our DV-approach represents an adaption of the TV-estimator to generate robustness against a specific type of deviation from the semimartingale paradigm.

To provide a more comprehensive overview of our results, we conclude this section by illustrating the performance of our DV-based forecasts in a polar opposite case from Table 6. We now focus on the two other ex-post volatility proxies, DV and SV, and the longer one-month forecast horizon. Table 8 displays results for the full sample, the days with (significant) indications of a persistent mean drift, and the remaining days.[17] The uniformly superior forecast performance of DV relative to TV is noteworthy. As before, we also find the RV- and BV-based forecasts to be much worse than those relying on the TV- and DV-statistics. Finally, we note that the MSE losses are lower for the SV proxies. This is natural, as the ex-post DV-measure provides a noisy measure of the volatility realization, while the SV value refers directly to a concurrent (option-implied) estimate of the underlying volatility level.[18] This suggests that SV provides a less noisy benchmark for assessment of forecast performance. In light of this observation, the extremely significant evidence in favor of DV- versus TV-forecasts, confirmed for this scenario in Table 7, is particularly striking, with p-values almost identically equal to zero.

Conceptually, the difference between HAR-TV and HAR-DV_{1-3} models is that the former provides robustness versus jumps, while the latter also controls for distortions induced by pockets of extreme return persistence, such as gradual jumps and flash crash scenarios. Our findings suggest that the DV-procedure is preferable, even for trading periods when the test for persistent noise is insignificant. This is likely due to the presence of violations that our test has insufficient

[17] The TV- and DV-values in the bottom rows of the top panel in Table 8 reflect the same underlying forecasts and volatility proxies that generate the first four p-values in the bottom row of the middle and bottom panel in Table 7.

[18] The large MSE discrepancies in Table 8 between forecasts of the DV and SV proxies also reflect the different sample periods, as the SV forecasts are initiated only during the great financial crisis of 2008–2009. However, we confirm that the difference remains substantial, even if we generate the forecasts over the identical time period.

Table 9
Aggregate weekly out-of-sample DV forecasts for individual DJIA stocks.

	Average				Median			
	RW		IW		RW		IW	
	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
Full sample								
HAR-RV	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
HAR-BV	1.0455	1.0152	1.0441	1.0316	1.0143	1.0170	1.0146	1.0227
HAR-TV	0.8169	0.8872	0.8270	0.8485	0.8310	0.9035	0.8552	0.8431
HAR-DV_{1-3}	0.8167	0.8771	0.8265	0.8277	0.8044	0.8842	0.8243	0.8236
Persistent Noise days								
HAR-RV	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
HAR-BV	1.0783	1.0073	1.0868	1.0305	1.0237	1.0099	1.0225	1.0169
HAR-TV	0.8688	0.9151	0.8547	0.8603	0.8529	0.8988	0.8521	0.8544
HAR-DV_{1-3}	0.8258	0.9085	0.8213	0.8455	0.8267	0.8967	0.8380	0.8563
Remaining days								
HAR-RV	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
HAR-BV	1.0181	1.0155	1.0164	1.0308	1.0123	1.0166	1.0151	1.0226
HAR-TV	0.8691	0.8862	0.8764	0.8494	0.9444	0.9018	0.9329	0.8488
HAR-DV_{1-3}	0.8773	0.8760	0.8834	0.8284	0.9466	0.8814	0.9472	0.8286

The forecast results for the future DV_{1-3} measure across all days from January 2, 2003 to July 2, 2018 (upper panel) are decomposed by Persistent-Noise days, whenever $T_t^n(2) > 2.325$ (middle panel), and remaining days, when $T_t^n(2) < 2.325$ (bottom panel). $T_t^n(k)$ is defined in Eq. (26). The threshold values for the TV and DV estimators are $(C_\zeta^{TV}, C_\zeta^{DV}) = (3, 3\sqrt{2})$.

power to identify. Our DV-procedure retains its robustness irrespectively, and this may more than compensate for minor efficiency losses incurred during regular periods.

5.4.2. DJIA stocks

The DJIA results are harder to convey in a concise manner. We report only a few representative findings, while deferring most results to the Supplementary Appendix. Table 9 provides a first overview. For readability, the loss functions are reported as ratios relative to the corresponding value for HAR-RV. The table refers to weekly forecasts based on DV_{1-3}. These results constitute a middle ground between those for daily and monthly forecasts – the relative performance of HAR-DV_{1-3} is slightly worse at the daily and slightly better at the monthly horizon.

