The drift burst hypothesis☆

Kim Christensen [a,*], Roel Oomen [b,c], Roberto Renò [d]

[a] *Department of Economics and Business Economics, CREATES, Aarhus University, Denmark*
[b] *Deutsche Bank, London, United Kingdom*
[c] *Department of Statistics, London School of Economics, United Kingdom*
[d] *Department of Economics, University of Verona, Italy*

A R T I C L E   I N F O

A B S T R A C T

The drift burst hypothesis postulates the existence of short-lived locally explosive trends in the price paths of financial assets. The recent U.S. equity and treasury flash crashes can be viewed as two high-profile manifestations of such dynamics, but we argue that drift bursts of varying magnitude are an expected and regular occurrence in financial markets that can arise through established mechanisms of liquidity provision. We show how to build drift bursts into the continuous-time Itô semimartingale model, elaborate on the conditions required for the process to remain arbitrage-free, and propose a nonparametric test statistic that identifies drift bursts from noisy high-frequency data. We apply the test and demonstrate that drift bursts are a stylized fact of the price dynamics across equities, fixed income, currencies and commodities. Drift bursts occur once a week on average, and the majority of them are accompanied by subsequent price reversion and can thus be regarded as "flash crashes." The reversal is found to be stronger for negative drift bursts with large trading volume, which is consistent with endogenous demand for immediacy during market crashes.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

The orderly functioning of financial markets is viewed by most regulators as their first and foremost objective. It is therefore unsurprising that the recent flash crashes in the U.S. equity and treasury markets are subject to intense debate and scrutiny, not least because they raise concerns around the stability of the market and the integrity of its design (see e.g. CFTC and SEC, 2010, 2011; U.S. Treasury, FRB, NYFED, SEC, and CFTC, 2015, and Fig. 1 for an illustration). Moreover,

Panel A: S&P500 equity index.
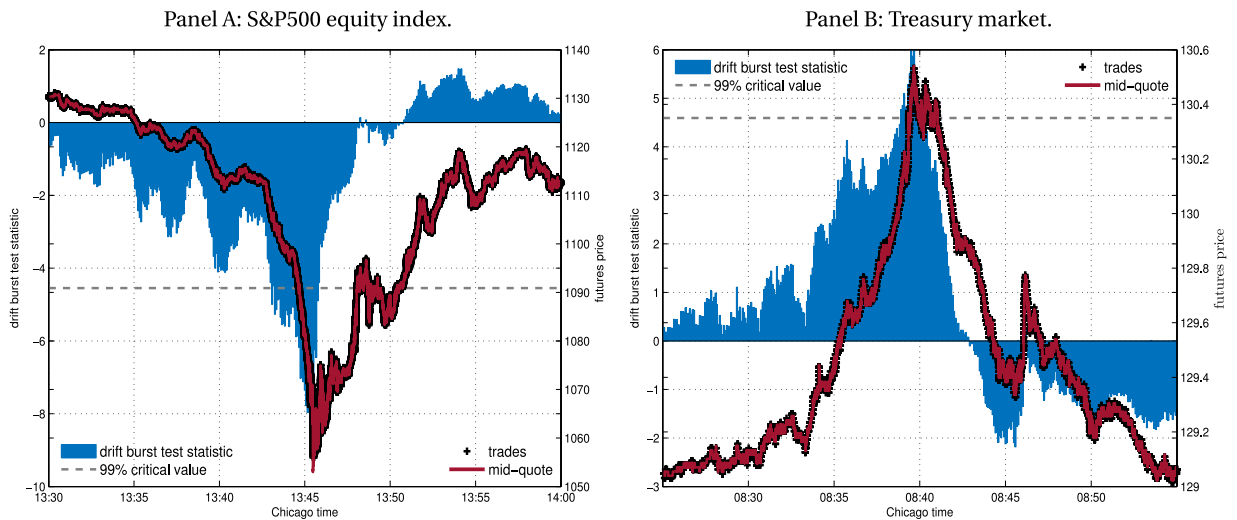
Panel B: Treasury market.

Fig. 1. The U.S. S&P500 equity index and treasury market flash crash. *Note.* This figure draws the mid-quote and traded price (right axis) of the E-mini S&P500 (in Panel A) and 10-Year Treasury Note (in Panel B) futures contracts over the flash crash episodes of May 6, 2010 and October 15, 2014. Superimposed is the nonparametric drift burst t-statistic (left axis) proposed in this paper.

there is growing consensus that flash crashes of varying magnitude are becoming more frequent across financial markets.[1] The distinct price evolution over such events – with highly directional and sustained price moves – poses three direct challenges to the academic community. First, how can one formally model such dynamics? The literature on continuous-time finance has focused extensively on the volatility and jump components of the price process, but as we show, these are not sufficient to explain the observed dynamics. Secondly, how can one identify or test for the presence of such features in the data? And third, can such events be reconciled within the theory of price formation in the presence of market frictions? This paper addresses all these challenges.

Suppose the log-price of a traded asset, $X = (X_t)_{t\geq 0}$, has the dynamic:

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

where μ_t is the drift, σ_t is the volatility, W_t a Brownian motion, and J_t is a jump process. In a conventional setup with locally bounded coefficients, over a vanishing time interval $\Delta \to 0$, the drift is $O_p(\Delta)$ (as is the jump term) and swamped by a diffusive component of larger order $O_p(\sqrt{\Delta})$. Hence, much of the infill asymptotics is unaffected by the presence of a drift and the theory therefore invariably neglects it. Also, in empirical applications, particularly those relying on intraday data over short horizons, the drift term is generally small and estimates of it are subject to considerable measurement error (e.g. Merton, 1980). Thus, the common recommendation is to ignore it. Yet, to explain such events as those in Fig. 1, it is hard to see how the drift component can be dismissed. Our starting point is therefore – what we refer to as – the drift burst hypothesis, which postulates the existence of short-lived locally explosive trends in the price paths of financial assets. The objective of this paper is to build theoretical and empirical support for the hypothesis, thereby contributing towards a better understanding of financial market dynamics. We show how drift bursts can be embedded in Eq. (1). Next, we develop a feasible nonparametric identification strategy that enables the online detection of drift bursts from high-frequency data. What is tested here is a drift explosion against the null hypothesis of a jump-diffusion model.[2] The empirical application demonstrates that drift bursts are a stylized fact of the price process.

An exploding drift is unconventional in continuous-time finance, but there are a number of theoretical models of price formation that back the idea. Grossman and Miller (1988) examine a collection of risk-averse market makers that provide immediacy in exchange for a positive expected (excess) return $\mu = E(P_1/P_0 - 1)$ of the form:

$$\frac{\mu}{\sigma} = \frac{s\gamma}{1+M}\sigma P_0, \tag{2}$$

where P_0 is the initial price, P_1 is the price at which the market maker trades, σ is the standard deviation of the price move, s is the size of the order that requires execution, γ is the risk aversion of market makers, and M is the number of market makers competing for the order. Eq. (2) illustrates that the drift can dominate the volatility when liquidity demand (s) is unusually high or the willingness or capacity of the collective market makers to absorb order flow is

---

[1] Nanex Research has reported hundreds of flash crashes across all major financial markets, see http://www.nanex.net/NxResearch/. Related work includes Golub et al. (2012).

[2] We assume that volatility is strictly positive, so pure-jump Itô semimartingale processes are ruled out.

impaired (i.e. increased risk aversion γ or fewer active market makers M). This prediction fits the 2010 equity flash crash in that the price drop appeared to be accompanied by increased risk aversion and a rapid decline in the number of participating market makers. CFTC and SEC (2010) write "some market makers and other liquidity providers widened their quote spreads, others reduced offered liquidity, and a significant number withdrew completely from the markets". The subsequent price reversal observed in Fig. 1 is also predicted by this model as the excess return is only temporary and the long-run price level returns to P_0. In related work, Campbell et al. (1993) show that as liquidity demand increases (as measured by trading volume), the price reaction and subsequent reversal grow in size. Alternative mechanisms that can generate these price dynamics include trading frictions as in Huang and Wang (2009), predatory trading and forced liquidation as in Brunnermeier and Pedersen (2005, 2009), or agents with tournament-type preferences and an aversion to missing out on trends as in Johnson (2016).[3] While this literature provides valuable insights and hypotheses regarding price dynamics, the testable implications often relate to confounding measures such as the unconditional serial correlation of price returns. Moreover, because the theory is typically cast as a two-period model, it does not readily translate into an econometric identification strategy of the impacted sample paths in continuous-time. In addition to the above, there is an extensive econometrics literature on bubble detection (see e.g. Phillips and Yu, 2011), which is potentially related to our paper, but it operates on a much longer time scale than the one analyzed here. We deliver a foundation that increases the depth of the empirical work that can be conducted in these areas.

With the drift burst hypothesis and the corresponding Itô semimartingale price process in place, we develop a nonparametric identification strategy for the online detection of drift burst sample paths from intraday noisy high-frequency data. Our approach aims to establish whether the observed price movement is generated by the drift rather than be the result of volatility. Unsurprisingly, it requires estimation of the local drift and volatility coefficients which is nontrivial for several reasons. First, from Merton (1980) we know that even when the drift term is a constant, it cannot be estimated consistently over a bounded time interval. Secondly, while infill asymptotics do provide consistent volatility estimates, in practice microstructure effects complicate inference. Building on Bandi (2002) and Kristensen (2010) for coefficient estimation and Newey and West (1987), Andrews (1991), Barndorff-Nielsen et al. (2008) and Jacod et al. (2009) for the robustification to microstructure noise, we formulate a nonparametric kernel-based filtering approach that delivers estimates of the local drift and volatility on the basis of which we construct the test statistic. Under the null hypothesis of no drift burst, the test is asymptotically standard normal, but it diverges – and therefore has power under the alternative – when the drift explodes sufficiently fast. When calculated sequentially and using potentially overlapping data, the critical values of the test are determined on the basis of extreme value theory, as in Lee and Mykland (2008). A simulation study confirms that the test is capable of identifying drift bursts. Interestingly, applying the test to high-frequency data for the days of the U.S. equity and treasury market flash crashes, displayed in Fig. 1, we find that they constitute highly significant drift bursts.

The introduction of drift bursts via an exploding drift coefficient provides an essential ingredient, which helps to reconcile a number of phenomena observed in financial markets. The first is the occurrence of flash crashes, where highly directional and sustained price movements are reversed shortly after. While there is a substantial body of research that looks at the 2010 equity market flash crash (a partial list includes Easley et al., 2011; Madhavan, 2012; Andersen et al., 2015; Kirilenko et al., 2017; Menkveld and Yueshen, 2018), there has been no attempt thus far to analyze these events in a more systematic fashion. Our test procedure lays down a framework that makes this possible. The second is that of "gradual jumps" – in Barndorff-Nielsen et al. (2009) terminology – where the price converges in a rapid but continuous fashion to a new level. This relates to a puzzle put forward by Christensen et al. (2014), who find that the total return variation that can be attributed to the jump component is an order of magnitude smaller than previously reported by extensive empirical literature (see also Bajgrowicz et al., 2016). In particular, they show that jumps identified using data sampled at a five-minute frequency often vanish when viewed at the highest available tick frequency and instead appear as sharp but continuous price movements. They show that spurious detection of jumps at low frequency can be explained by an erratic volatility process. However, because volatility merely leads to wider price dispersion, it fails to reconcile the often steady and directional price evolution over such episodes. We argue that the drift burst hypothesis constitutes a more intuitive and appealing mechanism that can explain the reported over-estimation of the total jump variation.

We undertake an empirical analysis to determine the prevalence of drift bursts in practice and to characterize their basic features. To that end, we employ a comprehensive set of high-quality tick data covering some of the most liquid futures contracts across the equity, fixed income, currency, and commodity markets. We calculate the drift burst test statistic at five-second intervals over a multi-year sample. Our findings demonstrate that drift bursts are an integral part of the price process across all asset classes. Over the full sample we identify more than one thousand highly significant events, or roughly one per week. For most of the drift bursts we detect, the magnitude of the price drop or surge typically ranges between 25 and 200 basis points, with only a handful of extreme moves between 3% and 8%. We find that roughly two thirds of the drift bursts are followed by price reversion, which means that many of the identified events resemble (mini) flash crashes that are symptomatic of liquidity shocks. Consistent with the literature on price formation, particularly that in Huang and Wang (2009) of endogenous trading imbalances generated by costly market presence, we find that trading volume during a drift burst is highly correlated with the subsequent price reversal. The post-drift burst return

---

[3] There are also a number of practical mechanisms that cause or amplify price drops and surges, such as margin calls on leveraged positions (i.e. forced liquidation), dynamic hedging of short-gamma positions, stop-loss orders, or technical momentum trading strategies.

can therefore – as predicted by the theory – be interpreted as a compensation for supplying immediacy during times of substantial market stress.

The remainder of the paper is organized as follows. Section 2 introduces the drift burst hypothesis and describes the theoretical framework. Section 3 develops the identification strategy on the basis of noisy high-frequency data. Section 4 includes an extensive simulation study that demonstrates the properties of the test. The empirical application is found in Section 5, while Section 6 concludes.

2. The hypothesis

The log-price process $X = (X_t)_{t \geq 0}$, introduced in Eq. (1), is an Itô semimartingale defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t \geq 0}, \mathscr{P})$. We assume the following about X.

Assumption 1. X is described by the dynamics in Eq. (1), where X_0 is $\mathscr{F}_0$-measurable, $\mu = (\mu_t)_{t \geq 0}$ is a locally bounded and predictable drift, $\sigma = (\sigma_t)_{t \geq 0}$ is an adapted, càdlàg, locally bounded and almost surely (a.s.) strictly positive volatility, $W = (W_t)_{t \geq 0}$ is a standard Brownian motion and $J = (J_t)_{t \geq 0}$ is a pure-jump process.

The above is a standard formulation for continuous-time arbitrage-free price processes. It represents our frictionless null, where the price is in a "normal" state with non-explosive (locally bounded) coefficients μ_t and σ_t. We do not restrict the model in any essential way, other than by imposing mild regularity conditions on the driving terms, which are listed in Assumption 3 in Appendix A. As such, it encompasses a wide range of specifications and is compatible with time-varying expected returns, stochastic volatility, leverage effects, and jumps (both of finite and infinite activity) in the log-price and volatility. Below, we further add pre-announced jumps, explosive volatility, and additive microstructure noise. Note, however, that pure-jump processes without a Brownian term are not analyzed, because we require σ_t to be strictly positive in Assumption 1.

To introduce the drift burst hypothesis, we momentarily enforce that X has continuous sample paths, i.e. $dJ_t = 0$. The jump process is fully reactivated and driving functions are generalized in our theoretical results.

As μ and σ are locally bounded under Assumption 1, it follows that for a fixed time point τ_{db}:

$$\int_{\tau_{db}-\Delta}^{\tau_{db}+\Delta} \mu_s ds = O_p(\Delta) \qquad \text{and} \qquad \int_{\tau_{db}-\Delta}^{\tau_{db}+\Delta} \sigma_s dW_s = O_p(\sqrt{\Delta}), \tag{3}$$

as $\Delta \to 0$. Thus, the drift is much smaller than the volatility, because $\Delta \ll \sqrt{\Delta}$. This is consistent with the notion that over short time intervals the main contributor to the log-return is volatility. It is this feature that has led the econometrics literature to largely neglect the drift.

However, the drift can prevail in an alternative model where, in a neighborhood of τ_{db}, μ is allowed to diverge in such a way that:

$$\int_{\tau_{db}-\Delta}^{\tau_{db}+\Delta} \mu_s ds = O_p(\Delta^{\gamma_\mu}), \tag{4}$$

with $0 < \gamma_\mu < 1/2$. We refer to an exploding drift coefficient as a *drift burst* and to τ_{db} as a *drift burst time*. The condition $\gamma_\mu > 0$ ensures continuity of the sample path.

An example of an exploding drift leading to a drift burst is:

$$\mu_t^{db} = \begin{cases} a_1 (\tau_{db} - t)^{-\alpha} & t < \tau_{db} \\ a_2 (t - \tau_{db})^{-\alpha} & t > \tau_{db} \end{cases}, \tag{5}$$

with $1/2 < \alpha < 1$ and a_1, a_2 constants. Setting $\gamma_\mu = 1 - \alpha$, this formulation is consistent with Eq. (4). This specification of the drift can capture flash crashes when a_1 and a_2 are of the opposite sign (see e.g. Panel A in Fig. 2). It can also accommodate gradual jumps without reversion, e.g. when $a_2 = 0$.[4]

The log-price with explosive coefficients, replacing the drift in Eq. (1) with a drift as in Eq. (5), is denoted by $\widetilde{X}$. As a process it is still a semimartingale, which is necessary – but not sufficient – to exclude arbitrage from the model (e.g. Delbaen and Schachermayer, 1994).[5] To prevent arbitrage, a further condition (imposed by Girsanov's Theorem) is necessary for the existence of an equivalent martingale measure:

$$\int_{\tau_{db}-\Delta}^{\tau_{db}+\Delta} \left(\frac{\mu_s}{\sigma_s}\right)^2 ds < \infty \quad (a.s.), \tag{6}$$

---

[4] The drift burst specification in Eq. (5) also allows for gradual jumps that start off strong and then decelerate (when $a_1 = 0$ and $a_2 \neq 0$), akin to price behavior observed around, for instance, scheduled news announcements, where the first order price impact tends to be realized immediately, but it may then be followed by a gradual continuation as the market interprets and fully incorporates the shock.

[5] While explosive drift does not impede the semimartingale structure, it can on the other hand negatively affect nonparametric estimation of volatility from high-frequency data, as unveiled by Example 3.4.2 in Jacod and Protter (2012), because the volatility is completely swamped by the drift. This is consistent with the findings of Li et al. (2015), who note that standard OLS estimation of their proposed jump regression is seriously affected by the inclusion of two outliers in the sample. Incidentally, these are the equity flash crash of May 6, 2010 (in Fig. 1) and the hoax tweet of April 23, 2014 (in Fig. 6).
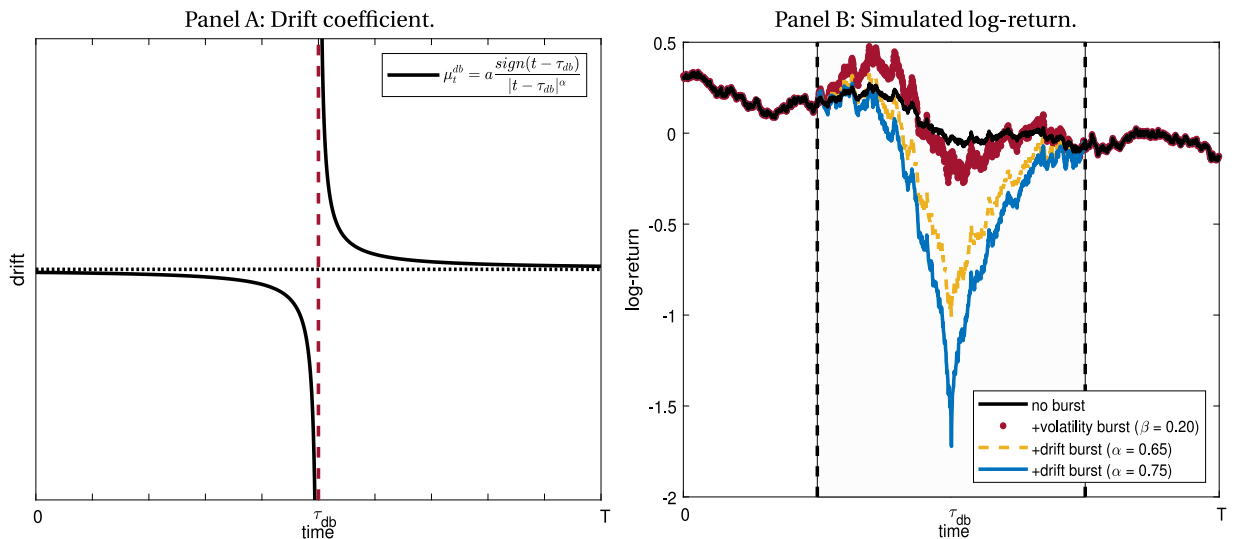
Panel A: Drift coefficient.

Panel B: Simulated log-return.

Fig. 2. Illustration of a log-price with a drift burst. *Note.* In Panel A, the drift coefficient is shown against time, while Panel B shows the evolution of a simulated log-price with a burst in: (i) nothing, (ii) volatility, and (iii) drift and volatility. The latter is based on Eqs. (5) and (8) with $-a_1 = a_2 = 3$, $b = 0.15$, $\alpha = 0.65$ or 0.75 and $\beta = 0.2$.

which is known as a "structural condition".[6] This cannot hold if the drift explodes in the neighborhood of τ_{db}, but the volatility remains bounded, as it allows for a so-called "free lunch with vanishing risk", see Definition 10.6 in Björk (2003). Thus, explosive volatility is a necessary condition for drift bursts in a market free of arbitrage.

We say there is a volatility burst, if

$$\int_{\tau_{db}-\Delta}^{\tau_{db}+\Delta} \sigma_s dW_s = O_p\left(\Delta^{\gamma_\sigma}\right),\tag{7}$$

with $0 < \gamma_\sigma < 1/2$. As above, an example of a bursting volatility is:

$$\sigma_t^{vb} = b\,|\tau_{db} - t|^{-\beta}\tag{8}$$

with $0 < \beta < 1/2$ and $b > 0$. We here restrict β to ensure that $\int_{\tau_{db}-\Delta}^{\tau_{db}+\Delta} \left(\sigma_s^{vb}\right)^2 ds < \infty$, so that the stochastic integral in Eq. (7) can be defined. In this case,

$$\int_{\tau_{db}-\Delta}^{\tau_{db}+\Delta} \sigma_s^{vb} dW_s = O_p\left(\Delta^{1/2-\beta}\right),\tag{9}$$

so that $\gamma_\sigma = 1/2 - \beta$.

We can thus introduce a "canonical" alternative model:

$$d\widetilde{X}_t = \mu_t^{db} dt + \sigma_t^{vb} dW_t,\tag{10}$$

for which $\mu_t^{db}/\sigma_t^{vb} \to \infty$ as $t \to \tau_{db}$ if $\alpha > \beta$. The structural condition in Eq. (6) is readily fulfilled when $\alpha - \beta < 1/2$. Thus, this example shows that the drift coefficient can explode locally (even after normalizing by an exploding volatility) either preserving absence of arbitrage (when $0 < \alpha - \beta < 1/2$), or allowing local arbitrage opportunities (when $\alpha - \beta > 1/2$). However, as our econometric analysis in Theorem 2 reveals, we are only able to detect an explosive drift with $\alpha - \beta > 1/2$, that is when a short-lived absence of arbitrage is found. The intuition for this result is that under the condition in Eq. (6) we can always switch to an equivalent probability measure with no drift and unaltered volatility. Hence, there is no way to detect an explosive drift in the arbitrage-free setting.

In Panel B of Fig. 2, we show simulated sample paths of the associated log-price in this framework. A flash crash is generated from Eq. (10). We set the drift burst rate at $\alpha = 0.65$ and $\alpha = 0.75$ with a volatility burst parameter $\beta = 0.2$. While the former parametrization preserves absence of arbitrage (since $\alpha - \beta < 1/2$), the latter does not. Visually, however, the price dynamics in both scenarios is pronounced and qualitatively identical.

The notion that volatility can burst during market turbulence or dislocation is uncontroversial. Kirilenko et al. (2017) and Andersen et al. (2015) report elevated levels of volatility during the equity flash crash (see also Bates, 2018). However,

---

[6] The structural condition is sufficient together with the exponential moment condition $E\left[\exp\left\{-\int_{\tau_{db}-\Delta}^{\tau_{db}+\Delta} \frac{\mu_s}{\sigma_s} dW_s - \frac{1}{2}\int_{\tau_{db}-\Delta}^{\tau_{db}+\Delta} \left(\frac{\mu_s}{\sigma_s}\right)^2 ds\right\}\right] = 1$, see Theorem 4.2 in Karatzas and Shreve (1998).

as illustrated by Fig. 2 and proved formally in Theorem 2, a volatility burst in itself is not sufficient to capture the gradual jump or flash crash dynamics regularly observed in practice. The introduction of a separate drift burst component as we propose in this paper is a convenient and effective tool to reconcile continuous-time semimartingale theory with such empirical observations.

3. Identification

We now develop a nonparametric approach to detect drift bursts in real data. We propose a test statistic that exploits the message of Eq. (4), namely if there is a drift burst in the price process at time τ_{db}, the drift can prevail over volatility and locally dominate log-returns in the vicinity of τ_{db}. The test statistic thus compares a suitably rescaled estimate of μ_t/σ_t based on high-frequency data in a neighborhood of t. Later, we prove that our "signal-to-noise" measure uncovers drift bursts if they are sufficiently strong.

We extend existing work on nonparametric kernel-based estimation of the coefficients of diffusion processes to estimate μ_t and σ_t (e.g. Bandi, 2002; Kristensen, 2010). We assume that X is recorded at times $0 = t_0 < t_1 < \cdots < t_n = T$, where $\Delta_{i,n} = t_i - t_{i-1}$ is the time gap between observations and T is fixed. The sampling times are potentially irregular, as formalized in Assumption 4 in Appendix A. The discretely sampled log-return over $[t_{i-1}, t_i]$ is $\Delta_i^n X = X_{t_i} - X_{t_{i-1}}$, and we define:

$$\hat{\mu}_t^n = \frac{1}{h_n} \sum_{i=1}^n K\left(\frac{t_{i-1} - t}{h_n}\right) \Delta_i^n X, \quad \text{for } t \in (0, T],$$ (11)

where h_n is the bandwidth of the mean estimator and K is a kernel. We also set:

$$\hat{\sigma}_t^n = \left(\frac{1}{h_n'} \sum_{i=1}^n K\left(\frac{t_{i-1} - t}{h_n'}\right) \left(\Delta_i^n X\right)^2\right)^{1/2}, \quad \text{for } t \in (0, T],$$ (12)

where h_n' is the bandwidth of the volatility estimator.

The bandwidths h_n, h_n' and the kernel K are assumed to fulfill some weak regularity conditions that are succinctly listed in Assumption 5 in Appendix A.

In absence of a drift burst, the proof of Theorem 1 stated below shows that, as $n \to \infty$:

$$\sqrt{h_n}\left(\hat{\mu}_t^n - \mu_{t-}^*\right) \xrightarrow{d} N\left(0, K_2 \sigma_{t-}^2\right),$$ (13)

where $\mu_t^* = \mu_t + \int_{\mathbb{R}} \delta(t, x) I_{\{|\delta(t,x)|>1\}} \lambda(dx)$, K_2 is a kernel-dependent constant, and the above convergence is stable in law.

As shown by Eq. (13), $\hat{\mu}_t^n$ is asymptotically unbiased for the (jump compensated) drift term. It is inconsistent, because the variance explodes as $h_n \to 0$. This appears to rule out drift burst detection via $\hat{\mu}_t^n$. On the other hand, if we rescale the left-hand side of Eq. (13) with $\hat{\sigma}_t^n \sqrt{K_2}$, it appears the right-hand side has a standard normal distribution.[7] It is this insight that facilitates the construction of a test statistic that can identify drift bursts, as we prove in Theorem 1, which describes the behavior of the t-statistic under the null of locally bounded coefficients, whereas Theorem 2 does it under the alternative of explosive drift and volatility.

The test statistic is defined as:

$$T_t^n = \sqrt{\frac{h_n}{K_2}} \frac{\hat{\mu}_t^n}{\hat{\sigma}_t^n}.$$ (14)

T_t^n has an intuitive interpretation with the indicator kernel. Here, it is the ratio of the drift part to the volatility part of the log-return over the interval $[t - h_n, t]$ (as $h_n, h_n' \to 0$, this holds for any valid kernel).

Theorem 1. *Assume that X is a semimartingale as defined by Eq. (1), and that Assumptions 1, 3–5 are fulfilled. As $n \to \infty$, it holds that:*

$$T_t^n \xrightarrow{d} N(0, 1).$$ (15)

Proof. See Appendix A. ∎

Theorem 1 shows that, in absence of a drift burst, the t-statistic in Eq. (14) has a limiting standard normal distribution.[8] We note that under the null, T_t^n does not depend on μ_{t-}^* and σ_{t-} for large n. Thus, although it is not possible to consistently

---

[7] Lemma 3 in Appendix A shows that $\hat{\sigma}_t^n$ is a consistent estimator of σ_{t-}.

[8] While this statement appears to follow trivially from Eq. (13) – i.e., via application of Slutsky's Theorem – this is not true. In general, we can only use Eq. (13) to deduce Eq. (15), if σ_{t-} is a constant. In our paper, where σ_{t-} is a random variable, the definition of convergence in distribution does not support such a conclusion. We therefore prove in Appendix A that the convergence in Eq. (13) is in law stably, which is a stronger form of convergence that helps to recover this feature (the concept is explained in e.g. Jacod and Protter, 2012). Moreover, we allow for leverage effects and jumps. In both these directions, Theorem 1 extends Kristensen (2010).

estimate μ_{t-}^*, we can exploit its asymptotic distribution to form a test of the drift burst hypothesis, since a large t-statistic signals that the realized log-return is mostly induced by drift.[9]

The drift burst alternative is formalized by an exploding μ_t term. We note that the alternative is broad enough to allow σ_t to co-explode with the drift.

Theorem 2. *Assume that $\widetilde{X}$ is of the form:*

$$d\widetilde{X}_t = dX_t + \frac{c_{1,t}}{(\tau_{db} - t)^\alpha} dt + \frac{c_{2,t}}{(\tau_{db} - t)^\beta} dW_t, \tag{16}$$

where $\tau_{db} > 0$, dX_t is the model with locally bounded coefficients in Eq. (1) for which the conditions of Theorem 1 hold, $c_{1,t}$ and $c_{2,t}$ are adapted stochastic processes adhering to identical conditions as for μ_t and σ_t listed in Assumption 3. Moreover, α and β are constants such that $0 \leq \beta < 1/2$ and $0 < \alpha < 1$. Then, as $n \to \infty$, it holds that:

$$T_{\tau_{db}}^n \begin{cases} \overset{a.s.}{\to} \pm\infty, & \text{if } \alpha - \beta > 1/2, \\ \overset{d}{\to} c_{K,\beta} N(0, 1) + d_{K,\beta,c_1,c_2}, & \text{if } \alpha - \beta = 1/2, \\ \overset{d}{\to} c_{K,\beta} N(0, 1), & \text{if } \alpha - \beta < 1/2, \end{cases} \tag{17}$$

where

$$c_{K,\beta} = \sqrt{\frac{\int_{\mathbb{R}} K^2(x)|x|^{-2\beta} dx}{K_2 \int_{\mathbb{R}} K(x)|x|^{-2\beta} dx}} \quad \text{and} \quad d_{K,\beta,c_1,c_2} = \frac{c_{1,\tau_{db}}}{c_{2,\tau_{db}}} \frac{\int_{\mathbb{R}} K^2(x)|x|^{-\beta-1/2} dx}{\sqrt{K_2 \int_{\mathbb{R}} K(x)|x|^{-2\beta} dx}}. \tag{18}$$

Proof. See Appendix A. ■

Theorem 2 implies the explosion of the t-statistic under the alternative, when the drift term explodes fast enough relative to the volatility (i.e. $\alpha - \beta > 1/2$). The condition $\alpha - \beta > 1/2$ is equivalent to require that, in a neighborhood of τ_{db}, the log-return is dominated by drift. In a frictionless economy this allows for a short-lived arbitrage around τ_{db}. In practice, to make the test statistic large, it is of course enough that the mean log-return is significantly nonzero over a small time interval. Our simulations confirm that the condition $\alpha - \beta > 1/2$ is not needed to achieve power in small samples.[10]

An implication of Theorem 2 is that $T_{\tau_{db}}^n$ does not explode because of a volatility burst, even if it occurs without a drift burst. With $\alpha - \beta < 1/2$ the t-statistic is normally distributed in the limit with a controllable asymptotic variance that depends on the kernel shape and β. With a left-sided exponential kernel $K(x) = \exp(-|x|)$, for $x \leq 0$, adopted below, $c_{K,\beta} = 2^\beta \leq \sqrt{2}$. At the border, $\alpha - \beta = 1/2$, the t-statistic is bounded but can take arbitrarily large values as $\alpha \to 1$ and $\beta \to 1/2$, because $\lim_{\beta \to 1/2} d_{K,\beta,c_1,c_2} = \infty$.[11]

Thus, the large values of the test statistic frequently observed in the empirical application are neither explained by volatility bursts nor by jumps. We conclude that – in our setting – a significant t-statistic can only be induced by a drift explosion.[12] These theoretical statements are also corroborated by our simulation analysis.

Remark 1. In this paper, the alternative model is a drift explosion. It is left to future research to determine whether different mechanisms (e.g. pure-jump or Hawkes-type processes, endogenous market microstructure noise, or perhaps rough models) can explain the departure from the classical jump-diffusion Itô semimartingale framework that our test statistic puts forward.

3.1. Inference via the maximum statistic

The asymptotic theory asserts that T_t^n is standard normal in absence of a drift burst, whereas it grows arbitrarily large under the alternative, as we approach a drift burst time. It suggests that a viable detection strategy is to compute the t-statistic progressively over time and reject the null when $|T_t^n|$ gets significantly large. This leads to a multiple testing problem, which can cause size distortions, if the quantile function of the standard normal distribution is used to determine a critical value of T_t^n.

---

[9] Todorov and Tauchen (2014) build a related type of self-normalized statistic to test whether the price process is jump-diffusion or of pure-jump type. Their procedure has power against the pure-jump alternative. It is therefore important in future work to establish the properties of our procedure in a pure-jump setting to ensure it has size control in such models.

[10] As detailed in Appendix C, we can estimate α and β using a parametric maximum likelihood approach based on Eq. (10). The sample averages across detected events in the empirical high-frequency data analyzed in Section 5 are $\bar{\hat{\alpha}}_{ML} = 0.6250$ and $\bar{\hat{\beta}}_{ML} = 0.1401$. Hence, assuming that bursts are of the form in Eq. (10), the process is found to be right at the margin of being arbitrage-free.

[11] Hengartner and Linton (1996) study nonparametric regression estimation at design poles and zeros. A nearly identical set of "bias" and "variance" kernel constants to d_{K,β,c_1,c_2} and $c_{K,\beta}$ also appear in the asymptotic distribution of their estimator.

[12] A likelihood ratio test based on the model in Eq. (10), also reported in Appendix C, strongly rejects $\mathscr{H}_0' : \beta = 0$ against $\mathscr{H}_1' : \beta > 0$ during a drift burst, yielding empirical support for the volatility co-exploding with the drift.

To control the family-wise error rate, we evaluate a standardized version of the maximum of the absolute value of our t-statistic using extreme value theory.[13] We compute $(T_{t_i^*}^n)_{i=1}^m$ at m equispaced time points $t_i^* \in (0, T]$, where T is fixed. We set:

$$T_m^* = \max_{t_i^*} |T_{t_i^*}^n|, \quad i = 1, \ldots, m. \tag{19}$$

The crucial point is that, in addition to $T_{t_i^*}^n \xrightarrow{d} N(0, 1)$ under the null, the $T_{t_i^*}^n$'s are also independent – up to error terms that are asymptotically negligible – if m does not grow too fast. It follows that a normalized version of T_m^* has a limiting Gumbel distribution, as $m \to \infty$ at a suitable rate.

Theorem 3. *Assume that the conditions of Theorem 1 hold. Then, if $n \to \infty$, $m \to \infty$ such that $mh_n \to 0$ and $\sqrt{\log(m)}\left(\frac{m}{\sqrt{nh_n}} + (mh_n)^{-B}\sqrt{\log(m)} + m^{-\Gamma/2}\sqrt{\log(m)}\right) \to 0$, it further holds that:*

$$(T_m^* - b_m)a_m \xrightarrow{d} \xi, \tag{20}$$

where

$$a_m = \sqrt{2\log(m)}, \qquad b_m = a_m - \frac{1}{2}\frac{\log(\pi \log(m))}{a_m}, \tag{21}$$

and the CDF of ξ is the Gumbel, i.e. $P(\xi \le x) = \exp(-\exp(-x))$.

Proof. See Appendix A. ∎

The alternative to Theorem 3 is that there is at least one drift burst (as formalized in Theorem 2) in the time interval $[0, T]$, whence the maximum statistic T_m^* diverges rendering the test consistent.

In practice, the Gumbel distribution is conservative, as the convergence in Eq. (20) is slow and because of residual dependence in the t-statistics due to small sample effects, microstructure noise and pre-averaging (introduced in Section 3.3). In Appendix B, we propose a simulation-based procedure to determine data-driven critical values.

3.2. Robustness to a pre-announced jump

We here study an extended model that – on top of the drift, volatility and jump component in Eq. (1) – has a "pre-announced" jump (see e.g. Jacod et al., 2017; Dubinsky et al., 2018), where the jump time is predictable. We show these types of jumps do not compromise drift burst detection.