We note that all values for TV- and DV-based forecasts indicate dramatic improvements relative to RV. Moreover, we again confirm that the BV-procedure performs poorly. As for the S&P 500 index, we also find the DV-forecasts to outperform TV for the full sample. The improvement is seen to stem primarily from the relatively few trading days with significant indications of persistent noise episodes, while the performance for the remaining days is comparable.

For a more detailed look, involving each individual stock, Table 10 reports p-values of Diebold–Mariano tests for the relative performance of DV- versus TV-based forecasts for future DV realizations.[19] Informally, we summarize the findings by counting the number of p-values that fall below versus above a given threshold for each criterion. For MSE obtained via RW (IW) estimation, the count is 5-0 (5-1) for p-values below 10% versus above 90%, while the corresponding numbers for the QLIKE are 17-0 (23-2). Thus, while the evidence is largely non-discriminatory for the MSE criterion, the QLIKE results point towards significant outperformance for DV-based forecasts. The results for the daily and monthly DV forecasts are similar. The corresponding test for DV- versus HARQ-forecasts also generate the same conclusions, except for weak evidence in favor of HARQ for daily forecasts assessed via MSE.[20] Finally, the identical analysis for RV forecasts also yields qualitatively similar results, with the DV forecasts being strongly favored according to the QLIKE criterion in almost all cases. The main deviation relative to Table 10 is that TV performs on par with DV at the one-day horizon, and TV generally improves its relative performance for RV forecasts according to the MSE criterion.

6. Conclusion

We propose a new family of jump-robust integrated volatility (IV) measures, labeled differenced-return volatility (DV) estimators, based on averaged truncated return differentials across low order lags. These DV estimators significantly reduce the bias in IV estimation caused by pockets of extreme return persistence, induced by episodic gradual jump

[19] Because both forecasts are based on jump-robust measures and ex-post RV measures are inherently noisy, the DV-forecasts constitute a natural benchmark. Nonetheless, we note that the findings for RV-forecast are qualitatively similar, as can be confirmed from the results tabulated in the Supplementary Appendix. Of course, an analysis based on SV would be preferable, but option-implied spot volatility measures are not available for the individual stocks.

[20] For example, in analogy to Table 10, the Diebold–Mariano tests regarding weekly DV predictions for DV- versus HARQ-based forecasts generate the following number of p-values below 10% (in favor of DV) and above 90% (favoring HARQ) for the RW (IW) procedure, MSE: 5-2 (6-1), QLIKE: 15-0 (18-1). The same comparison at the daily level, where HARQ performs relatively better, yields MSE: 0-5 (0-3), QLIKE: 16-1 (14-4).

Table 10
Diebold–Mariano p-values for weekly DV_{1-3} forecasts on DJIA stocks.

	RW		IW			RW		IW	
	MSE	QLIKE	MSE	QLIKE		MSE	QLIKE	MSE	QLIKE
AXP	0.2473	0.0005	0.2512	0.0000	KO	0.5719	0.5457	0.5727	0.0630
BA	0.6330	0.6282	0.6095	0.6866	MCD	0.2902	0.0190	0.3303	0.0003
BAC	0.7683	0.6314	0.7635	0.8642	MMM	0.5227	0.3173	0.5395	0.0019
CAT	0.8715	0.7161	0.8634	0.9195	MRK	0.8916	0.0712	0.9155	0.0151
CSCO	0.6236	0.0028	0.6345	0.0000	MSFT	0.3964	0.0650	0.4757	0.0345
CVX	0.2844	0.0054	0.2882	0.0028	NKE	0.1988	0.0333	0.2037	0.0234
DD	0.8262	0.6923	0.8282	0.3554	PFE	0.7101	0.5142	0.6909	0.7715
DIS	0.6143	0.1313	0.5032	0.1165	PG	0.6883	0.0556	0.6600	0.0140
GE	0.0762	0.2320	0.0804	0.0151	T	0.0936	0.0000	0.0887	0.0000
GS	0.1653	0.7950	0.1653	0.0000	TRV	0.8480	0.8050	0.8485	1.0000
HD	0.0944	0.0004	0.0953	0.0000	UNH	0.1312	0.0533	0.1331	0.0000
IBM	0.4019	0.0251	0.3555	0.0439	UTX	0.2361	0.0000	0.2199	0.0000
INTC	0.2168	0.4259	0.2144	0.0063	VZ	0.0769	0.0000	0.0790	0.0000
JNJ	0.7545	0.1497	0.7721	0.0217	WMT	0.1879	0.0000	0.1968	0.0000
JPM	0.0712	0.0000	0.0743	0.0000	XOM	0.1427	0.0000	0.1440	0.0000