Theorem 4. *Assume that $\widetilde{\widetilde{X}}$ is of the form:*

$$d\widetilde{\widetilde{X}}_t = dX_t + dJ'_t, \tag{22}$$

where dX_t is the model in Eq. (1) such that the conditions of Theorem 1 hold, while $J'_t = J \cdot I_{\{0 < \tau_J \le t\}}$, τ_J is a stopping time, and J is $\mathscr{F}_{\tau_J}$-measurable. Then, as $n \to \infty$, it holds that:

$$T_{\tau_J}^n \xrightarrow{p} \sqrt{\frac{K(0)}{K_2}} \cdot \text{sign}(J). \tag{23}$$

Proof. See Appendix A. ∎

We can readily select a kernel that can tell apart the occurrence of a fixed jump from a drift explosion. In particular, the left-sided exponential kernel has $\sqrt{K(0)/K_2} = \sqrt{2}$, so that $|T_{\tau_J}^n| \xrightarrow{p} \sqrt{2}$. Thus, our proposed t-statistic is – asymptotically – small under the null (standard normal distributed) and pre-announced jump alternative ($\sqrt{2}$ in absolute value), while it is large (diverging) under the drift burst alternative.

3.3. Robustness to microstructure noise

In practice, we do not measure the true, efficient log-price X_{t_i}, because transaction and quotation data are disrupted by multiple layers of "noise" or "friction" (e.g. Black, 1986; Stoll, 2000). In this section, we show how to modify our test

---

[13] Bajgrowicz et al. (2016) and Lee and Mykland (2008) also exploit these ideas in the high-frequency framework to devise an unbiased jump-detection test, while in a related context Andersen et al. (2007) propose a Bonferroni correction. The latter was another viable tool to avoid systematic overrejection of the null hypothesis.

for drift bursts, so it is resistant to such features of the market microstructure at the tick level. To incorporate noise, we suppose that:

$$Y_{t_i} = X_{t_i} + \epsilon_{t_i}, \qquad \text{for } i = 0, 1, \ldots, n, \tag{24}$$

where ϵ_{t_i} is an additive error term.

Assumption 2. $(\epsilon_{t_i})_{i=0}^n$ is adapted and independent of X. Moreover, $E[\epsilon_{t_i}] = 0$, $E[\epsilon_{t_i}^4] < \infty$, and denoting the autocovariance function by $\gamma_k = E[\epsilon_{t_i}\epsilon_{t_{i+k}}]$ for any integer $k \geq 0$, we further assume γ_k is finite, independent of i and n, and such that $\gamma_k = 0$ for $k > Q$, where $Q \geq 0$ is an integer (i.e. Q-dependent noise).

This is a standard noise model in financial econometrics. It allows for autocorrelation in the noise process, but it rules out dependence between the noise and fundamental price, as in Li and Linton (2020). The difficulty brought by noise is then that in order to do inference about drift bursts in X, we are forced to work with the contaminated high-frequency record of Y.

A direct application of the t-statistic in Eq. (14) to the noise-contaminated returns $\Delta_i^n Y$ is powerless, because the noise asymptotically dominates the other shocks and overwhelms the signal of an exploding drift. A solution to this problem is to slightly slow down the accumulation of noise by pre-averaging Y_{t_i}, as in Jacod et al. (2009). We define a pre-averaged increment for any stochastic process V:

$$\Delta_i^n \overline{V} = \sum_{j=1}^{k_n-1} g_j^n \Delta_{i+j}^n V = -\sum_{j=0}^{k_n-1} H_j^n V_{t_{i+j}}, \tag{25}$$

where k_n is the pre-averaging window, $g_j^n = g(j/k_n)$ and $H_j^n = g_{j+1}^n - g_j^n$ with $g : [0,1] \mapsto \mathbb{R}$ continuous and piecewise continuously differentiable with a piecewise Lipschitz derivative g', such that $g(0) = g(1) = 0$ and $\int_0^1 g^2(s)ds < \infty$. Absent a drift burst, it follows that:

$$\Delta_i^n \overline{X} = O_p\left(\sqrt{\frac{k_n}{n}}\right) \qquad \text{and} \qquad \Delta_i^n \overline{\epsilon} = O_p\left(\frac{1}{\sqrt{k_n}}\right). \tag{26}$$

As Eq. (26) shows, the noise is reduced by a factor $\sqrt{k_n}$. The drift and volatility of X are enhanced by $\sqrt{k_n}$, while leaving their relative order unchanged. Intuitively, it therefore suffices with minimal pre-averaging to bring down the noise enough and make a fast drift burst with α close to one dominate the divergence of the asymptotic variance in the drift estimator.

The drift burst t-statistic in Eq. (14) is then refined as a noise-robust version, where the drift estimator and its long-run variance are computed from the pre-averaged return series:

$$\overline{T}_t^n = \sqrt{h_n}\frac{\hat{\overline{\mu}}_t^n}{\sqrt{\hat{\overline{\sigma}}_t^n}}, \tag{27}$$

with

$$\hat{\overline{\mu}}_t^n = \frac{1}{h_n}\sum_{i=1}^{n-k_n+2} K\left(\frac{t_{i-1}-t}{h_n}\right)\Delta_{i-1}^n \overline{Y}, \tag{28}$$

and

$$\hat{\overline{\sigma}}_t^n = \frac{1}{h_n'}\left[\sum_{i=1}^{n-k_n+2}\left(K\left(\frac{t_{i-1}-t}{h_n'}\right)\Delta_{i-1}^n \overline{Y}\right)^2 + 2\sum_{L=1}^{L_n} w\left(\frac{L}{L_n}\right)\sum_{i=1}^{n-k_n-L+2} K\left(\frac{t_{i-1}-t}{h_n'}\right)K\left(\frac{t_{i+L-1}-t}{h_n'}\right)\Delta_{i-1}^n \overline{Y}\Delta_{i-1+L}^n \overline{Y}\right], \tag{29}$$

where $w : \mathbb{R}_+ \to \mathbb{R}$ is a kernel with $w(0) = 1$ and $w(x) \to 0$ as $x \to \infty$, and L_n is the lag length that determines the number of autocovariances in Eq. (29). $\hat{\overline{\sigma}}_t^n$ is a heteroscedasticity and autocorrelation consistent (HAC)-type statistic (e.g. Newey and West, 1987; Andrews, 1991). The extra complexity is required to account for any noise dependence and the serial correlation induced by pre-averaging to consistently estimate the asymptotic variance of $\hat{\overline{\mu}}_t^n$.

Theorem 5. Set $Y_{t_i} = X_{t_i} + \epsilon_{t_i}$, where X is defined by Eq. (1) and ϵ is defined by Assumption 2. Suppose that Assumptions 1–5 are fulfilled. For every fixed $t \in (0, T]$, as $n \to \infty$, $k_n \to \infty$, $L_n \to \infty$ such that $k_n h_n \to 0$, $k_n h_n' \to 0$, $\frac{k_n}{nh_n} \to 0$, $\frac{k_n}{nh_n'} \to 0$, and $\frac{L_n}{nh_n'} \to 0$, it holds that:

$$\overline{T}_t^n \xrightarrow{d} N(0, 1).$$

Proof. See Appendix A. ∎

The conditions $k_n h_n \to 0$ and $\frac{k_n}{nh_n} \to 0$ (and the associated ones with h'_n) call for moderate pre-averaging, so that the number of pre-averaged terms is not too large. The condition $\frac{L_n}{nh'_n} \to 0$ means the lag length also cannot grow too fast when estimating the long-run variance of the drift estimator.

Theorem 6. *Set $Y_{t_i} = \widetilde{X}_{t_i} + \epsilon_{t_i}$, where $\widetilde{X}$ is defined as in Theorem 2 and everything else is maintained as in Theorem 5. As $n \to \infty$, $k_n \to \infty$, $L_n \to \infty$ such that $k_n h_n \to 0$, $k_n h'_n \to 0$, $\frac{k_n}{nh_n} \to 0$, $\frac{k_n}{nh'_n} \to 0$, and $\frac{L_n}{nh'_n} \to 0$, and $k_n h_n^{2(1-\alpha)} \to \infty$, it holds that $|\bar{T}^n_{\tau_{db}}| \overset{p}{\to} \infty$ for $\alpha - \beta > 1/2$.*

Proof. See Appendix A. ∎

This confirms that in presence of noise the pre-averaged test statistic converges in law to a standard normal under the null of no drift burst, while it again diverges at a drift burst time, if $\alpha - \beta > 1/2$. Under the alternative, pre-averaging cannot be too moderate ($k_n h_n^{2(1-\alpha)} \to \infty$), otherwise the signal of an exploding drift is not enhanced enough relative to the noise. The condition also shows that with α closer to 1, we need less pre-averaging.

4. Simulation study

In this section, we adopt a Monte Carlo approach to further explore the t-statistic proposed in Eq. (27) as a tool to uncover drift bursts in X. The overall goal is to investigate the size and power properties of our test and figure out how "small" drift bursts we are able to detect under the alternative, amid also an exploding volatility, the presence of infinity-activity price jump processes, and microstructure noise.

We simulate a driftless Heston (1993)-type stochastic volatility (SV) model:

$$\begin{aligned} \mathrm{d}X_t &= \sigma_t \mathrm{d}W_t, \\ \mathrm{d}\sigma_t^2 &= \kappa\left(\theta - \sigma_t^2\right)\mathrm{d}t + \xi\sigma_t \mathrm{d}B_t, \quad t \in [0,1], \end{aligned} \tag{30}$$

where W and B are standard Brownian motions with $E(\mathrm{d}W_t \mathrm{d}B_t) = \rho \mathrm{d}t$. Thus, the drift-to-volatility ratio of the efficient log-price is $\mu_t/\sigma_t = 0$.

We configure the variance process to match key features of real financial high-frequency data. As consistent with prior work (e.g. Aït-Sahalia and Kimmel, 2007), we assume the annualized parameters of the model are $(\kappa, \theta, \xi, \rho) = (5, 0.0225, 0.4, -\sqrt{0.5})$. We note θ implies an unconditional standard deviation of log-returns of roughly 15% p.a., which aligns with what we observe across assets in our empirical study. A total of 1,000 repetitions is generated via an Euler discretization. In each simulation, σ_t^2 is initiated at random from its stationary law $\sigma_t^2 \sim \text{Gamma}(2\kappa\theta\xi^{-2}, 2\kappa\xi^{-2})$. The sample size is $n = 23{,}400$, which is representative of the liquidity in the futures contracts analyzed in Section 5 (see Table 2). It corresponds to second-by-second sampling in a 6.5 hours trading session.

We create drift and volatility bursts with the parametric model:

$$\mu_t^{\mathrm{db}} = a\frac{\mathrm{sign}(t - \tau_{\mathrm{db}})}{|\tau_{\mathrm{db}} - t|^\alpha}, \qquad \sigma_t^{\mathrm{vb}} = b\frac{\sqrt{\theta}}{|\tau_{\mathrm{db}} - t|^\beta}, \quad \text{for } t \in [0.475, 0.525], \tag{31}$$

with $\tau_{\mathrm{db}} = 0.5$ fixed. Here, the price experiences a short-lived flash crash at τ_{db}, as consistent with our empirical finding that most of the identified drift bursts are followed by partial or full recovery.[14] The window $[0.475, 0.525]$ can be interpreted as making the duration of the bursts last about 20 minutes. We set $\alpha = (0.55, 0.65, 0.75)$ and $\beta = (0.1, 0.2, 0.3, 0.4)$ to gauge their impact on our t-statistic.[15] In particular, fixing $a = 3$ we induce a cumulative return $\int_0^{\tau_{\mathrm{db}}} \mu_t^{\mathrm{db}}\mathrm{d}t$ of about -0.5% (with opposite sign after the crash) for $\alpha = 0.55$ to slightly less than -1.5% for $\alpha = 0.75$, as comparable to what we observe in the real data. Also, with $b = 0.15$ our choices of β produce a 25% ($\beta = 0.1$) to more than 100% ($\beta = 0.4$) increase in the standard deviation of log-returns in the drift burst window relative to its unconditional level across simulations. A drift burst is therefore accompanied by highly elevated volatility, making it challenging to detect the signal.

To explicate the robustness of our t-statistic, in each simulation we superposition on top of the continuous sample path model in Eq. (30) a Lévy process with jump density given by:

$$\nu(\mathrm{d}x) = \psi\frac{e^{-\lambda x}}{x^{1+\upsilon}}\mathrm{d}x, \tag{32}$$

where $\upsilon > 0$ and $\psi > 0$. This defines a pure-jump tempered stable process with activity index υ. We assume $\upsilon = 0.5$, such that an infinite-activity finite-variation process with both small and large jumps appearing randomly in X

---

[14] To ensure X reverts during a pure volatility burst, we recenter the log-return series associated with σ_t^{vb}, so that $\int_0^T \sigma_t^{\mathrm{vb}}\mathrm{d}W_t = 0$. This has almost no impact on the outcome of the t-statistic, but it makes the price processes comparable across settings.

[15] Note that as $\alpha - \beta > 1/2$ for some of these combinations, the model is not always devoid of arbitrage.

is produced. We further set $\lambda = 3$ and calibrate ψ so that on average 20% of the quadratic variation is induced by the jump component. This is broadly in line with previous studies, e.g. Aït-Sahalia et al. (2012) and Aït-Sahalia and Xiu (2015). The process is simulated as the difference between two positive tempered stable processes, as outlined in Todorov et al. (2014). These are generated with the acceptance–rejection algorithm of Baeumer and Meerschaert (2010). Note that the discretization is exact for the selected value of υ. Also, in agreement with our theoretical exposition the jump process is active both under the null and alternative hypothesis.

The noisy log-price is:

$$Y_{i/n} = X_{i/n} + \epsilon_{i/n}, \quad i = 0, 1, \ldots, n, \tag{33}$$

where $\epsilon_{i/n} \sim N\left(0, \omega_{i/n}^2\right)$ and $\omega_{i/n} = \gamma \frac{\sigma_{i/n}}{\sqrt{n}}$, so the noise is both conditionally heteroscedastic, serially dependent (via σ), and positively related to the riskiness of the efficient log-price (e.g. Bandi and Russell, 2006; Oomen, 2006; Kalnina and Linton, 2008). We should note that this is a more general microstructure noise than allowed by Assumption 2. γ is the noise-to-volatility ratio. We set $\gamma = 0.5$, which amounts to a medium contamination level (e.g. Christensen et al., 2014). To reduce the noise, we pre-average $Y_{i/n}$ locally within a block of length $k_n = 3$ and based on the weight function $g(x) = \min(x, 1 - x)$.[16,17] $\hat{\bar{\mu}}_t^n$ and $\hat{\bar{\sigma}}_t^n$ are constructed from $(\Delta_i^n \bar{Y})_{i=0}^{n-2k_n+1}$ based on Eqs. (28) and (29) with a left-sided exponential kernel $K(x) = \exp(-|x|)$, for $x \leq 0$.

A Parzen kernel is selected for w:

$$w(x) = \begin{cases} 1 - 6x^2 + 6|x|^3, & \text{for } 0 \leq |x| \leq 1/2, \\ 2(1 - |x|)^3, & \text{for } 1/2 < |x| \leq 1, \\ 0, & \text{otherwise.} \end{cases} \tag{34}$$

This choice has some profound advantages in our framework. First, the Parzen kernel ensures that $\hat{\bar{\sigma}}_t^n$ is positive, so we can always compute the t-statistic, which is not true for a general weight function. Secondly, the efficiency of the Parzen kernel is near-optimal, e.g. Andrews (1991) and Barndorff-Nielsen et al. (2009). The slight loss of efficiency brings the distinct merit that $\hat{\bar{\sigma}}_t^n$ can be computed on the back of the first L_n lags of the autocovariance function, whereas more efficient weight functions typically require all n lags. In the high-frequency framework n is large, so the latter can be prohibitively slow to compute. In contrast, L_n is typically small compared to n in practice, rendering our choice of kernel much less time-consuming.

We set $L_n = Q^* + 2(k_n - 1)$ to estimate $\hat{\bar{\sigma}}_t^n$. Here, $2(k_n - 1)$ is due to pre-averaging and we compute Q^* from $(\Delta_i^n Y)_{i=1}^n$ as a data-driven measure of noise dependence based on automatic lag selection (see, e.g. Newey and West, 1994; Barndorff-Nielsen et al., 2009).[18] The bandwidth for $\hat{\bar{\mu}}_t^n$ is varied in $h_n = (120, 300, 600)$ seconds. We use a larger bandwidth of $h_n' = 5h_n$ for $\hat{\bar{\sigma}}_t^n$ to better capture persistence in volatility and estimate the microstructure-induced return variation.

A new value of $\bar{T}_t^n$ is recorded at every 60th transaction update.[19] We extract $T_m^* = \max_{i=1,\ldots,m} |\bar{T}_{t_i^*}^n|$ based on the resulting $m = 341$ tests in each sample. The simulation-based approach explained in Appendix B is adopted to find a critical value of T_m^*.

Fig. 3 reports Q–Q plots of the distribution of $\bar{T}_t^n$ in absence of a drift burst. Panel A is based on the Heston (1993)-type SV+jump model. As readily seen, the Gaussian curve is an accurate description of the sampling variation of $\bar{T}_t^n$, although the t-statistic is slightly thin-tailed with $h_n = 120$. This is caused by a modest correlation between $\hat{\bar{\mu}}_t^n$ and $\hat{\bar{\sigma}}_t^n$, which are computed from partly overlapping data; an effect that is more pronounced for small bandwidths. In Panel B, the process featuring a large volatility burst ($\beta = 0.4$) is plotted.[20] While the volatility burst puts more mass into the tails of the distribution of $\bar{T}_t^n$, the normal continues to be a good approximation.

This is corroborated by Table 1, where we compute how often T_m^* leads to rejection of the null hypothesis of no drift burst for three significance levels $c = 5\%, 1\%, 0.5\%$. There are several interesting findings. Look at the columns with $\mu_t^{db} \equiv 0$, which report the results in absence of a drift burst (i.e., size). Without a volatility burst ($\beta = 0.0$), the test is conservative compared to the nominal level if h_n is small, as also reflected in Fig. 3. As β increases, T_m^* is mildly inflated yielding a tiny size distortion, but the effect is only present for large β and h_n. Otherwise, the test is roughly unbiased. This suggests our t-statistic is adaptive and highly robust to even substantial shifts in spot variance, so that we do not falsely pick up an explosion in volatility as a significant drift burst.

---

[16] In the Online Appendix, we present a comprehensive analysis with $\gamma = 0.5, 2$ and 5 and pre-averaging horizon $k_n = 1, \ldots, 10$. The results do not differ materially from those reported here. A modest loss of power is noted, however, if h_n is small and k_n is large.

[17] With equidistant data, it follows that if k_n is even and $g(x) = \min(x, 1 - x)$, the pre-averaged return in Eq. (25) can be rewritten as $\Delta_i^n \bar{Y} = \frac{1}{k_n} \sum_{j=1}^{k_n/2} Y_{\frac{i+k_n/2+j}{n}} - \frac{1}{k_n} \sum_{j=1}^{k_n/2} Y_{\frac{i+j}{n}}$. Thus, the sequence $(2\Delta_i^n \bar{Y})_{i=1}^{n-k_n+2}$ can be interpreted as constituting a new set of increments from a price process that is constructed by averaging of the rescaled noisy log-price series, $(Y_{i/n})_{i=0}^n$, in a neighborhood of i/n, thus making the use of the term pre-averaging and the associated notation transparent.