H_0: HAR-DV_{1-3} is less accurate than HAR-TV with threshold values $(C_\zeta^{TV}, C_\zeta^{DV}) = (3, 3\sqrt{2})$. The forecasts are for the future weekly DV_{1-3} measures. The forecast period is from January 2, 2003 to July 2, 2018.

and flash crash incidents. We capture such scenarios through either the drift burst model of Christensen et al. (2020) or a related new persistent noise model, but our inference procedure is robust to a much larger set of (persistent noise) deviations from the standard framework. We show, both theoretically and via simulation, that our DV estimators significantly reduce the IV estimation biases suffered by the standard truncation volatility (TV) estimator, while they are subject to only minor efficiency losses under the usual Itô semimartingale assumption. Capitalizing on this feature, we use DV as the basis for predictive volatility regressions within a Heterogeneous Autoregressive (HAR) type model. This empirical application suggests that the resulting HAR-DV model can provide substantial improvements relative to commonly adopted forecast models in the literature under both MSE and QLIKE losses for the S&P 500 index futures and Dow Jones Individual stocks.

Appendix A

A.1. Proofs

Proof of Theorem 1. The theorem is proved in a similar way to the proof of Theorem 13.2.4 in Jacod and Protter (2012). First, as in the proof of that theorem, we can show that the difference between *TV* and *DV* and their counterparts formed by the continuous part of the process X and without truncation is asymptotically negligible. Second, exactly as Lemma 5.4.9 in Jacod and Protter (2012), we can show that the difference between the latter statistics and

$$\sqrt{n}\sum_{i=(t-1)n+2}^{tn}\begin{pmatrix}\sigma^2_{(i-2)\Delta_n}(\Delta_i^n W)^2\\ \frac{1}{2}\sigma^2_{(i-2)\Delta_n}(\Delta_i^n W-\Delta_{i-1}^n W)^2\end{pmatrix}, \tag{42}$$

is asymptotically negligible. Hence, we are left with establishing the CLT for the above process. The latter result can be shown as in the proof of Lemma 5.4.10 in Jacod and Protter (2012) and an application of big block–small block idea (skipping one increment to break the time series correlation in the summands). □

Proof of Theorem 2. In the proof, we will assume that $\tau_{db} = 0$ and we will focus attention on the asymptotic properties of TV_1^n and DV_1^n. The more general case $\tau_{db} > 0$ can be handled similarly.

Convergence result for TV

Direct calculation shows

$$\Delta_i^n H = c_r[i^{1-\alpha} - (i-1)^{1-\alpha}]\Delta_n^{1-\alpha}.$$

The function $f(x) = x^{1-\alpha} - (x-1)^{1-\alpha}$ is decreasing in the interval $[1, \infty)$ with $f(1) = 1$ and $\lim_{x\to\infty} f(x) = 0$. Therefore, in the case $\alpha > 1 - \varpi$, we have that for Δ_n sufficiently small, $f(x) = \frac{\chi}{c_r}\Delta_n^{\varpi+\alpha-1}$ has a unique solution denoted with $K_n(\chi)$ for arbitrary $\chi > 0$ and such that $K_n \asymp \Delta_n^{\frac{1-\varpi-\alpha}{\alpha}}$. With this result, we can make the following decomposition

$$TV_1^n = \sum_{i=1}^{K_n(\zeta/2)}(\Delta_i^n Y)^2 \mathbb{1}_{\{|\Delta_i^n Y|\le\zeta\Delta_n^\varpi\}} + \sum_{K_n(\zeta/2)+1}^{n}(\Delta_i^n Y)^2\mathbb{1}_{\{|\Delta_i^n Y|\le\zeta\Delta_n^\varpi\}} := A_1^n + A_2^n, \tag{43}$$

and we note that for Δ_n small enough, we have $1 < K_n(\zeta/2) < n$. We look at each of the three terms separately. Starting with A_1^n, we have

$$\left| A_1^n - \int_0^{K_n(\zeta/2)\Delta_n} \sigma_s^2 ds \right| \leq \zeta^2 \Delta_n^{2\varpi} K_n(\zeta/2) + \int_0^{K_n(\zeta/2)\Delta_n} \sigma_s^2 ds = O_p\left(\Delta_n^{2\varpi + \frac{1-\varpi-\alpha}{\alpha}}\right). \tag{44}$$

We are left with A_2^n. First, by noting that $|\Delta_i^n Y| \leq \zeta \Delta_n^{\varpi}$ implies $|\Delta_i^n X| \leq \frac{3}{2}\zeta \Delta_n^{\varpi}$ for $i > K_n(\zeta/2)$, we have

$$\sum_{K_n(\zeta/2)+1}^{n} (\Delta_i^n X)^2 \mathbb{1}_{\{|\Delta_i^n Y| \leq \zeta \Delta_n^{\varpi}\}} - \int_{K_n(\zeta/2)\Delta_n}^{1} \sigma_s^2 ds = O_p(\sqrt{\Delta_n}). \tag{45}$$

Next, using Taylor expansion, we have

$$\begin{aligned} \sum_{i=K_n(\zeta/2)+1}^{n} (\Delta_i^n H)^2 &\leq (1-\alpha)^2 c_r^2 \Delta_n^{2(1-\alpha)} \sum_{i=K_n(\zeta/2)+1}^{n} i^{-2\alpha} \\ &\sim \Delta_n^{2(1-\alpha)} K_n(\zeta/2)^{1-2\alpha} \int_1^{n/K_n(\zeta/2)} s^{-2\alpha} ds \\ &\sim \Delta_n^{2\varpi + \frac{1-\varpi-\alpha}{\alpha}}. \end{aligned} \tag{46}$$

Using this result and Burkholder–Davis–Gundy inequality, we have

$$\sum_{i=K_n(\zeta/2)+1}^{n} \Delta_i^n X \Delta_i^n H \mathbb{1}_{\{|\Delta_i^n Y| \leq \zeta \Delta_n^{\varpi}\}} = O_p\left(\sqrt{\Delta_n}\Delta_n^{\varpi + \frac{1-\varpi-\alpha}{2\alpha}}\right) = o_p(\sqrt{\Delta_n}). \tag{47}$$

Altogether, we have

$$A_2^n - \int_{K_n(\zeta/2)\Delta_n}^{1} \sigma_s^2 ds = O_p\left(\Delta_n^{2\varpi + \frac{1-\varpi-\alpha}{\alpha}} \bigvee \sqrt{\Delta_n}\right). \tag{48}$$

The results for A_1^n and A_2^n imply the result about *TV* in the theorem.

Convergence result for DV

Direct calculation shows

$$\Delta_i^n H - \Delta_{i-1}^n H = c_r[i^{1-\alpha} - 2(i-1)^{1-\alpha} + (i-2)^{1-\alpha}]\Delta_n^{1-\alpha}.$$

The function $g(x) = x^{1-\alpha} - 2(x-1)^{1-\alpha} + (x-2)^{1-\alpha}$ is increasing in the interval $[2, \infty)$ with $g(2) = 2^{1-\alpha} - 2$ and $\lim_{x\to\infty} g(x) = 0$. Therefore, in the case $\alpha > 1 - \varpi$, we have that for Δ_n sufficiently small, $g(x) = \frac{\chi}{c_r}\Delta_n^{\varpi+\alpha-1}$ has a unique solution denoted with $K_n(\chi)$ for arbitrary $\chi > 0$ and such that $K_n \asymp \Delta_n^{\frac{1-\varpi-\alpha}{1+\alpha}}$. Now, we can mirror the steps followed in the proof of the result for *TV* by making the appropriate modifications along the way. We start with making the following decomposition

$$\begin{aligned} DV_1^n &= \frac{1}{2}\sum_{i=2}^{K_n(\zeta/2)} (\Delta_i^n Y - \Delta_{i-1}^n Y)^2 \mathbb{1}_{\{|\Delta_i^n Y - \Delta_{i-1}^n Y| \leq \zeta \Delta_n^{\varpi}\}} + \frac{1}{2}\sum_{K_n(\zeta/2)+1}^{n} (\Delta_i^n Y - \Delta_{i-1}^n Y)^2 \mathbb{1}_{\{|\Delta_i^n Y - \Delta_{i-1}^n Y| \leq \zeta \Delta_n^{\varpi}\}} \\ &:= B_1^n + B_2^n, \end{aligned} \tag{49}$$

and as in the case for the proof of *TV*, we have for Δ_n small enough that $1 < K_n(\zeta/2) < n$. First, exactly as the result for A_1^n in the proof of the result for TV_1^n, we have

$$\left| B_1^n - \int_0^{K_n(\zeta/2)\Delta_n} \sigma_s^2 ds \right| \leq \zeta^2 \Delta_n^{2\varpi} K_n(\zeta/2) + \int_0^{K_n(\zeta/2)\Delta_n} \sigma_s^2 ds = O_p\left(\Delta_n^{2\varpi + \frac{1-\varpi-\alpha}{\alpha+1}}\right). \tag{50}$$