[18] In our simulations, the average value of Q^* is 11.9, while its interquartile range is 8–17.

[19] We set a burn-in period of a full volatility bandwidth before testing to allow for a sufficient number of observations to construct $\bar{T}_t^n$.

[20] The results for other values of β fall in between those of Panel A and B and are therefore not reported.

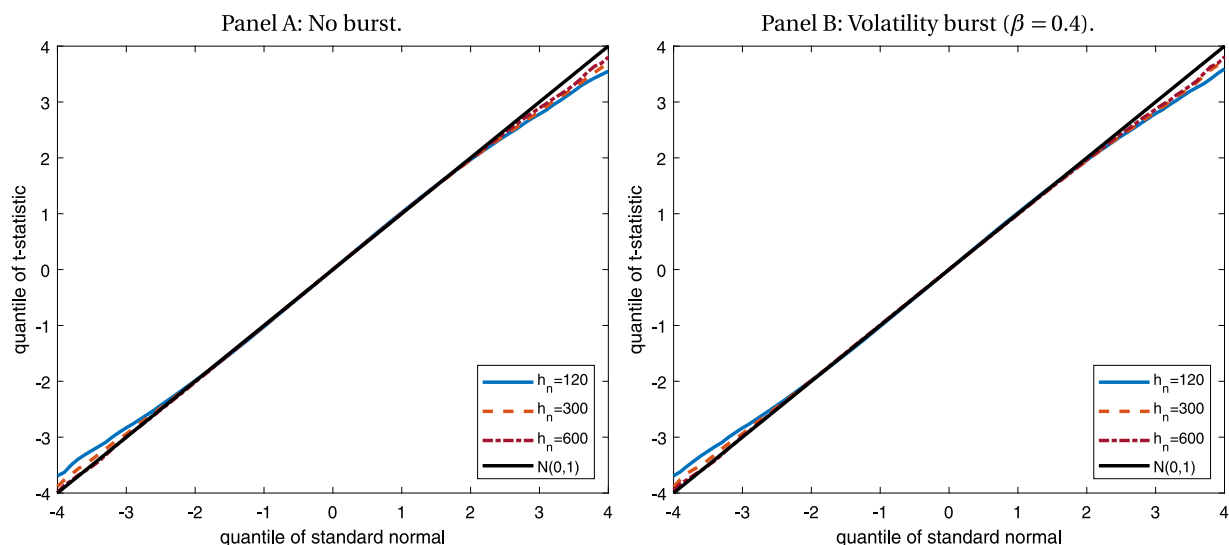Panel A: No burst. Panel B: Volatility burst ($\beta = 0.4$).

Fig. 3. Q–Q plot of $\bar{T}_t^n$ without drift burst. *Note.* We present a Q–Q plot of $\bar{T}_t^n$ under the null hypothesis of no drift burst. Panel A is from the Heston (1993)-type SV+jump model with no burst in neither drift nor volatility, while Panel B adds a volatility burst using the parametric model in Eq. (31) with $b = 0.15$ and $\beta = 0.4$.

Table 1
Size and power of drift burst t-statistic T_m^* with $\gamma = 0.5$ and $k_n = 3$.

		$\Pr(T_m^* > q_{0.950})$				$\Pr(T_m^* > q_{0.990})$				$\Pr(T_m^* > q_{0.995})$			
		vb (size)	db (power)			vb (size)	db (power)			vb (size)	db (power)		
		$\mu_t^{db} \equiv 0$	$\alpha =$	0.55	0.65	0.75	$\mu_t^{db} \equiv 0$	$\alpha =$	0.55	0.65	0.75	$\mu_t^{db} \equiv 0$	$\alpha =$	0.55	0.65	0.75

		vb (size) $\mu_t^{db} \equiv 0$	db (power) $\alpha =$ 0.55	0.65	0.75	vb (size) $\mu_t^{db} \equiv 0$	db (power) $\alpha =$ 0.55	0.65	0.75	vb (size) $\mu_t^{db} \equiv 0$	db (power) $\alpha =$ 0.55	0.65	0.75
Panel A: $h_n = 120$													
$\beta =$	0.0*	0.8	44.9	83.0	98.0	0.1	29.0	66.6	90.7	0.0	23.9	58.1	85.8
	0.1	0.9	27.9	75.7	96.8	0.1	13.8	53.4	87.6	0.0	9.9	44.4	80.6
	0.2	1.0	18.2	66.4	94.9	0.3	6.4	44.1	84.4	0.0	4.2	35.2	75.9
	0.3	0.9	9.5	48.9	92.1	0.3	2.1	26.9	76.7	0.1	1.2	19.1	64.0
	0.4	0.9	5.4	25.3	80.5	0.2	0.8	8.2	51.2	0.0	0.4	4.6	38.6
Panel B: $h_n = 300$													
$\beta =$	0.0*	2.1	51.6	86.7	98.9	0.6	39.7	78.8	96.9	0.3	33.6	74.5	95.7
	0.1	2.0	41.8	82.4	98.5	0.6	29.2	71.8	95.8	0.3	24.8	67.1	93.7
	0.2	2.1	36.0	78.3	98.1	0.5	21.9	66.3	94.1	0.3	18.5	60.1	92.5
	0.3	2.6	26.1	70.3	96.7	0.7	13.2	54.4	92.4	0.3	10.5	48.9	90.2
	0.4	4.1	16.7	52.5	93.4	0.9	7.9	34.6	84.4	0.5	5.5	27.9	79.2
Panel C: $h_n = 600$													
$\beta =$	0.0*	4.6	46.4	82.6	98.0	1.0	33.8	73.3	95.6	0.4	30.0	68.8	94.5
	0.1	4.2	40.1	79.4	97.6	1.0	28.6	68.6	94.6	0.3	25.2	63.7	93.1
	0.2	4.4	36.4	77.1	97.2	1.0	24.6	64.7	93.8	0.2	21.5	60.2	92.6
	0.3	4.7	30.3	71.1	96.0	1.1	19.0	57.4	92.2	0.3	14.0	52.2	90.1
	0.4	5.8	23.2	57.9	93.9	1.8	11.9	42.6	87.6	0.9	9.2	37.3	83.6

Note. $\Pr(T_m^* > q_{1-c})$ is the rejection rate (in percent, across Monte Carlo replications) of the drift burst t-statistic T_m^* defined in Eq. (19), where q_{1-c} is a simulated $(1 - c)$-level quantile from the finite sample extreme value distribution of T_m^* under the null of no drift burst, as explained in Appendix B. α is the explosion rate of the drift burst (db), while β is the explosion rate of the volatility burst (vb). *$\beta = 0.0$ represents the Heston (1993)-type SV+jump model with no volatility burst. h_n is the bandwidth of $\hat{\bar{\mu}}_t^n$ (measured as effective sample size), while the bandwidth of $\hat{\bar{\sigma}}_t^n$ is $5h_n$. γ is the level of noise-to-volatility (per increment) and k_n is the pre-averaging horizon.

Turn next to the alternative with a drift burst (i.e., power). As expected, the power is increasing in α, holding β fixed, while it is decreasing in β, holding α fixed. In general, the test has decent power and is capable of identifying a true explosion in the drift coefficient, except those causing a minuscule cumulative log-return and that are coupled with a large volatility burst. While the test has excellent ability to discover the largest drift bursts, which from a practical point of view are arguably also the most important, it is intriguing that we can uncover many of the smaller ones as well. At last, higher values of h_n improve the rejection rate under the alternative, but the marginal gain of going from $h_n = 300$

Table 2
CME futures data.

Code	Name	# days	Volume		# quote	Inside	Sub-sample retained	
			# contracts	Notional ($bn)	Updates	Spread (bps)	By volume	By quotes
6E	Euro FX	1536	209,281	31.9	62,357	0.69	92.87%	87.25%
CL	Crude oil	1545	299,821	18.8	69,580	1.70	95.58%	92.28%
ES	E-mini S&P500	2053	1,763,100	145.5	27,513	1.54	97.33%	91.94%
GC	Gold	1544	162,056	22.0	56,562	0.97	88.14%	84.59%
ZC	Corn	1528	112,612	2.6	4870	5.89	87.23%	71.18%
ZN	10-Year T-Note	1542	1,121,122	112.1	6372	1.22	94.33%	89.10%

Note. This table reports for each futures contract, the number of days in the sample, the average daily volume by number of contracts and notional traded, the average daily number of top-of-book quote updates, and the average daily median spread in basis points calculated from 09:00–10:00 Chicago time. The sample period is January 2012–December 2017 for all contracts, except ES where we start in January 2010. In the empirical analysis, we restrict attention to the most active trading hours from 01:00–15:15 Chicago time for all contracts, except for ZC where the interval is restricted to 08:30–13:20 Chicago time. The fraction of volume and quote updates retained after removing the most illiquid parts of the day is reported in the last two columns.

to $h_n = 600$ is negligible. This suggests – on the one hand – that h_n should not be too narrow, as it erodes the power, while – on the other – it should neither be too wide, as this creates a small size distortion.

5. Drift bursts in financial markets

5.1. Data

We apply the drift burst test statistic to a comprehensive set of intraday tick data, covering a broad range of financial assets. We use trades and quotes with milli-second precision timestamps for futures contracts traded on the Chicago Mercantile Exchange (CME). We select the most actively traded futures contract for each of the main assets classes, namely the Euro FX for currencies (6E), Crude oil for energy (CL), the E-mini S&P500 for equities (ES), Gold for precious metals (GC), Corn for agricultural commodities (ZC), and the 10-Year Treasury Note for rates (ZN). These futures contracts are amongst the most liquid financial instruments in the world. To illustrate, the average daily notional volume traded in just a single ES contract on the CME is comparable to the trading volume of the entire U.S. cash equity market covering over 5000 stocks traded across more than ten different exchanges.[21] The sample period is January 2012–December 2017, except we backdate ES to January 2010 in order to capture the May 2010 flash crash. While the CME is open nearly all day, we restrict attention to the more liquid European and U.S. trading sessions: from 01:00–15:15 Chicago time or 07:00–21:15 London time. The only exception is Corn, where we use data from 08:30–13:20 Chicago time. Outside of these hours, trading is minimal in this contract. Table 2 provides informative summary statistics of the data.

5.2. Implementation of test

We construct for each series a mid-quote as the average of the best bid and offer available at any point in time. A quote update is retained if the mid-quote changes. The remaining data are pre-averaged with $k_n = 3$.[22] We then calculate the drift burst test statistic on a regular five-second grid and only include values that are preceded by a mid-quote revision. As our primary interest is to identify short-lived drift bursts, we continue with a five-minute bandwidth for the drift. We base the spot volatility on a 25-minute bandwidth with the Parzen kernel and $L_n = 2(k_n - 1) + 10$ lags for the HAC-correction. As in the simulations, a left-sided exponential kernel is adopted: $K(x) = \exp(-|x|)$, for $x \leq 0$.[23]

Table 3 reports selected descriptive measures of the calculated drift burst t-statistics over the full sample. Judging by the standard deviation and kurtosis of the test statistic, the distribution is close to standard normal, as consistent with the asymptotic theory under the null of no drift burst. This is remarkable, because the test is applied to relatively short intraday intervals across a wide range of asset classes and is therefore exposed to substantial changes in liquidity conditions, diurnal effects, or to futures contracts where the minimum price increment – and hence the microstructure noise – is relatively large (e.g. ZN). Concentrating on the tails, we identify a large number of drift bursts.[24] At a critical value of 4.5, for instance, there are 202 drift bursts in the E-mini S&P500 futures, or about one every two weeks. They are more prevalent in Euro FX, Gold, and Oil contracts but less frequent in the Treasury and Corn futures. The number of expected false positives (computed as described in Appendix B) is practically zero.

---

[21] See https://batstrading.com/market_summary/ for daily U.S. equity market volume statistics.

[22] The Online Appendix contains the empirical analysis with $k_n = 1$ (i.e., no pre-averaging), 5 and 10. The results are broadly speaking in line with those reported here.

[23] In practice, a backward-looking kernel is preferred to allow for real-time updating. Moreover, it enhances the power for testing the drift burst hypothesis. The intuition is that after a drift burst, the pre-dominant tendency of the price to reverse combined with a high and persistent level of volatility, delivers a lower value of the test statistic if a two-sided kernel is employed.

[24] To account for the rolling calculation of the test statistic and avoid double counting of events, we allow for at most one drift burst to be established over any five-minute window at which the test statistic attains a local extremum and exceeds a set critical value.

Table 3
Drift burst test summary statistics.

Code	Name	Empirical distribution			# of identified drift bursts				
		σ	σ_q	kurtosis	$|T| > 4.0$	>5.0	>5.0	>5.5	>6.0
6E	Euro FX	1.04	1.05	3.6	1329	615	292	145	64
CL	Crude oil	1.04	1.07	3.8	1606	667	300	127	51
ES	E-mini S&P500	1.07	1.07	3.1	549	202	78	29	11
GC	Gold	1.02	1.04	3.7	1324	542	227	83	35
ZC	Corn	1.15	1.13	2.9	97	29	15	5	2
ZN	10-Year T-Note	1.22	1.10	2.4	105	46	21	10	4

Note. This table reports for each futures contract, the standard deviation and kurtosis of the empirical drift burst t-statistic. We calculate the test every five seconds across the sample if there was a mid-quote update over that interval. The standard deviation is also calculated by rescaling the 5/95-percentile of the empirical distribution by that of a standard normal ("σ_q"). The number of drift bursts identified for critical values ranging between 4 and 6 is reported. The number of false positives we expect, which can be computed using the technique described in Appendix B, is virtually zero.

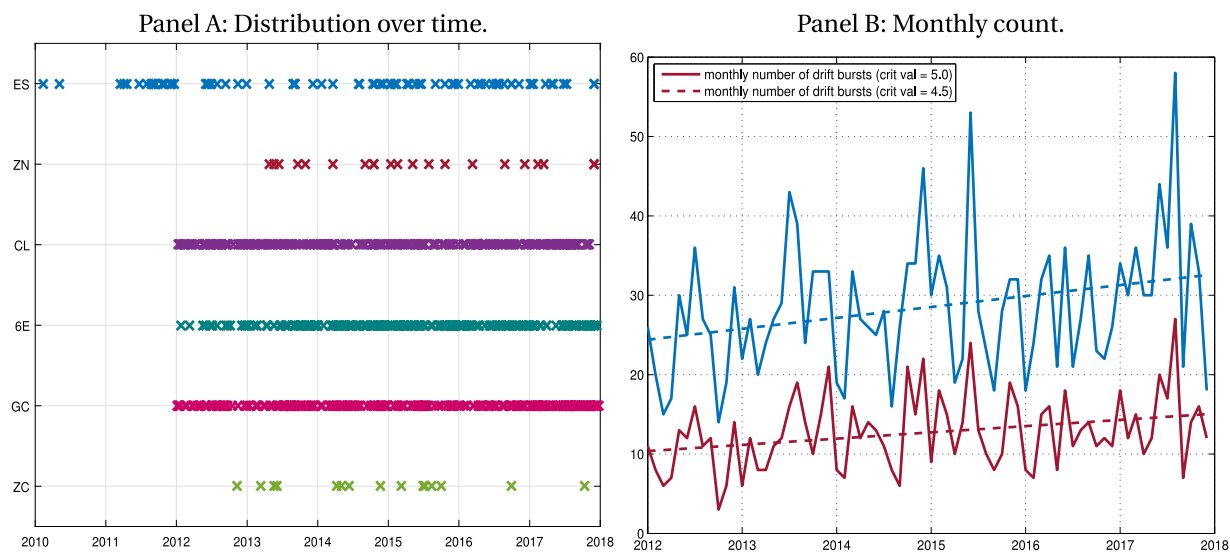**Fig. 4.** Time series of drift bursts. *Note.* In Panel A, we plot the cross-sectional distribution of drifts bursts across asset classes and over time (each cross represents a day with $|T_m^*| > 5.0$), while Panel B shows the associated number of significant daily events aggregated to a monthly level.

Panel A of Fig. 4 indicates the location of drift bursts for the various securities, while Panel B reports the pooled monthly counts. The results lend support to the perception that flash crashes are increasing over time. A time trend is positive for each asset but lacks statistical significance for some due to the limited number of observations. The pooled results, however, indicate a statistically significant time trend of 5%–10% per annum in the number of identified drift bursts with a regression t-statistic of 2.5 for both scenarios drawn in Panel B.

Fig. 5 shows some examples of single-asset drift bursts identified by the test. A multi-asset drift burst is presented in Fig. 6, which plots the evolution of the six asset prices during the Twitter hoax flash crash of April 23, 2013. The figures show that drift bursts are a stylized feature of the price process and, in some instances, systemic to the market. It is evident that neither jumps nor volatility are driving the price dynamics observed in these examples. The drift burst hypothesis is a plausible alternative to model the data.