Next, using Taylor expansion (twice), we have

$$\begin{aligned} \sum_{i=K_n(\zeta/2)+1}^{n} (\Delta_i^n H - \Delta_{i-1}^n H)^2 &\leq \alpha^2(1-\alpha)^2 c_r^2 \Delta_n^{2(1-\alpha)} \sum_{i=K_n(\zeta/2)+1}^{n} i^{-2(\alpha+1)} \\ &\sim \Delta_n^{2(1-\alpha)} K_n(\zeta/2)^{1-2(\alpha+1)} \int_1^{n/K_n(\zeta/2)} s^{-2(\alpha+1)} ds \\ &\sim \Delta_n^{2\varpi + \frac{1-\varpi-\alpha}{\alpha+1}}. \end{aligned} \tag{51}$$

Using this result and Burkholder–Davis–Gundy inequality, we have

$$\sum_{i=K_n(\zeta/2)+1}^{n}(\Delta_i^n X-\Delta_{i-1}^n X)(\Delta_i^n H-\Delta_{i-1}^n H)\mathbb{1}_{\{|\Delta_i^n Y-\Delta_{i-1}^n Y|\le\zeta\Delta_n^{\varpi}\}}=O_p\left(\sqrt{\Delta_n}\Delta_n^{\varpi+\frac{1-\varpi-\alpha}{2(\alpha+1)}}\right)=o_p(\sqrt{\Delta_n}). \tag{52}$$

Altogether, we have

$$B_2^n-\int_{K_n(\zeta/2)\Delta_n}^{1}\sigma_s^2 ds=O_p\left(\Delta_n^{2\varpi+\frac{1-\varpi-\alpha}{\alpha+1}}\bigvee\sqrt{\Delta_n}\right). \tag{53}$$

The results for B_1^n and B_2^n imply the result about DV in the theorem. □

Proof of Theorem 3. We will prove the result for TV_1^n and DV_1^n. We denote with X^c the continuous part of X. Then, on an increment without jump in X, we have $\Delta_i^n Y=\Delta_i^n X^c+\Delta_i^n H$. Taking into account then the fact that jumps are of finite activity, i.e., $r=0$, we have

$$\sum_{i=1}^{n}(\Delta_i^n Y)^2\mathbb{1}_{\{|\Delta_i^n Y|\le\zeta\Delta_n^{\varpi}\}}-\sum_{i=1}^{n}(\Delta_i^n X^c+\Delta_i^n H)^2\mathbb{1}_{\{|\Delta_i^n X^c+\Delta_i^n H|\le\zeta\Delta_n^{\varpi}\}}=O_p(\Delta_n^{2\varpi})=o_p(\sqrt{\Delta_n}), \tag{54}$$

and

$$\begin{aligned}&\sum_{i=2}^{n}(\Delta_i^n Y-\Delta_{i-1}^n Y)^2\mathbb{1}_{\{|\Delta_i^n Y-\Delta_{i-1}^n Y|\le\zeta\Delta_n^{\varpi}\}}\\&-\sum_{i=2}^{n}(\Delta_i^n X^c-\Delta_{i-1}^n X^c+\Delta_i^n H-\Delta_{i-1}^n H)^2\mathbb{1}_{\{|\Delta_i^n X^c-\Delta_{i-1}^n X^c+\Delta_i^n H-\Delta_{i-1}^n H|\le\zeta\Delta_n^{\varpi}\}}=o_p(\sqrt{\Delta_n}),\end{aligned} \tag{55}$$

because of the restriction $\varpi>\frac{1}{4}$. We next note that for $\alpha\ge 0$, we have

$$\int_{t_1}^{t_2}(s-\tau)^{\alpha-1}ds=\frac{1}{\alpha}(t_2-\tau)^{\alpha}-\frac{1}{\alpha}(t_1-\tau)^{\alpha},\quad 0\le\tau\le t_1<t_2, \tag{56}$$

$$\int_{t_1}^{t_2}(\tau-s)^{\alpha-1}ds=\frac{1}{\alpha}(\tau-t_1)^{\alpha}-\frac{1}{\alpha}(\tau-t_2)^{\alpha},\quad 0\le t_1<t_2\le\tau. \tag{57}$$

In addition, we note that $\Delta_n K_n(\zeta)$ in the proof of Theorem 2 is $o_p(1)$ (both for TV and DV). Therefore, with probability approaching one, the sets of increments for which $\Delta_j^n H^{(i)}$ exceeds in absolute value $\zeta\Delta_n^{\varpi}$, for $i=1,\dots,N$, are disjoint. From here, the proof can proceed following similar steps as the proof of Theorem 2 to establish that

$$\sum_{i=1}^{n}(\Delta_i^n X^c+\Delta_i^n H)^2\mathbb{1}_{\{|\Delta_i^n X^c+\Delta_i^n H|\le\zeta\Delta_n^{\varpi}\}}-\sum_{i=1}^{n}(\Delta_i^n X^c)^2=o_p(\sqrt{\Delta_n}), \tag{58}$$

and

$$\begin{aligned}&\sum_{i=2}^{n}(\Delta_i^n X^c-\Delta_{i-1}^n X^c+\Delta_i^n H-\Delta_{i-1}^n H)^2\mathbb{1}_{\{|\Delta_i^n X^c-\Delta_{i-1}^n X^c+\Delta_i^n H-\Delta_{i-1}^n H|\le\zeta\Delta_n^{\varpi}\}}-\sum_{i=2}^{n}(\Delta_i^n X^c-\Delta_{i-1}^n X^c)^2\\&=o_p(\sqrt{\Delta_n}).\quad\square\end{aligned} \tag{59}$$

Proof of Theorems 4–6. The proof follows similar steps as for those of Theorems 1–3. □

Proof of Theorem 7. We will consider the interval [0, 1]. First, exactly as the proof of Theorem 3, it suffices to consider only the case when X does not contain jumps in the interval [0, 1], there is only one τ and it is equal to 0. Using the notation in the proof of the result about DV in Theorem 2, we have

$$\begin{aligned}RQ_1^n&=\frac{1}{12\Delta_n}\sum_{i=2}^{K_n(\zeta/2)}(\Delta_i^n Y-\Delta_{i-1}^n Y)^4\mathbb{1}_{\{|\Delta_i^n Y-\Delta_{i-1}^n Y|\le\zeta\Delta_n^{\varpi}\}}+\frac{1}{12\Delta_n}\sum_{K_n(\zeta/2)+1}^{n}(\Delta_i^n Y-\Delta_{i-1}^n Y)^4\mathbb{1}_{\{|\Delta_i^n Y-\Delta_{i-1}^n Y|\le\zeta\Delta_n^{\varpi}\}}\\&:=C_1^n+C_2^n,\end{aligned} \tag{60}$$

and as in the case for the proof of Theorem 2, we have for Δ_n small enough that $1<K_n(\zeta/2)<n$. For C_1^n, we have

$$C_1^n-\int_0^{K_n(\zeta/2)\Delta_n}\sigma_s^4 ds=O_p\left(\Delta_n^{4\varpi-1+\frac{1-\varpi-\alpha}{\alpha+1}}\right), \tag{61}$$

which is $o_p(1)$ because $\alpha < 1$ and $\varpi > \frac{2}{7}$ (using again the notation for α in the proof of Theorem 2). For C_2^n, first exactly as in the proof of Theorem 2, we have

$$\frac{1}{\Delta_n}\sum_{i=K_n(\zeta/2)+1}^{n}\left(\Delta_i^n H - \Delta_{i-1}^n H\right)^4 = O_p\left(\Delta_n^{4\varpi-1+\frac{1-\varpi-\alpha}{1+\alpha}}\right), \tag{62}$$

which is again $o_p(1)$ because $\alpha < 1$ and $\varpi > \frac{2}{7}$. From here the asymptotic negligibility of $C_1^n - \int_{K_n(\zeta/2)\Delta_n}^1 \sigma_s^4 ds$ follows by application of Holder's inequality, Burkholder–Davis–Gundy inequality and Theorem 3.4.1 of Jacod and Protter (2012). □

Proof of Theorem 8. The CLT result to be proved is based on the following two results:

(1) We have

$$\sqrt{n}\sum_{i=(t-1)n+2}^{n}\Delta_i^n Y\Delta_{i-1}^n Y \xrightarrow{\mathcal{L}-s} \int_{t-1}^{t}\sigma_s^2 d\overline{W}_s + \sum_{s\in[t-1,t]}\Delta X_s(\sigma_{s-}\eta_s^- + \sigma_s\eta_s^+), \tag{63}$$

where $\overline{W}$ and the sequence $\{\eta_s^-, \eta_s^+)_{s\geq 0}$ are defined on an extension of the original probability space, are independent of it and of each other, $\overline{W}$ is a Brownian motion and $\{\eta_s^-, \eta_s^+)_{s\geq 0}$ is an i.i.d. sequence of standard normal random variables.