5.3. Reversion after the drift burst

We now investigate in more detail whether the mean reversion experienced during the flash crashes in Figs. 1 and 6 is more broadly associated with the price dynamics of a drift burst. Let $\{t_j\}$ denote the set of time points, where drift bursts are identified. The start of a drift burst is set to five minutes before t_j.[25] We sample the mid-quote process at this frequency pre- and post-drift burst and calculate:

$$R_{t_j}^- = X_{t_j} - X_{t_j-5m} \quad \text{and} \quad R_{t_j}^+ = X_{t_j+5m} - X_{t_j}, \tag{35}$$

---

[25] The 5-minute frequency is arbitrary, but often used in practice and consistent with our bandwidth. As a robustness check, we also applied an endogenous event window, where the duration of the drift burst is defined relative to the latest point in time prior to t_j, where the absolute value of the t-statistic is below one. The results are in line with those we report here and are available at request.
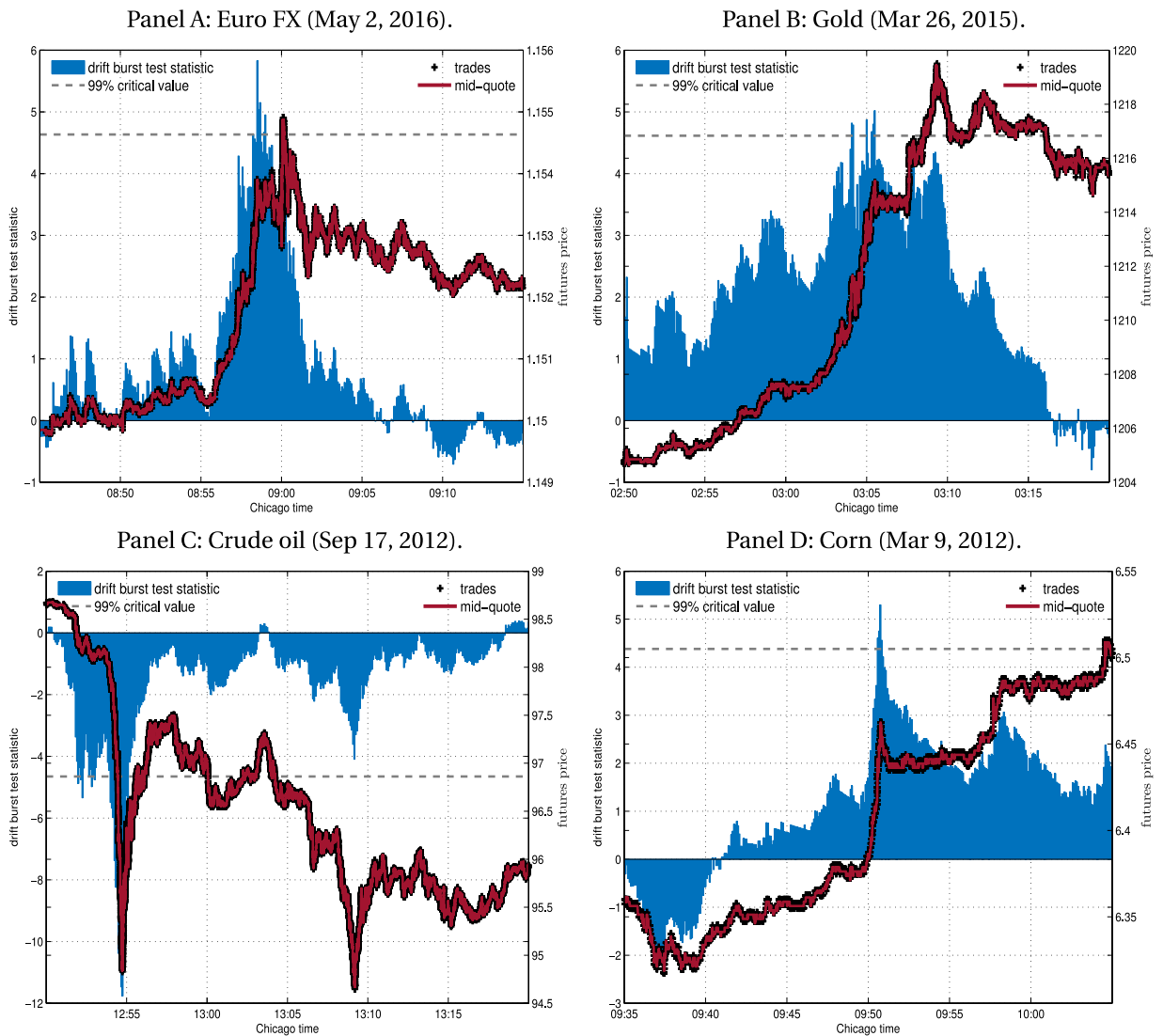
Fig. 5. Drift burst examples. *Note.* This figure draws, for some identified drift bursts, the sample path of the mid-quote and traded price (right axis) together with the *t*-statistic (left axis) over a 30-minute window that includes the peak of the drift burst.

where $R_{t_j}^-$ is the five-minute log-return during the jth drift burst, while $R_{t_j}^+$ is the corresponding post-drift burst log-return. In Fig. 7, we plot $R_{t_j}^+$ against $R_{t_j}^-$ pooled across the various asset markets. We observe that drift bursts can be associated with both positive and negative returns, but most of them are reversals and the percentage of "gradual jumps" is roughly one third.

To gauge the magnitude of the reversion and evaluate whether drift bursts are subject to short-term return predictability, we run the following regression:

$$R_{t_j}^+ = a + bR_{t_j}^- + \epsilon_{t_j}. \tag{36}$$

A value of b different from zero indicates predictability conditional on a drift burst, and $b < 0$ means a drift burst tends to be followed by a retracement of the price.

In Table 4, we show the regression results for each futures contract separately and with all assets pooled together. We find a strong mean reversion with estimates of b that are negative and highly significant (the intercept estimate is insignificant and omitted). This is consistent across asset classes, also for those with a relatively small number of observations. The regression R^2 tends to lie in the 25%–40% range, with a few exceptions, and is about 30% on average. This suggests a substantial predictive power of the pre-drift burst return to forecast the post-drift burst return. The fraction of reversals, as measured by counting the relative occurrence where $R_{t_j}^+$ and $R_{t_j}^-$ are of opposite sign, rarely drops below
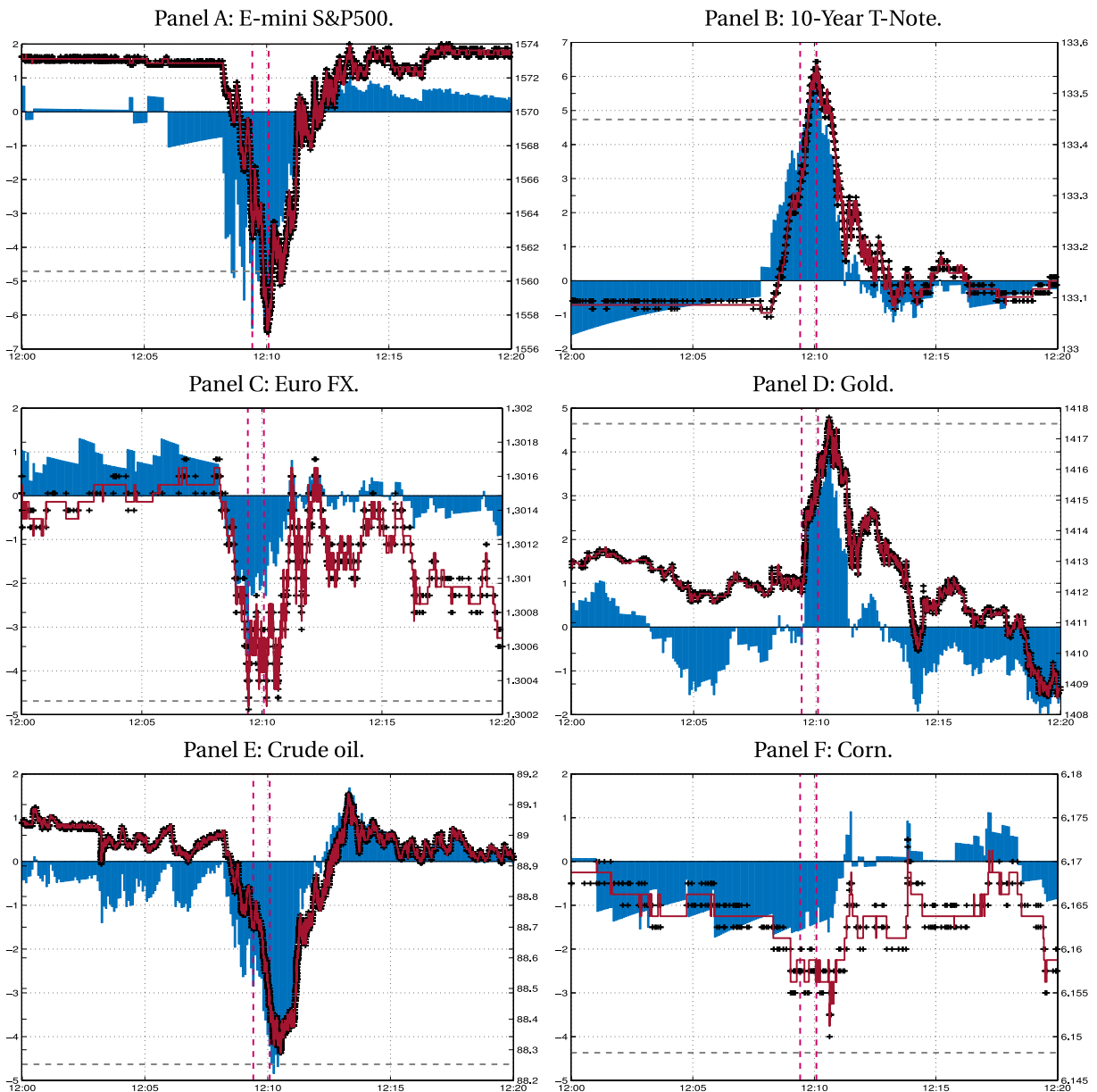
Fig. 6. The Twitter hoax of April 23, 2013: A synchronous flash event. *Note.* This figure draws the mid-quote and traded price (right axis) together with the *t*-statistic (left axis) for our six asset classes over a 20-minute window around the Twitter flash crash of April 23, 2013. At 12:07 p.m. Chicago time, Associated Press' Twitter account was hacked and a fake news was released, suggesting there had been explosions in the White House and that President Obama was injured. This led to perilous but short-lived sell-offs in the S&P500, Euro FX and Crude oil futures contracts, while the 10-Year T-Note and Gold contracts rallied. Only Corn was largely unaffected.

65%. As a robustness check, we remove the top decile of the most significant drift bursts and rerun the regression. The outcome is in the right-hand side of the table. As expected, the regression R^2 drops but the finding remains overly evident in the data with frequent reversals and significant mean reversion. The critical value is set to 4.0 and 4.5 here, but the results do not change much for larger and more conservative values.

5.4. Asymmetric reversal, trading volume and liquidity

Grossman and Miller (1988) suggest a large drift-to-volatility can be caused by exogenous demand for immediacy. In this framework, the return reversal represents a premium paid to market makers supplying liquidity against one-sided order flow (e.g. Nagel, 2012, also shows that returns on short-term reversal strategies can be thought of as a liquidity
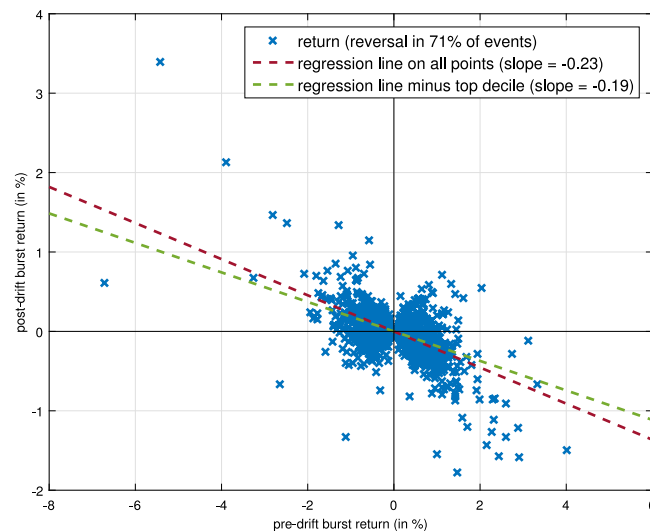
Fig. 7. Reversion in the drift burst. *Note.* This figure draws on the horizontal axis the percentage price move from the start to the peak of a drift burst against the subsequent price move (as defined by $R_{t_j}^+$ and $R_{t_j}^-$ in Eq. (35)) on the vertical axis. The horizon is five minutes. The number of drift bursts in each quadrant are Q1: 324, Q2: 670, Q3: 279 and Q4: 828.

Table 4
Reversion in the drift burst.

Code	All					Without top decile			
	#	b	R^2	%R		#	b	R^2	%R
Panel A: Critical value = 4.0									
ES	549	−0.26 (−12.05)	21.0%	67.2%		494	−0.14 (−5.59)	6.0%	66.8%
ZN	105	−0.25 (−5.30)	21.3%	65.7%		95	−0.16 (−3.09)	9.2%	64.2%
6E	1329	−0.22 (−21.05)	25.0%	68.2%		1196	−0.21 (−19.64)	24.4%	68.1%
GC	1324	−0.19 (−20.65)	24.4%	78.1%		1192	−0.18 (−18.36)	22.1%	78.2%
CL	1606	−0.21 (−29.33)	34.9%	71.3%		1445	−0.18 (−24.30)	29.0%	70.6%
ZC	97	−0.20 (−8.25)	41.5%	76.3%		87	−0.21 (−8.05)	43.0%	77.0%
ALL	5010	−0.21 (−46.62)	30.3%	71.8%		4509	−0.18 (−38.40)	24.6%	71.6%
Panel B: Critical value = 4.5									
ES	202	−0.36 (−14.45)	51.0%	70.8%		182	−0.20 (−8.12)	26.7%	69.8%
ZN	46	−0.29 (−3.89)	25.1%	71.7%		41	−0.13 (−1.76)	7.2%	70.7%
6E	615	−0.23 (−15.65)	28.5%	66.3%		554	−0.23 (−15.31)	29.8%	65.0%
GC	542	−0.17 (−11.86)	20.6%	77.9%		488	−0.15 (−9.57)	15.8%	77.3%
CL	667	−0.23 (−20.51)	38.7%	70.0%		600	−0.20 (−16.02)	30.0%	68.7%
ZC	29	−0.16 (−4.61)	43.1%	86.2%		26	−0.16 (−3.98)	38.8%	84.6%
ALL	2101	−0.23 (−34.40)	36.0%	71.3%		1891	−0.19 (−26.91)	27.7%	70.2%

Note. This table reports for each security, the number of identified drift bursts (#) using a critical value of 4.0 and 4.5, the estimated slope coefficient b of Eq. (36) with its associated t-statistic in parenthesis, the regression R^2, and the probability of reversion (%R) calculated as the fraction of drift bursts, where the price direction after the t-statistic peaks is opposite in sign compared to that in the run-up. To confirm robustness of the results, the right-hand part of the table removes the decile of the strongest drift bursts as measured by the absolute value of the t-statistic.

signature). The empirical results in Section 5.3 align with this interpretation. Their model, however, is symmetric in the order imbalance, as expressed by Eq. (2).

Huang and Wang (2009) show how trading imbalances arise endogenously due to costly market presence. Such a mechanism always leads to selling pressure, which tends to attenuate rallies and exacerbate sell-offs, resulting in market crashes absent news on fundamentals (Result 1–3 in their paper). As in Grossman and Miller (1988), the model induces negative serial correlation in observed returns (Result 4a), but it is asymmetric. Negative returns exhibit stronger

Table 5
ES return and volume dynamics.

Conditional information	$E[V_{t_j}^-]$	$E[R_{t_j}^-]$	$E[R_{t_j}^+]$	ρ	σ_p
Panel A: Unconditional					
All	1.00	0.03	0.02	−1.81%	9.41
$R_{t_j}^- > E[R_{t_j}^-]$	1.06	4.93	−0.08	0.05%	9.99
$R_{t_j}^- < E[R_{t_j}^-]$	0.96	−3.51	0.10	−2.27%	8.97
$V_{t_j}^- > E[V_{t_j}^-]$	1.54	−0.01	0.01	−1.93%	12.15
$V_{t_j}^- < E[V_{t_j}^-]$	0.46	0.06	0.04	−1.34%	5.42
Panel B: Negative drift burst					
All	3.53	−37.39	8.01	−79.58%	40.15
$R_{t_j}^- > E[R_{t_j}^-]$	2.70	−23.23	4.17	−15.90%	24.78
$R_{t_j}^- < E[R_{t_j}^-]$	5.46	−70.34	16.94	−85.53%	75.95
$V_{t_j}^- > E[V_{t_j}^-]$	6.22	−55.25	13.29	−84.26%	59.57
$V_{t_j}^- < E[V_{t_j}^-]$	1.94	−26.83	4.89	−53.52%	28.67
Panel C: Positive drift burst					
All	4.22	36.41	−5.44	−25.73%	38.29
$R_{t_j}^- > E[R_{t_j}^-]$	6.07	56.42	−9.02	−15.03%	59.58
$R_{t_j}^- < E[R_{t_j}^-]$	3.09	24.15	−3.25	−17.96%	25.25
$V_{t_j}^- > E[V_{t_j}^-]$	8.17	47.32	−8.52	−18.94%	50.26
$V_{t_j}^- < E[V_{t_j}^-]$	2.34	31.22	−3.98	−24.87%	32.60

Note. This table reports descriptive measures of the pre- and post-drift burst return, $R_{t_j}^-$ and $R_{t_j}^+$, and a normalized volume measure $V_{t_j}^-$. $E[\cdot]$ refers to the average taken across the entire sample (unconditional in Panel A) or over the number of identified positive and negative drift bursts (Panel B and C). ρ is the correlation between $R_{t_j}^-$ and $R_{t_j}^+$, while $\sigma_p = \sqrt{(R_{t_j}^-)^2 + (R_{t_j}^+)^2}$ is the volatility in the drift burst window. Returns are expressed in basis points (bps).

correlation than positive returns (Result 4b); volatility is higher during a negative return and high volume event (Result 5); abnormal volume implies larger future returns and stronger reversion (Result 6); and negative returns accompanied by high volume exhibit stronger reversals (Result 7). We here conduct an empirical assessment to check whether our sample of pre- and post-drift burst returns are consistent with these testable implications. We restrict attention to the E-mini S&P500 futures contract ES (Huang and Wang, 2009, also base their analysis on a stock paying a risky dividend). The results for other assets are available in the Online Appendix.

We construct a 5-minute grid as above. We then compute for each interval j:

- $R_{t_j}^-$ and $V_{t_j}^-$: the 5-minute log-return and traded volume,[26]
- $R_{t_j}^+$: the subsequent 5-minute log-return,
- ρ: correlation between $R_{t_j}^-$ and $R_{t_j}^+$,
- $\sigma_p = \sqrt{(R_{t_j}^-)^2 + (R_{t_j}^+)^2}$: a volatility measure.

Huang and Wang (2009) report $R_{t_j}^-$, $R_{t_j}^+$, $V_{t_j}^-$, ρ and σ_p based on *simulated* data.[27] In contrast, Table 5 – shaped as Table 1 in that paper for comparison – is computed from *actual* market returns. The full sample results in Panel A are calculated from all available 5-minute intervals. We observe a tiny negative return autocorrelation, which is slightly stronger if the preceding return is less than average or volume is above average. Volatility increases marginally with volume. The predicted effects are thus present, but extremely weak.

We now condition on t_j being a drift burst time (with critical value set to 4.0), allowing at most for one significant event per day. Grouping by the sign of $R_{t_j}^-$, we are left with 253 negative and 208 positive drift bursts. In Panel B–C of Table 5, we tabulate the corresponding averages. The results are striking. A drift burst of either sign is associated with high volatility, large volume, and substantial negative serial correlation. The typical negative drift burst entails a price

---

[26] We construct a normalized gross trading volume measure, which is defined by taking the gross trading volume in notional value and normalizing it by the average trading rate at that time of the day. This allows to depurate our volume series for the pronounced intraday swings in trading intensity, which makes the analysis more robust to time-of-the-day effects. Still, the results based on raw gross trading volume are broadly consistent with what we report here.

[27] They employ the notation $R_{1/2}$ for $R_{t_j}^-$, R_1 for $R_{t_j}^+$, and $V_{1/2}$ for $V_{t_j}^-$.