(2) We have

$$RQ_t^n + \sum_{i=(t-1)n+2}^{n}(\Delta_i^n Y - \Delta_i^n Y)^2 \mathbb{1}_{\{|\Delta_i^n Y - \Delta_{i-1}^n|>\zeta\Delta_n^{\varpi}\}}(\widehat{c}_{i,-}^n + \widehat{c}_{i,+}^n) \xrightarrow{\mathbb{P}} \int_{t-1}^{t}\sigma_s^4 ds + \sum_{s\in[t-1,t]}(\Delta X_s)^2(c_{s-} + c_s). \tag{64}$$

Starting with the limit result (1), its proof follows essentially the scheme of proving the CLT for realized quadratic variation in the presence of jumps given in Theorem 5.4.6 in Jacod and Protter (2012). In particular, using a standardization localization procedure, it suffices to prove the result when certain processes in the dynamics of X (volatility, drift term and jump size) are bounded. Another simplification then can be further made by looking at the case when the jump size is above a positive threshold. The latter means that jumps in X are of finite activity. In that case, on a set of probability approaching one, every increment contains at most one jump and no two consecutive increments contain two jumps. On that probability set, the autocovariance can be split into two sums, one containing the increments without jumps and one with the jumps. For the former sum, Theorem 1 provides the CLT result. For the part of the autocovariance that contains the increments with jumps, we can apply Theorem 4.3.1 in Jacod and Protter (2012) which shows that the sum of the two parts of the autocovariance converge jointly to the limit given above.

The proof of (2) is done exactly the same way as the proof of Theorems 9.2.1 and 9.5.1 in Jacod and Protter (2012). □

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2020.11.005.

References

Aït-Sahalia, Y., Mykland, P.A., Zhang, L., 2011. Ultra high frequency volatility estimation with dependent microstructure noise. J. Econometrics 160 (1), 160–175.
Andersen, T.G., Archalov, I., Cebiroglu, G., Hautsch, N., 2020. Local Mispricing and Microstructural Noise: A Parametric Perspective. Working Paper, Kellogg School of Management, Northwestern University.
Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. Internat. Econom. Rev. 39 (4), 885–905.
Andersen, T.G., Bollerslev, T., Diebold, F.X., 2007. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. Rev. Econ. Stat. 89 (4), 701–720.
Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2000. Great realisations. Risk Mag. 18, 105–108.
Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71 (2), 579–625.
Andersen, T.G., Dobrev, D., Schaumburg, E., 2012. Jump-robust volatility estimation using nearest neighbor truncation. J. Econometrics 169 (1), 75–93.
Bandi, F.M., Russell, J.R., 2008. Microstructure noise, realized variance, and optimal sampling. Rev. Econom. Stud. 75 (2), 339–369.
Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 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., Hansen, P.R., Lunde, A., Shephard, N., 2009. Realized kernels in practice: Trades and quotes. Econom. J. 12 (3), C1–C32.
Barndorff-Nielsen, O.E., Shephard, N., 2002. Estimating quadratic variation using realized variance. J. Appl. Econometrics 17, 457–477.
Barndorff-Nielsen, O.E., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. J. Financ. Econ. 2 (1), 1–37.
Barndorff-Nielsen, O.E., Shephard, N., 2006. Econometrics of testing for jumps in financial economics using bipower variation. J. Financ. Econ. 4 (1), 1–30.
Barndorff-Nielsen, O.E., Shephard, N., Winkel, M., 2006. Limit theorems for multipower variation in the presence of jumps. Stochastic Process. Appl. 116 (5), 796–806.
Bender, C., 2012. Simple arbitrage. Ann. Appl. Probab. 22 (5), 2067–2085.
Bollerslev, T., Patton, A.J., Quaedvlieg, R., 2016. Exploiting the errors: A simple approach for improved volatility forecasting. J. Econometrics 192 (1), 1–18.