Table 6
ES post-drift burst return conditional on pre-drift burst return and volume.

	Negative drift burst				Positive drift burst			
Sorting variable	$V_{t_j}^-$							
$R_{t_j}^-$	Low	Medium	High	High-low	Low	Medium	High	High-low
Low	14.3	8.5	42.8	28.4	−0.9	−3.0	−2.7	−1.8
Medium	5.1	4.9	6.8	1.7	−3.1	−5.2	−4.5	−1.4
High	1.8	4.0	0.5	−1.3	−8.1	−12.0	−9.6	−1.5

Note. We report the average post-drift burst return $R_{t_j}^+$ after double-sorting first by pre-drift burst return $R_{t_j}^-$ and second by normalized gross trading volume $V_{t_j}^-$. As in Huang and Wang (2009), the sample is divided in three subsets: "low" is defined as the 1st quartile, "medium" is the inter-quartile range, and "high" the 4th quartile. The column "high-low" reports the difference between the high and low subgroup. Returns are expressed in basis points (bps).

move of −37.39 bps compared to +36.41 bps for a positive one.[28] The subsequent 5-minute log-return is +8.01 bps and −5.44 bps, on average. The serial correlation is −0.7958 (−0.2573) for negative (positive) drift bursts. Looking further at price drops, we split the sample into negative drift bursts with a smaller and larger return or volume than that of an average event. The negative serial correlation is most pronounced in drops that are below average (more negative) or accompanied by high trading volume. These effects are not present in positive drift bursts. At last, volatility is markedly higher for tail drift bursts (in either direction) and those with large volume.

Table 6 reports conditional averages of the post-drift burst return $R_{t_j}^+$, which is double-sorted by pre-drift burst return $R_{t_j}^-$ and volume $V_{t_j}^-$. As before, the table is shaped as Table 2 in Huang and Wang (2009) to facilitate comparison. It reinforces that the reversion is stronger if the price change or volume is large. The results are again more pronounced for negative drift bursts.

As an alternative way to describe the double-conditioning of Table 6, we estimate the forecasting equation of Campbell et al. (1993):

$$R_{t_j}^+ = a + b R_{t_j}^- + c R_{t_j}^- V_{t_j}^- + \epsilon_{t_j}, \tag{37}$$

where an interaction term $R_{t_j}^- V_{t_j}^-$ is added to Eq. (36). This approach is suggested by Huang and Wang (2009) in order to assess if the return predictability is driven by trading volume. The parameter estimates of Eq. (37) are displayed in Table 7 along with *t*-statistics (in parenthesis) based on Newey and West (1987)-robust standard errors. Once we control for volume, the serial correlation observed during a negative drift burst is largely subsumed by the interaction term and the coefficient on $R_{t_j}^-$ is heavily reduced, albeit it remains significant for drops below average. This does not occur for positive drift bursts. We again split the sample in half according to the size of the pre-drift burst return and – as predicted by Huang and Wang (2009) – we note that the volume-return variable is more important for large negative drops. On the other hand, for positive drift bursts it is largely irrelevant in explaining the post-drift burst return.

To summarize, Tables 5–7 confirm the theoretical predictions made by Grossman and Miller (1988) and Huang and Wang (2009). The reversion in drift bursts is consistent with market makers absorbing large orders and charging a fee for this service. The asymmetry between negative and positive drift bursts is a symptom of endogenous demand for liquidity due to costly market participation. In the latter case, we confirm several testable implications on trading volume (regarded as a proxy of order imbalance), in that compensation increases with volume, which is larger during negative drift bursts. Overall, our findings suggest the regular occurrence of "flash crashes" documented here is consistent with existing theories of liquidity provision.

6. Conclusion

The drift burst hypothesis is proposed as a theoretical framework for the modeling of distinct and sustained trends in the price paths of financial assets. We show how drift bursts – defined as a short-lived locally explosive drift coefficient – can be embedded into standard continuous-time models and demonstrate that the arbitrage-free property is preserved if the volatility co-explodes during a drift burst, something we provide strong empirical support for. Applying a novel methodology for drift burst identification to a comprehensive set of tick data covering six major asset classes, we deliver unprecedented insights into potentially disruptive but poorly understood events, such as flash crashes. In contrast to the existing literature, which has mostly regarded these events as market glitches, we show that they are instead a regular, stylized fact in the markets, whose dynamic features match theoretical predictions from the market microstructure literature of liquidity provision.

---

[28] These price changes are far out in the tails of the return distribution. A −37.39 bps drop in five minutes translates to a loss of −31.41% in seven hours. In our sample it represents a tenfold increase in magnitude compared to a typical 5-minute negative stock market return and corresponds to a four standard deviations draw, or about 1 in 25,000.

Table 7
ES post-drift burst return forecasting equation.

	Negative drift burst			Positive drift burst		
	a	b	c	a	b	c
All	−11.06	−0.51	–	0.56	−0.16	–
	(−3.50)	(−5.34)		(0.20)	(−1.78)	
	−2.69	−0.17	−0.02	0.45	−0.16	−0.00
	(−1.41)	(−2.17)	(−6.06)	(0.16)	(−1.66)	(−0.46)
$R_{t_j}^- > E[R_{t_j}^-]$	0.54	−0.13	−0.01	−1.34	−0.13	−0.00
	(0.28)	(−1.10)	(−0.78)	(−0.15)	(−0.72)	(−0.51)
$R_{t_j}^- < E[R_{t_j}^-]$	−14.58	−0.34	−0.01	0.75	−0.15	−0.00
	(−3.52)	(−3.96)	(−3.86)	(0.36)	(−1.36)	(−0.37)

Note. This table shows the estimated OLS coefficients from the Campbell et al. (1993) forecasting equation $R_{t_j}^+ = a + b R_{t_j}^- + c R_{t_j}^- V_{t_j}^- + \epsilon_{t_j}$, where $R_{t_j}^+$ is the post-drift burst return, $R_{t_j}^-$ the pre-drift burst return, and $V_{t_j}^-$ is a pre-drift burst normalized gross trading volume. t-statistics of the parameter estimates – based Newey and West (1987)-corrected standard errors – are reported in parenthesis. The regression is fitted separately for negative (left panel) and positive (right panel) detected drift burst with $c = 0$ (first row); with $c \neq 0$ (second row); after conditioning on $R_{t_j}^-$ being above (third row) and below (fourth row) its average value $E[R_{t_j}^-]$ in each of the two subsamples (negative and positive drift bursts).

The paper contributes towards a better understanding of the microstructure dynamics of financial markets and can help to inform the regulatory policy agenda going forward by shedding light on a number of important questions: Who triggers a flash crash? Who supplies liquidity during the event? What is the role played by high-frequency traders? And what, if anything, can be done to prevent flash crashes in the future?

Appendix A. Mathematical appendix

In this section, C is a generic positive constant, whose value may change from line to line.

Assumption 3. With the notation of Assumption 1, we further impose that
(i) The jump process J_t is of the form:

$$J_t = \int_0^t \int_{\mathbb{R}} \delta(s, x) I_{\{|\delta(s,x)| \leq 1\}} (\nu(ds, dx) - \tilde{\nu}(ds, dx)) + \int_0^t \int_{\mathbb{R}} \delta(s, x) I_{\{|\delta(s,x)| > 1\}} \nu(ds, dx), \tag{38}$$

where ν is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}$, $\tilde{\nu}(ds, dx) = \lambda(dx)ds$ a compensator, and λ is a σ-finite measure on $\mathbb{R}$. Moreover, $\delta : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ is a predictable function such that there exists a sequence $(\tau_n)_{n \geq 1}$ of $\mathscr{F}_t$-stopping times with $\tau_n \to \infty$ and, for each n, a deterministic and non-negative function Γ_n with $\min(|\delta(t,x)|, 1) \leq \Gamma_n(x)$ for $t \leq \tau_n$ and $\int_{\mathbb{R}} \Gamma_n(x)^2 \lambda(dx) < \infty$ and, for all $\kappa \in (0, 1)$, $\int_{\{x: \Gamma_n(x) \leq \kappa\}} \Gamma_n(x)\lambda(dx) < \infty$, for all (t, x) and $n \geq 1$.
(ii) Fix $t \in (0, T]$. We assume there exists a $\Gamma > 0$ and a sequence $(\tau_m)_{m \geq 1}$ of $\mathscr{F}_t$-stopping times with $\tau_m \to \infty$ and constants $C_t^{(m)}$ such that for all m,

$$E_s \left[|\mu_u - \mu_s|^2 + |\sigma_u - \sigma_s|^2 \right] \leq C_t^{(m)} |u - s|^\Gamma, \tag{39}$$

for all $0 \leq s \leq u \leq T \wedge \tau_m$, where $E_t[\cdot] = E[\cdot|\mathscr{F}_t]$.

Remark 2. With Assumptions 1 and 3 in hand, the localization procedure in Jacod and Protter (2012, Section 4.4.1) implies that we can assume μ_t, σ_t, and $\delta(t, x)$ are bounded, as (ω, t, x) vary within $\Omega \times [0, T] \times \mathbb{R}$, and that $|\delta(t, x)| \leq \overline{\Gamma}(x)$, where $\overline{\Gamma}(x)$ is bounded and such that $\int_{\mathbb{R}} \overline{\Gamma}(x)^2 \lambda(dx) < \infty$ and, for all $\kappa \in (0, 1)$, $\int_{\{x: \overline{\Gamma}(x) \leq \kappa\}} \overline{\Gamma}(x)\lambda(dx) < \infty$.

Assumption 4. $(t_i)_{i=0}^n$ is a deterministic sequence. We denote by $\Delta_{i,n} = t_i - t_{i-1}$, $\Delta_n^- = \min_{i=1,\ldots,n}\{\Delta_{i,n}\}$, $\Delta_n^+ = \max_{i=1,\ldots,n}\{\Delta_{i,n}\}$ and assume that, for a sufficiently large n, and suitable constants c and C (that do not depend on n),

$$c\Delta_n \leq \Delta_n^- \leq \Delta_n^+ \leq C\Delta_n,$$

where $\Delta_n = T/n$. Moreover, denoting the "quadratic variation of time up to t" as $H(t) = \lim_{n\to\infty} H_n(t)$, where $H_n(t) = \frac{1}{\Delta_n} \sum_{t_i \leq t} (\Delta_{i,n})^2$, we assume $H(t)$ exists and is Lebesgue-almost surely differentiable in $(0, T)$ with derivative H' such that $|H'(t_i) - \Delta_{i,n}/\Delta_n| \leq C\Delta_{i,n}$, for any t_i in which H is differentiable, where C does not depend on i and n.

Assumption 5. The bandwidths h_n, h_n' are sequences of positive real numbers, such that, as $n \to \infty$, $h_n \to 0$, $h_n' \to 0$, $nh_n \to \infty$, and $nh_n' \to \infty$. The kernel $K : \mathbb{R} \to \mathbb{R}_+$ is any function with the properties:

(K0) $K(x) = 0$ for $x > 0$;

(K1) K is bounded and differentiable with bounded first derivative. Further, $xK(x) \to 0$ and $xK'(x) \to 0$ as $x \to -\infty$;

(K2) $\int_{-\infty}^{0} K(x)dx = 1$ and $K_2 = \int_{-\infty}^{0} K^2(x)dx < \infty$;

(K3) It holds that for every positive sequence $G_{n,t} \to \infty$, $\int_{-\infty}^{-G_{n,t}} K(x)dx \le CG_{n,t}^{-B}$ for some $B > 0$ and $C > 0$ (i.e., K has a fast vanishing tail);

(K4) $m_K(\alpha) = \int_{-\infty}^{0} K(x)|x|^{\alpha}dx < \infty$, for all $\alpha > -1$; $m_K'(\alpha) = \int_{-\infty}^{0} K^2(x)|x|^{\alpha}dx < \infty$, for all $\alpha > -1$.

Remark 3. Condition (K0) can be extended to allow for two-sided kernels without changing any of the theoretical results and only minor modifications in the proofs. We impose it to make the mathematical exposition less cumbersome and aligned with the implementation in the empirical application.

Without loss of generality, in the proofs we set $h_n' = h_n$.

Lemma 1. *Assume that the conditions of Assumptions 1 and 4–5 hold. Then, for every fixed $t \in (0, T]$ as $n \to \infty$ and $h_n \to 0$, it holds that:*

$$A_n = \frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) \int_{t_{i-1}}^{t_i} \mu_s ds - \int_0^T \frac{1}{h_n} K\left(\frac{s - t}{h_n}\right) \mu_s ds = O_p\left(\frac{1}{nh_n}\right),$$

$$B_n = \frac{1}{\Delta_n h_n} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right)\left(\int_{t_{i-1}}^{t_i} \mu_s ds\right)^2 - \int_0^T \frac{1}{h_n} K\left(\frac{s - t}{h_n}\right) \mu_s^2 ds = O_p\left(\frac{1}{nh_n}\right).$$

This also applies if μ_t is replaced by σ_t.

Proof. Write:

$$A_n = \frac{1}{h_n} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left(K\left(\frac{t_{i-1} - t}{h_n}\right) - K\left(\frac{s - t}{h_n}\right)\right) \mu_s ds.$$

The mean value theorem – together with the boundedness of K' and μ_t – implies that, for each interval $[t_{i-1}, t_i]$, there exists a ξ_{t_{i-1}, t_i} such that:

$$|A_n| \le \frac{1}{h_n} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left|K'\left(\frac{\xi_{t_{i-1}, t_i} - t}{h_n}\right)(s - t_{i-1})\right| |\mu_s| ds \le C \frac{T}{n} \frac{1}{h_n} \int_0^T |\mu_s| ds \le C \frac{1}{nh_n},$$

where the second inequality follows from (K1). The proof for B_n is nearly identical. ∎

Lemma 2. *Assume that the conditions of Assumptions 1 and 4–5 hold. Then, for every fixed $t \in (0, T]$ as $n \to \infty$ and $h_n \to 0$, it holds that:*

$$A_n' = \int_0^T \frac{1}{h_n} K\left(\frac{s - t}{h_n}\right) \mu_s ds - \mu_{t-} = O_p\left(h_n^{\Gamma/2} + h_n^B\right).$$

This also applies if μ_t is replaced by σ_t.

Proof. Note that, by the properties of the kernel,

$$\mu_{t-} = \mu_{t-} \int_{-\infty}^{0} K(x)dx = \mu_{t-}\left(\int_{-\infty}^{-t/h_n} K(x)dx + \int_{-t/h_n}^{0} K(x)dx\right),$$

so we can write:

$$|A_n'| = \left|\int_0^T \frac{1}{h_n} K\left(\frac{s - t}{h_n}\right)(\mu_s - \mu_{t-})ds - \mu_{t-} \int_{-\infty}^{-t/h_n} K(x)dx\right| \le \int_0^T \frac{1}{h_n} K\left(\frac{s - t}{h_n}\right) |\mu_s - \mu_{t-}| ds + Ch_n^B,$$

where B is the constant in (K3). Then, by Jensen's inequality and Eq. (39):

$$E_{s \wedge t}\left[|\mu_s - \mu_{t-}|\right] \le C|s - t|^{\Gamma/2}.$$

Together with (K4) and a change of variable, this implies that:

$$E\left[\int_0^T \frac{1}{h_n} K\left(\frac{s - t}{h_n}\right) |\mu_s - \mu_{t-}| ds\right] \le \int_0^T \frac{1}{h_n} K\left(\frac{s - t}{h_n}\right) |s - t|^{\Gamma/2} ds = \int_{-t/h_n}^{0} K(x)|x|^{\Gamma/2} h_n^{\Gamma/2} dx \le Ch_n^{\Gamma/2}.$$

This concludes the proof. ∎

Lemma 3. *Assume that the conditions of Assumptions 1 and 3–5 hold. Then, for every fixed $t \in (0, T]$, as $n \to \infty$ and $h_n \to 0$ such that $nh_n \to \infty$, it holds that $\hat{\sigma}_t^n \xrightarrow{p} \sigma_{t-}$.*

Proof. The lemma extends Jacod and Protter (2012, Theorem 9.3.2) to a general kernel (as defined in Assumption 5). We compensate the large jump term and write $X_t' = \int_0^t \mu_s^* ds + \int_0^t \sigma_s dW_s$, where $\mu_t^* = \mu_t + \int_{\mathbb{R}} \delta(t, x) I_{\{|\delta(t,x)|>1\}} \lambda(dx)$ is bounded, and $X_t'' = X_t - X_t' = \int_{\mathbb{R}} \delta(s, x) (\nu(ds, dx) - \tilde{\nu}(ds, dx))$. Now, Mancini et al. (2015) established the convergence in probability:

$$\frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) \left(\Delta_i^n X'\right)^2 \xrightarrow{p} \sigma_{t-}^2. \tag{40}$$

Thus, it is enough to show that:

$$R_n^{\hat{\sigma}} = \frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) \left(\left(\Delta_i^n X\right)^2 - \left(\Delta_i^n X'\right)^2\right) \xrightarrow{p} 0.$$

Note that for $\kappa \in (0, 1)$, we may write $\Delta_i^n X = \Delta_i^n X' + \Delta_i^n X_1''(\kappa) + \Delta_i^n X_2''(\kappa)$, where

$$\Delta_i^n X_1''(\kappa) = \int_{\mathbb{R}} \delta(s, x) I_{\{\overline{\Gamma}(x) \leq \kappa\}} (\nu(ds, dx) - \tilde{\nu}(ds, dx)) \quad \text{and} \quad \Delta_i^n X_2''(\kappa) = \int_{\mathbb{R}} \delta(s, x) I_{\{\overline{\Gamma}(x) > \kappa\}} (\nu(ds, dx) - \tilde{\nu}(ds, dx)).$$

Applying the decomposition in Jacod and Protter (2012, Equation 9.3.9):

$$|R_n^{\hat{\sigma}}| \leq \frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) \left| \left(\Delta_i^n X\right)^2 - \left(\Delta_i^n X'\right)^2 \right|$$

$$\leq \frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) \left(\epsilon \left(\Delta_i^n X'\right)^2 + \frac{C}{\epsilon} \left(\left(\Delta_i^n X_1''(\kappa)\right)^2 + \left(\Delta_i^n X_2''(\kappa)\right)^2\right)\right),$$

for $0 < \epsilon \leq 1$ and C independent of ϵ.