Christensen, K., Oomen, R.C., Renò, R., 2020. The drift burst hypothesis. J. Econometrics forthcoming.
Corsi, F., 2009. A simple approximate long-memory model of realized volatility. J. Financ. Econ. 7 (2), 174–196.
Corsi, F., Pirino, D., Renò, R., 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. J. Econometrics 159 (2), 276–288.
Da, R., Xiu, D., 2019. When Moving-Average Models Meet High-Frequency Data: uniform Inference on Volatility. Chicago Booth Research Paper.
Delbaen, F., Schachermayer, W., 1994. A general version of the fundamental theorem of asset pricing. Math. Ann. 300 (1), 463–520.
Delbaen, F., Schachermayer, W., 1995. The existence of absolutely continuous local martingale measures. Ann. Appl. Probab. 926–945.
Diebold, F.X., 2015. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests. J. Bus. Econom. Statist. 33 (1), 1.
Farmer, L., Schmidt, L., Timmermann, A., 2019. Pockets of predictability. Available at SSRN 3152386.
Guasoni, P., 2006. No arbitrage under transaction costs, with fractional brownian motion and beyond. Math. Finance 16 (3), 569–582.
Hansen, P.R., Lunde, A., 2006. Realized variance and market microstructure noise. J. Bus. Econom. Statist. 24 (2), 127–161.
Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., Vetter, M., 2009. Microstructure noise in the continuous case: the pre-averaging approach. Stochastic Process. Appl. 119 (7), 2249–2276.
Jacod, J., Li, Y., Zheng, X., 2017. Statistical properties of microstructure noise. Econometrica 85 (4), 1133–1174.
Jacod, J., Li, Y., Zheng, X., 2019. Estimating the integrated volatility with tick observations. J. Econometrics 208 (1), 80–100.
Jacod, J., Protter, P., 2012. Discretization of Processes. Springer.
Jacod, J., Todorov, V., 2014. Efficient estimation of integrated volatility in presence of infinite variation jumps. Ann. Statist. 42 (3), 1029–1069.
Jacod, J., Todorov, V., 2018. Limit theorems for integrated local empirical characteristic exponents from noisy high-frequency data with application to volatility and jump activity estimation. Ann. Appl. Probab. 28 (1), 511–576.
Jarrow, R.A., Protter, P., Sayit, H., 2009. No arbitrage without semimartingales. Ann. Appl. Probab. 19 (2), 596–616.
Jiang, J., Oomen, R., 2008. Testing for jumps when asset prices are observed with noise- a "swap variance" approach. J. Econometrics 144 (2), 352–370.
Kinnebrock, S., Podolskij, M., 2008. A note on the central limit theorem for bipower variation of general functions. Stoch. Process. Appl. 118 (6), 1056–1070.
Kirilenko, A., Kyle, A.S., Samadi, M., Tuzun, T., 2017. The flash crash: High-frequency trading in an electronic market. J. Finance 72 (3), 967–998.
Li, Z.M., Linton, O.B., 2019. A remedi for microstructure noise. Available at SSRN 3423607.
Li, Y., Xie, S., Zheng, X., 2016. Efficient estimation of integrated volatility incorporating trading information. J. Econometrics 195 (1), 33–50.
Li, Y., Zhang, Z., Li, Y., 2018. A unified approach to volatility estimation in the presence of both rounding and random market microstructure noise. J. Econometrics 203 (2), 187–222.
Mancini, C., 2009. Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. Scand. J. Stat. 36 (2), 270–296.
Patton, A.J., 2011. Volatility forecast comparison using imperfect volatility proxies. J. Econometrics 160 (1), 246–256.
Patton, A.J., Sheppard, K., 2015. Good volatility, bad volatility: Signed jumps and the persistence of volatility. Rev. Econ. Stat. 97 (3), 683–697.
Strasser, E., 2005. Characterization of arbitrage-free markets. Ann. Appl. Probab. 15 (1A), 116–124.
Swanson, N.R., White, H., 1997. Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. Int. J. Forecast. 13 (4), 439–461.
Todorov, V., 2013. Power variation from second order differences for pure jump semimartingales. Stochastic Process. Appl. 123 (7), 2829–2850.
Todorov, V., 2019. Nonparametric spot volatility from options. Ann. Appl. Probab. 29 (6), 3590–3636.
Xiu, D., 2010. Quasi-maximum likelihood estimation of volatility with high frequency data. J. Econometrics 159 (1), 235–250.
Zhou, B., 1996. High-frequency data and volatility in foreign-exchange rates. J. Bus. Econom. Statist. 14 (1), 45–52.