Next, define $\Omega_n(\psi, \kappa) \subseteq \Omega$ the set such that the Poisson process $\nu([0, t] \times \{x : \overline{\Gamma}(x) > \kappa\})$ has no jumps in the interval $(t - h_n^\psi, t]$, for $0 < \psi < 1$ and $0 < \kappa < 1$. Note that $\Omega_n(\psi, \kappa) \to \Omega$, as $n \to \infty$, since $h_n^\psi \to 0$. Now write, for all $c > 0$, $\mathscr{P}\left(|R_n^{\hat{\sigma}}| > c\right) = \mathscr{P}\left(|R_n^{\hat{\sigma}}| > c \mid \Omega_n^{\complement}(\psi, \kappa)\right) + \mathscr{P}\left(|R_n^{\hat{\sigma}}| > c \mid \Omega_n(\psi, \kappa)\right) \leq \mathscr{P}\left(\Omega_n^{\complement}(\psi, \kappa)\right) + E\left[|R_n^{\hat{\sigma}}| \mid \Omega_n(\psi, \kappa)\right]/c$ by Markov's inequality.

On $\Omega_n(\psi, \kappa)$, we can bound the "large" jumps before $t - h_n^\psi$ as follows

$$E\left[\frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) \frac{C}{\epsilon} \left(\Delta_i^n X_2''(\kappa)\right)^2 \mid \Omega_n(\psi, \kappa)\right] = \frac{1}{h_n} \sum_{t_{i-1} \leq t - h_n^\psi} K\left(\frac{t_{i-1} - t}{h_n}\right) \frac{C}{\epsilon} E\left[\left(\Delta_i^n X_2''(\kappa)\right)^2\right]$$

$$\leq \frac{1}{h_n} \sum_{t_{i-1} \leq t - h_n^\psi} K\left(\frac{t_{i-1} - t}{h_n}\right) \frac{C}{\epsilon} \Delta_{i,n} \leq \frac{C}{\epsilon} h_n^{B(1-\psi)}$$

by a Riemann approximation and (K3). Moreover,

$$E\left[\left(\Delta_i^n X_1''(\kappa)\right)^2\right] \leq C \Delta_{i,n} \int_{\{x : \overline{\Gamma}(x) \leq \kappa\}} \overline{\Gamma}(x)^2 \lambda(dx),$$

so that, using again the convergence in Eq. (40);

$$E\left[|R_n^{\hat{\sigma}}| \mid \Omega_n(\psi, \kappa)\right] \leq C\epsilon + \frac{C}{\epsilon} \left(\int_{\{x : \overline{\Gamma}(x) \leq \kappa\}} \overline{\Gamma}(x)^2 \lambda(dx) + h_n^{B(1-\psi)}\right).$$

It follows that for all ϵ and $\kappa \in (0, 1)$,

$$\limsup_{n \to \infty} \mathscr{P}\left(|R_n^{\hat{\sigma}}| > c\right) \leq \frac{1}{c} \left(C\epsilon + \frac{C}{\epsilon} \left(\int_{\{x : \overline{\Gamma}(x) \leq \kappa\}} \overline{\Gamma}(x)^2 \lambda(dx) + h_n^{B(1-\psi)}\right)\right).$$

Setting $\epsilon = \left(\int_{\{x : \overline{\Gamma}(x) \leq \kappa\}} \overline{\Gamma}(x)^2 \lambda(dx) + h_n^{B(1-\psi)}\right)^{\psi'}$ with $0 < \psi' < 1$ and noticing that $\left(\int_{\{x : \overline{\Gamma}(x) \leq \kappa\}} \overline{\Gamma}(x)^2 \lambda(dx) + h_n^{B(1-\psi)}\right) \to 0$ as $\kappa \to 0$ and $n \to \infty$, we deduce that $\mathscr{P}\left(|R_n^{\hat{\sigma}}| > c\right) \to 0$. The convergence in probability $\hat{\sigma}_t^n \xrightarrow{p} \sigma_{t-}$ then follows from these results in combination with Slutsky's Theorem. ∎

Proof of Theorem 1. We decompose T_t^n into:

$$T_t^n = \underbrace{\sqrt{\frac{h_n}{K_2}} \frac{\left(\hat{\mu}_t^n - \mu_{t-}^*\right)}{\hat{\sigma}_t^n}}_{T_1} + \underbrace{\sqrt{\frac{h_n}{K_2}} \frac{\mu_{t-}^*}{\hat{\sigma}_t^n}}_{T_2},$$

where μ_t^* is the jump compensated drift as defined in the proof of Lemma 3. This, together with the boundedness of μ_t^*, σ_t and $\delta(t, x)$, yields the following:

$$|T_2| \leq C \frac{\sqrt{h_n}}{\hat{\sigma}_t^n} = O_p\left(\sqrt{h_n}\right).$$

Now, from Lemmas 1–2 we can write:

$$\hat{\mu}_t^n - \mu_{t-}^* = \frac{1}{h_n} \sum_{i=1}^n K\left(\frac{t_{i-1} - t}{h_n}\right) \Delta_i^n X - \frac{1}{h_n} \sum_{i=1}^n K\left(\frac{t_{i-1} - t}{h_n}\right) \int_{t_{i-1}}^{t_i} \mu_s^* ds + O_p\left(\frac{1}{nh_n} + h_n^{\Gamma/2} + h_n^B\right).$$

Hence,

$$\sqrt{h_n}\left(\hat{\mu}_t^n - \mu_{t-}^*\right) = \underbrace{\frac{1}{\sqrt{h_n}} \sum_{i=1}^n K\left(\frac{t_{i-1} - t}{h_n}\right) \int_{t_{i-1}}^{t_i} \sigma_s dW_s}_{G_{n,t}}$$

$$+ \underbrace{\frac{1}{\sqrt{h_n}} \sum_{i=1}^n K\left(\frac{t_{i-1} - t}{h_n}\right) \int_{t_{i-1}}^{t_i} \int_{\mathbb{R}} \delta(s, x)\left(\nu(ds, dx) - \tilde{\nu}(ds, dx)\right)}_{G'_{n,t}} + O_p\left(\frac{\sqrt{h_n}}{nh_n} + h_n^{\Gamma/2+1/2} + h_n^{B+1/2}\right).$$

The $O_p(\cdot)$ term is asymptotically negligible, as $n \to \infty$, $h_n \to 0$ and $nh_n \to \infty$. $G'_{n,t}$ also vanishes, which we show as in the proof of Lemma 3 by writing: $\int_{t_{i-1}}^{t_i} \int_{\mathbb{R}} \delta(s, x)\left(\nu(ds, dx) - \tilde{\nu}(ds, dx)\right) = \Delta_i^n X_1''(\kappa) + \Delta_i^n X_2''(\kappa)$ with $\kappa \in (0, 1)$ and writing $G'_{n,t} = G'_{n,t,1} + G'_{n,t,2}$, where $G'_{n,t,j} = \frac{1}{\sqrt{h_n}} \sum_{i=1}^n K\left(\frac{t_{i-1} - t}{h_n}\right) \Delta_i^n X_j''(\kappa)$ for $j = 1, 2$. Now, on the set $\Omega_n(\psi, \kappa)$:

$$E\left[\frac{1}{\sqrt{h_n}} \sum_{i=1}^n K\left(\frac{t_{i-1} - t}{h_n}\right) |\Delta_i^n X_2''(\kappa)| \mid \Omega_n(\psi, \kappa)\right]$$

$$= \frac{1}{\sqrt{h_n}} \sum_{t_{i-1} \leq t - h_n^\psi} K\left(\frac{t_{i-1} - t}{h_n}\right) E\left[|\Delta_i^n X_2''(\kappa)|\right]$$

$$\leq \frac{1}{\sqrt{h_n}} \sum_{t_{i-1} \leq t - h_n^\psi} K\left(\frac{t_{i-1} - t}{h_n}\right) C\Delta_{i,n} \int_{\{x:\overline{\Gamma}(x)>\kappa\}} \overline{\Gamma}(x)\lambda(dx) \leq C\sqrt{h_n} h_n^{B(1-\psi)},$$

where the first inequality is by virtue of Jacod and Protter (2012, Lemma 2.1.5) applied with $p = 1$. Since $\mathscr{P}\left(|G'_{n,t,2}| > c\right) \leq \mathscr{P}\left(\Omega_n^{\complement}(\kappa, \psi)\right) + E\left[|G'_{n,t,2}| \mid \Omega_n(\kappa, \psi)\right]/c$, it again follows that

$$\limsup_{n \to \infty} \mathscr{P}\left(|G'_{n,2}| > c\right) \leq \frac{1}{c} C\sqrt{h_n} h_n^{B(1-\psi)},$$

so that we deduce $G'_{n,t,2} \xrightarrow{p} 0$ as $n \to \infty$.

Next, from Lemma 2.1.5 in Jacod and Protter (2012),

$$E\left[|G'_{n,t,1}|\right] \leq \frac{1}{\sqrt{h_n}} \sum_{i=1}^n K\left(\frac{t_{i-1} - t}{h_n}\right) E\left[|\Delta_i^n X_1''(\kappa)|\right] \leq \frac{C}{\sqrt{h_n}} \sum_{i=1}^n K\left(\frac{t_{i-1} - 1}{h_n}\right) \Delta_{i,n} \int_{\{x:\overline{\Gamma}(x)\leq\kappa\}} \overline{\Gamma}(x)\lambda(dx),$$

and the bound has order $O_p(\sqrt{h_n})$ so that $G'_{n,t,1} \xrightarrow{p} 0$ and, therefore, $G'_{n,t} = o_p(1)$.

$G_{n,t}$ is the leading term, which we write as $G_{n,t} = \sum_{i=1}^n \Delta_i^n u$ with $\Delta_i^n u = \frac{1}{\sqrt{h_n}} K\left(\frac{t_{i-1} - t}{h_n}\right) \int_{t_{i-1}}^{t_i} \sigma_s dW_s$. The aim is to prove that $G_{n,t}$ – and hence $\sqrt{h_n}\left(\hat{\mu}_t^n - \mu_{t-}^*\right)$ – converges stably in law to $N\left(0, K_2\sigma_{t-}^2\right)$. We exploit Theorem 2.2.14 in Jacod and Protter (2012), which lists four sufficient conditions for this to hold:

$$\sum_{i=1}^n E_{t_{i-1}}\left[\Delta_i^n u\right] \xrightarrow{p} 0, \tag{41}$$

$$\sum_{i=1}^{n} E_{t_{i-1}}\left[(\Delta_i^n u)^2\right] \xrightarrow{p} K_2 \sigma_{t_-}^2, \tag{42}$$

$$\sum_{i=1}^{n} E_{t_{i-1}}\left[(\Delta_i^n u)^4\right] \xrightarrow{p} 0, \tag{43}$$

$$\sum_{i=1}^{n} E_{t_{i-1}}\left[\Delta_i^n u \Delta_i^n Z\right] \xrightarrow{p} 0, \tag{44}$$

where either $Z_t = W_t$ or $Z_t = W_t'$ with W_t' being orthogonal to W_t. The condition in Eq. (41) is immediate. Next, from Itô's Lemma, we deduce that:

$$\left(\int_{t_{i-1}}^{t_i} \sigma_s dW_s\right)^2 = \int_{t_{i-1}}^{t_i} \sigma_s^2 ds + 2\int_{t_{i-1}}^{t_i} \sigma_s \left(\int_{t_{i-1}}^{s} \sigma_u dW_u\right) dW_s,$$

so that

$$\begin{aligned}
\sum_{i=1}^{n} E_{t_{i-1}}\left[(\Delta_i^n u)^2\right] &= \sum_{i=1}^{n} \frac{1}{h_n} K^2\left(\frac{t_{i-1}-t}{h_n}\right) E_{t_{i-1}}\left[\left(\int_{t_{i-1}}^{t_i} \sigma_s dW_s\right)^2\right] \\
&= \sum_{i=1}^{n} \frac{1}{h_n} K^2\left(\frac{t_{i-1}-t}{h_n}\right) E_{t_{i-1}}\left[\int_{t_{i-1}}^{t_i} \sigma_s^2 ds\right] \\
&= \sum_{i=1}^{n} \frac{1}{h_n} K^2\left(\frac{t_{i-1}-t}{h_n}\right)\left(\sigma_{t_{i-1}}^2 \Delta_{i,n} + E_{t_{i-1}}\left[\int_{t_{i-1}}^{t_i}\left(\sigma_s^2 - \sigma_{t_{i-1}}^2\right) ds\right]\right).
\end{aligned}$$

The first term converges to $K_2 \sigma_{t_-}^2$, as shown in Mancini et al. (2015). The second term is negligible by the Lipschitz condition in Eq. (39):

$$\sum_{i=1}^{n} \frac{1}{h_n} K^2\left(\frac{t_{i-1}-t}{h_n}\right) E_{t_{i-1}}\left[\int_{t_{i-1}}^{t_i}\left(\sigma_s^2 - \sigma_{t_{i-1}}^2\right) ds\right] \le \sum_{i=1}^{n} \frac{1}{h_n} K^2\left(\frac{t_{i-1}-t}{h_n}\right) \Delta_{i,n} \Delta_{i,n}^\Gamma = O_p\left(\Delta_n^\Gamma\right).$$

To deal with the third condition in Eq. (43), we note that by the Burkholder–Davis–Gundy inequality and from the boundedness of σ_t, it holds that:

$$E_{t_{i-1}}\left[\left(\int_{t_{i-1}}^{t_i} \sigma_s dW_s\right)^4\right] \le C(\Delta_{i,n})^2,$$

which leads to

$$\begin{aligned}
\sum_{i=1}^{n} E_{t_{i-1}}\left[(\Delta_i^n u)^4\right] &= \sum_{i=1}^{n} \frac{1}{h_n^2} K^4\left(\frac{t_{i-1}-t}{h_n}\right) E_{t_{i-1}}\left[\left(\int_{t_{i-1}}^{t_i} \sigma_s dW_s\right)^4\right] \\
&\le C\sum_{i=1}^{n} \frac{1}{h_n^2} K^4\left(\frac{t_{i-1}-t}{h_n}\right)(\Delta_{i,n})^2 = O\left(\frac{\Delta_n}{h_n}\right).
\end{aligned}$$

To deal with Eq. (44), we first set $Z_t = W_t$. Then, using the Cauchy–Schwarz inequality:

$$\begin{aligned}
E_{t_{i-1}}\left[\Delta_i^n W \int_{t_{i-1}}^{t_i} \sigma_s dW_s\right] &\le \sqrt{E_{t_{i-1}}\left[(\Delta_i^n W)^2\right]} \sqrt{E_{t_{i-1}}\left[\left(\int_{t_{i-1}}^{t_i} \sigma_s dW_s\right)^2\right]} \\
&= \sqrt{\Delta_{i,n}} \sqrt{E_{t_{i-1}}\left[\int_{t_{i-1}}^{t_i} \sigma_s^2 ds\right]} = O_p\left(\Delta_n\right),
\end{aligned}$$

and therefore

$$\sum_{i=1}^{n} E_{t_{i-1}}\left[\Delta_i^n u \Delta_i^n W\right] \le C\frac{1}{\sqrt{h_n}} \sum_{i=1}^{n} K\left(\frac{t_{i-1}-t}{h_n}\right) \Delta_{i,n} \to 0.$$

If $Z_t = W'_t$, the process $W'_t \int_0^t \sigma_s dW_s$ is a martingale by orthogonality, so that:

$$E_{t_{i-1}} \left[\Delta_i^n W' \int_{t_{i-1}}^{t_i} \sigma_s dW_s \right] = 0.$$

This verifies that $\sqrt{h_n} \left(\hat{\mu}_t^n - \mu_{t-}^* \right) \xrightarrow{d} N \left(0, K_2 \sigma_{t-}^2 \right)$, where the convergence is stable in law. Combined with Lemma 3, this yields $T_t^n \xrightarrow{d} N(0, 1)$. ∎

Proof of Theorem 2. Without loss of generality, we set $\tau_{db} = 1$. We write $\widetilde{X}_t = X_t + D_t + V_t$, where $D_t = \int_0^t c_{1,s}(1-s)^{-\alpha} ds$ and $V_t = \int_0^t c_{2,s}(1-s)^{-\beta} dW_s$.

Look at the term $\hat{\mu}_t^n$. In Theorem 1, we already showed that $\frac{1}{h_n} \sum_{i=1}^n K \left(\frac{t_{i-1}-1}{h_n} \right) \Delta_i^n X = O_p \left(\frac{1}{\sqrt{h_n}} \right)$. Now,

$$A_n = \frac{1}{h_n} \sum_{i=1}^n K \left(\frac{t_{i-1}-1}{h_n} \right) \int_{t_{i-1}}^{t_i} c_{1,s}(1-s)^{-\alpha} ds = \frac{1}{h_n} \sum_{i=1}^n K \left(\frac{t_{i-1}-1}{h_n} \right) \Delta_{i,n} c_{1,\xi_{t_{i-1},t_i}} (1-\xi_{t_{i-1},t_i})^{-\alpha},$$

where $t_{i-1} \leq \xi_{t_{i-1},t_i} \leq t_i$. The last term is, following a change of variable and Riemann summation, asymptotically equivalent to $c_{1,1} h_n^{-\alpha} m_K(-\alpha)$, where $m_K(-\alpha)$ is the constant in (K4).

As for the second term:

$$B_n = \frac{1}{h_n} \sum_{i=1}^n K \left(\frac{t_{i-1}-1}{h_n} \right) \int_{t_{i-1}}^{t_i} c_{2,s}(1-s)^{-\beta} dW_s,$$

we apply again Jacod and Protter (2012, Theorem 2.2.14). The crucial part is the variance (we omit the other three conditions that are uncomplicated here), and for this term we find:

$$\frac{1}{h_n^2} \sum_{i=1}^n K^2 \left(\frac{t_{i-1}-1}{h_n} \right) E \left[\left(\int_{t_{i-1}}^{t_i} c_{2,s}(1-s)^{-\beta} dW_s \right)^2 \right] = \frac{1}{h_n^2} \sum_{i=1}^n K^2 \left(\frac{t_{i-1}-1}{h_n} \right) E \left[\int_{t_{i-1}}^{t_i} c_{2,s}^2 (1-s)^{-2\beta} ds \right]$$
$$\overset{p}{\sim} h_n^{-1-2\beta} m_K'(-2\beta) c_{2,1}^2,$$

which implies that

$$h_n^{1/2+\beta} B_n \xrightarrow{d} N \left(0, m_K'(-2\beta) c_{2,1}^2 \right),$$

where the convergence is stable in law. Thus, $\hat{\mu}_t^n$ is dominated by A_n when $\alpha - \beta > 1/2$, by B_n when $\alpha - \beta < 1/2$, whereas both terms influence the limit with $\alpha - \beta = 1/2$.

We turn to the denominator and set $(\hat{\sigma}_t^n)^2 = \frac{1}{h_n} \sum_{i=1}^n K \left(\frac{t_{i-1}-1}{h_n} \right) \left[(\Delta_i^n D)^2 + (\Delta_i^n V)^2 \right] + R_n'$, where

$$R_n' = \frac{1}{h_n} \sum_{i=1}^n K \left(\frac{t_{i-1}-1}{h_n} \right) \left((\Delta_i^n X + \Delta_i^n D + \Delta_i^n V)^2 - (\Delta_i^n D)^2 - (\Delta_i^n V)^2 \right).$$

Since for all $\epsilon > 0$ and a, b and c real: $(a+b+c)^2 - a^2 - b^2 \leq \epsilon(a^2 + b^2) + \frac{1+\epsilon}{\epsilon} c^2$, we deduce the bound:

$$|R_n'| \leq \frac{1}{h_n} \sum_{i=1}^n K \left(\frac{t_{i-1}-1}{h_n} \right) \left(\epsilon \left((\Delta_i^n D)^2 + (\Delta_i^n V)^2 \right) + \frac{1+\epsilon}{\epsilon} \left(\Delta_i^n X \right)^2 \right)$$
$$= \epsilon \frac{1}{h_n} \sum_{i=1}^n K \left(\frac{t_{i-1}-1}{h_n} \right) \left((\Delta_i^n D)^2 + (\Delta_i^n V)^2 \right) + \frac{1+\epsilon}{\epsilon} O_p(1).$$

Setting $\epsilon = h_n^c$, for a suitable $c > 0$, R_n' is at most $O_p\left(h_n^{-c}\right)$ and therefore negligible compared to the two leading terms. Next, setting $\Theta_n = \frac{1}{h_n}\sum_{i=1}^n K\left(\frac{t_{i-1}-1}{h_n}\right)(\Delta_i^n V)^2$ and using Itô's Lemma:

$$(\Delta_i^n V)^2 = \left(\int_{t_{i-1}}^{t_i} c_{2,s}(1-s)^{-\beta}dW_s\right)^2 = 2\int_{t_{i-1}}^{t_i}\left(\int_{t_{i-1}}^s c_{2,u}(1-u)^{-\beta}dW_u\right)c_{2,s}(1-s)^{-\beta}dW_s + \int_{t_{i-1}}^{t_i} c_{2,s}^2(1-s)^{-2\beta}ds,$$

and we split $\Theta_n = \Theta_{1,n} + \Theta_{2,n}$ accordingly. As for the term A_n, $\Theta_{2,n}$ is asymptotically equivalent to $c_{2,1}^2 h_n^{-2\beta}m_K(-2\beta)$. We write $\Theta_{1,n} = \sum_{i=1}^n \Delta u_i'$, where

$$\Delta u_i' = \frac{1}{h_n}K\left(\frac{t_{i-1}-1}{h_n}\right)2\int_{t_{i-1}}^{t_i} c_{2,s}(1-s)^{-\beta}\left(\int_{t_{i-1}}^s c_{2,u}(1-u)^{-\beta}dW_u\right)dW_s,$$

so $\Theta_{1,n}$ is a sum of martingale differences. Now, using a Taylor expansion on $c_{2,s}$,

$$\sum_{i=1}^n E_{t_{i-1}}\left[(\Delta u_i')^2\right] = \frac{1}{h_n^2}\sum_{i=1}^n K^2\left(\frac{t_{i-1}-1}{h_n}\right)(c_{2,t_{i-1}}^4 + o_p(1))E_{t_{i-1}}\left[\left(2\int_{t_{i-1}}^{t_i}(1-s)^{-\beta}\left(\int_{t_{i-1}}^s(1-u)^{-\beta}dW_u\right)dW_s\right)^2\right]$$

$$= \frac{4}{h_n^2}\sum_{i=1}^n K^2\left(\frac{t_{i-1}-1}{h_n}\right)(c_{2,t_{i-1}}^4 + o_p(1))\int_{t_{i-1}}^{t_i}(1-s)^{-2\beta}E_{t_{i-1}}\left[\left(\int_{t_{i-1}}^s(1-u)^{-\beta}dW_u\right)^2\right]ds$$

$$= \frac{4}{h_n^2}\sum_{i=1}^n K^2\left(\frac{t_{i-1}-1}{h_n}\right)(c_{2,t_{i-1}}^4 + o_p(1))\int_{t_{i-1}}^{t_i}(1-s)^{-2\beta}\left(\int_{t_{i-1}}^s(1-u)^{-2\beta}du\right)ds$$

$$= \frac{2}{h_n^2}\sum_{i=1}^n K^2\left(\frac{t_{i-1}-1}{h_n}\right)(c_{2,t_{i-1}}^4 + o_p(1))\left((1-t_{i-1})^{-4\beta}\Delta_{i,n}^2 + O(\Delta_{i,n}^3)\right),$$

where the $o_p(1)$ term is uniform in i. If $\beta < 1/4$, Lemma A.1 (ii) in Mancini et al. (2015) implies:

$$\frac{h_n^{1+4\beta}}{\Delta_n}\sum_{i=1}^n E_{t_{i-1}}\left[(\Delta u_i')^2\right] \xrightarrow{p} 2K_2 m_K'(-4\beta)H'(1)c_{2,1}^4,$$

which means that $\Theta_{1,n}$ is dominated by $\Theta_{2,n}$. On the other hand, if $\beta \geq 1/4$, we deduce from Assumption 4 and the boundedness of K and $c_{2,t}$ that

$$\frac{2}{h_n^2}\sum_{i=1}^n K^2\left(\frac{t_{i-1}-1}{h_n}\right)c_{2,t_{i-1}}^4(1-t_{i-1})^{-4\beta}\Delta_{i,n}^2 \leq C\frac{1}{h_n^2}\sum_{i=1}^n(\Delta_n^-(n-(i-1)))^{-4\beta}(\Delta_n^+)^2$$

$$\leq C\frac{\Delta_n^{2-4\beta}}{h_n^2}\sum_{j=1}^n j^{-4\beta} \leq C\frac{\Delta_n^{2-4\beta}}{h_n^2}\log(n),$$

so this term is also dominated by $\Theta_{2,n}$.

With the remaining term in $(\hat{\sigma}_t^n)^2$, it holds that:

$$A_n' = \frac{1}{h_n}\sum_{i=1}^n K\left(\frac{t_{i-1}-1}{h_n}\right)(\Delta_i^n D)^2 = \frac{1}{h_n}\sum_{i=1}^n K\left(\frac{t_{i-1}-1}{h_n}\right)(c_{1,t_{i-1}}^2 + o_p(1))\left(\int_{t_{i-1}}^{t_i}(1-s)^{-\alpha}ds\right)^2$$

$$\overset{p}{\sim} \frac{1}{h_n}\sum_{i=1}^n K\left(\frac{t_{i-1}-1}{h_n}\right)c_{1,t_{i-1}}^2(1-t_{i-1})^{-2\alpha}\Delta_{i,n}^2.$$

As for the term A_n above, when $\alpha < 1/2 : \frac{h_n^{2\alpha}}{\Delta_n}A_n' \xrightarrow{p} c_{1,1}^2 H'(1)m_K(-2\alpha)$. When $\alpha > 1/2$, we deduce (via the logic used to handle the term $\Theta_{1,n}$ above) that $A_n' \leq C\frac{\Delta_n^{2-2\alpha}}{h_n}$, which competes with $\Theta_{2,n}$. This bound is $A_n' \leq C\frac{\Delta_n\log(n)}{h_n}$ (and is then dominated by $\Theta_{2,n}$) for $\alpha = 1/2$.

We combine the various settings as follows. If $\alpha - \beta > 1/2$ (implying $\alpha > 1/2$), $\hat{\mu}_t^n$ is dominated by A_n and has order $h_n^{-\alpha}$, whereas $(\hat{\sigma}_t^n)^2$ is at most of order $\Delta_n^{2-2\alpha}/h_n$ (if A_n' dominates), or $h_n^{-2\beta}$ (if $\Theta_{2,n}$ dominates). In the first instance, the t-statistic is of order $(h_n/\Delta_n)^{1-\alpha} \to \infty$, while in the second it is $h_n^{1/2-\alpha+\beta} \to \infty$. In the arbitrage-free setting $\alpha - \beta < 1/2$, the leading term in $\hat{\mu}_t^n$ is B_n. Note that $(\hat{\sigma}_t^n)^2$ is then dominated by $\Theta_{2,n}$, since the divergence rate of $A_n' \to \infty$ is always slower than $h_n^{-2\beta}$, so the t-statistic is asymptotically normally distributed as stated in the theorem. In the borderline setting $\alpha - \beta = 1/2$, $(\hat{\sigma}_t^n)^2$ is still dominated by $\Theta_{2,n}$, while $\hat{\mu}_t^n$ is influenced by both A_n and B_n. The limit in distribution

is then:

$$c_{K,\beta} N(0, 1) + \frac{c_{1,1}}{c_{2,1}} \frac{m_K(-\alpha)}{(K_2 m_K(-2\beta))^{1/2}}.$$

This concludes the proof. ∎

Proof of Theorem 3. As in the proof of Theorem 1, we write:

$$T_t^n = \sqrt{\frac{h_n}{K_2}} \frac{(\hat{\mu}_t^n - \mu_{t-}^*)}{\hat{\sigma}_t^n} + \sqrt{\frac{h_n}{K_2}} \frac{\mu_{t-}^*}{\hat{\sigma}_t^n} = \underbrace{\sqrt{\frac{h_n}{K_2}} \frac{(\hat{\mu}_t^n - \mu_{t-}^*)}{\sigma_{t-}}}_{T_{t,1}^n} + \underbrace{\sqrt{\frac{h_n}{K_2}} \frac{(\hat{\mu}_t^n - \mu_{t-}^*)}{\sigma_{t-}} \left(\frac{\sigma_{t-}}{\hat{\sigma}_t^n} - 1 \right)}_{T_{t,2}^n} + \underbrace{\sqrt{\frac{h_n}{K_2}} \frac{\mu_{t-}^*}{\hat{\sigma}_t^n}}_{O_p(\sqrt{h_n})}. \tag{45}$$

The last term is handled as in the proof of Theorem 1, which shows that $|T_{t,3}^n| = O_p(\sqrt{h_n})$. This order is uniform in t, so that $\max_{j=1,\ldots,m} |T_{t_j^*,3}^n| = O_p(m\sqrt{h_n})$ by Boole's inequality. Next, we note that $T_{t,2}^n = T_{t,1}^n \left(\frac{\sigma_{t-}}{\hat{\sigma}_t^n} - 1 \right) = T_{t,1}^n o_p(1)$ by Lemma 3, so this term is negligible relative to $T_{t,1}^n$.

Turning to $T_{t,1}^n$, it follows from the proof of Theorem 1:

$$T_{t,1}^n = \sqrt{\frac{1}{K_2}} \frac{G_{n,t} + G_{n,t}' + O_p \left(\frac{\sqrt{h_n}}{n h_n} + h_n^{\Gamma/2+1/2} + h_n^{B+1/2} \right)}{\sigma_{t-}},$$

where the $O_p \left(\frac{\sqrt{h_n}}{n h_n} + h_n^{\Gamma/2+1/2} + h_n^{B+1/2} \right)$ term is uniform in t. Moreover, by Jacod and Protter (2012, Lemma 2.1.5) $G_{n,t}' = O_p \left(\sqrt{h_n} \right)$, uniformly in t.

We decompose $G_{n,t_j^*}/\sigma_{t_j^*-}$, with $j = 1, \ldots, m$, as follows:

$$\frac{G_{n,t_j^*}}{\sigma_{t_j^*-}} = \frac{1}{\sigma_{t_j^*-}\sqrt{h_n}} \sum_{i=1}^n K \left(\frac{t_{i-1} - t_j^*}{h_n} \right) I_{\left\{ t_{j-1}^* < t_i \le t_j^* \right\}} \int_{t_{i-1}}^{t_i} \sigma_s dW_s + \frac{1}{\sigma_{t_j^*-}\sqrt{h_n}} \sum_{i=1}^n K \left(\frac{t_{i-1} - t_j^*}{h_n} \right) I_{\left\{ t_i \le t_{j-1}^* \right\}} \int_{t_{i-1}}^{t_i} \sigma_s dW_s$$

$$= \underbrace{\frac{1}{\sqrt{h_n}} \sum_{i=1}^n K \left(\frac{t_{i-1} - t_j^*}{h_n} \right) I_{\left\{ t_{j-1}^* < t_i \le t_j^* \right\}} \left(W_{t_i} - W_{t_{i-1}} \right)}_{G_{t_j^*,1}^n}$$

$$+ \underbrace{\frac{1}{\sigma_{t_j^*-}\sqrt{h_n}} \sum_{i=1}^n K \left(\frac{t_{i-1} - t_j^*}{h_n} \right) I_{\left\{ t_{j-1}^* < t_i \le t_j^* \right\}} \int_{t_{i-1}}^{t_i} \left(\sigma_s - \sigma_{t_j^*-} \right) dW_s}_{G_{t_j^*,2}^n}$$

$$+ \underbrace{\frac{1}{\sigma_{t_j^*-}\sqrt{h_n}} \sum_{i=1}^n K \left(\frac{t_{i-1} - t_j^*}{h_n} \right) I_{\left\{ t_i \le t_{j-1}^* \right\}} \int_{t_{i-1}}^{t_i} \sigma_s dW_s}_{G_{t_j^*,3}^n}.$$

Note $G_{t_{j_1}^*,1}^n$ and $G_{t_{j_2}^*,1}^n$ are independent for $j_1 \ne j_2$. Moreover, $\sqrt{\frac{1}{K_2}} G_{n,t_j^*}$ is, for each j, normally distributed with mean zero and unit asymptotic variance, so as in the proof of Lemma 1 in Lee and Mykland (2008):

$$a_m \left(\max_{j=1,\ldots,m} \sqrt{\frac{1}{K_2}} G_{t_j^*,1}^n - b_m \right) \xrightarrow{d} \xi.$$

For the second term, we write:

$$\mathscr{P} \left(\max_{j=1,\ldots,m} \left| G_{t_j^*,2}^n \right| \ge \psi_{n,m} \right) \le \mathscr{P} \left(\max_{j=1,\ldots,m} \left| G_{t_j^*,2}^n \right| \ge \psi_{n,m}, \max_{j=1,\ldots,m} \left[G_{t_j^*,2}^n \right] \le \beta_{n,m} \right) + \mathscr{P} \left(\max_{j=1,\ldots,m} \left[G_{t_j^*,2}^n \right] > \beta_{n,m} \right),$$

with

$$
\left[G_{t_j^*,2}^n\right] = \frac{1}{\sigma_{t_j^*-}h_n^2}\sum_{i=1}^n K^2\left(\frac{t_{i-1}-t_j^*}{h_n}\right)I_{\left\{t_{j-1}^*<t_i\le t_j^*\right\}}\int_{t_{i-1}}^{t_i}\left(\sigma_s-\sigma_{t_j^*-}\right)^2 ds
$$

$$
\le Cm^{-\Gamma}\frac{1}{h_n}\sum_{i=1}^n K^2\left(\frac{t_{i-1}-t_j^*}{h_n}\right)\Delta_{i,n}\le Cm^{-\Gamma},
$$

where C is uniform over j.

Applying the exponential inequality from Dzhaparidze and van Zanten (2001), we conclude that:

$$
\mathscr{P}\left(\max_{j=1,\dots,m}\left|G_{t_j^*,2}^n\right|\ge C_n,\ \max_{j=1,\dots,m}\left[G_{t_j^*,2}^n\right]\le\beta_n\right)\le 2m\exp\left(-\frac{C_n^2}{2\beta_n}\right),
$$

while Markov's inequality yields

$$
\mathscr{P}\left(\max_{j=1,\dots,m}\left[G_{t_j^*,2}^n\right]>\beta_{n,m}\right)\le\frac{E\left(\max_{j=1,\dots,m}\left[G_{t_j^*,2}^n\right]\right)}{\beta_{n,m}}\le C\frac{m^{-\Gamma}}{\beta_{n,m}}.
$$

Setting $\beta_{n,m}=cm^{-\Gamma}$ and $\psi_{n,m}=cm^{-\Gamma/2}\sqrt{\log m}$, this reduces to:

$$
\mathscr{P}\left(\frac{1}{\psi_{n,m}}\max_{j=1,\dots,m}\left|G_{t_j^*,2}^n\right|\ge c\right)\le 2m^{1-c/2}+\frac{C}{c},
$$

which shows that

$$
\max_{j=1,\dots,m}\left|G_{t_j^*,2}^n\right|=O_p\left(m^{-\Gamma/2}\sqrt{\log(m)}\right).
$$

The fast vanishing tails of the kernel means that for $G_{t_j^*,3}^n$:

$$
\left[G_{t_j^*,3}^n\right]=\frac{1}{\sigma_{t_j^*-}^2 h_n}\sum_{i=1}^n K^2\left(\frac{t_{i-1}-t_j^*}{h_n}\right)I_{\left\{t_i\le t_{j-1}^*\right\}}\int_{t_{i-1}}^{t_i}\sigma_s^2 ds
$$

$$
\le C(mh_n)^{-2B},
$$

which implies, arguing as before,

$$
\max_{j=1,\dots,m}\left|G_{t_j^*,3}^n\right|=O_p\left((mh_n)^{-B}\sqrt{\log(m)}\right).
$$

Thus, if $a_m\left(\frac{m}{\sqrt{nh_n}}+(mh_n)^{-B}\sqrt{\log(m)}+m^{-\Gamma/2}\sqrt{\log(m)}\right)\to 0$, $a_m\left(T_m^*-b_m\right)\xrightarrow{d}\xi$. ∎

Proof of Theorem 4. We again set $\tau_{\mathrm{db}}=\tau_J=T=1$. Then,

$$
T_{\tau_J}^n=\sqrt{\frac{h_n}{K_2}}\frac{\hat{\mu}_t^n}{\hat{\sigma}_t^n}=\sqrt{\frac{h_n}{K_2}}\frac{\frac{1}{h_n}\sum_{i=1}^n K\left(\frac{t_{i-1}-1}{h_n}\right)\Delta_i^n X+\frac{1}{h_n}K\left(\frac{\Delta_{n,n}}{h_n}\right)J}{\left(\frac{1}{h_n}\sum_{i=1}^n K\left(\frac{t_{i-1}-1}{h_n}\right)(\Delta_i^n X)^2+\frac{2}{h_n}K\left(\frac{\Delta_{n,n}}{h_n}\right)J\Delta_n^n X+\frac{1}{h_n}K\left(\frac{\Delta_{n,n}}{h_n}\right)J^2\right)^{1/2}}
$$

$$
=\frac{N\left(0,\sigma_{t-}^2\right)+\sqrt{\frac{1}{K_2 h_n}}K\left(\frac{\Delta_{n,n}}{h_n}\right)J+o_p(1)}{\left(\sigma_{t-}^2+\frac{2}{h_n}K\left(\frac{\Delta_{n,n}}{h_n}\right)JO_p\left(\sqrt{\Delta_n}\right)+\frac{1}{h_n}K\left(\frac{\Delta_{n,n}}{h_n}\right)J^2+o_p(1)\right)^{1/2}}\xrightarrow{p}\sqrt{\frac{K(0)}{K_2}}\cdot\mathrm{sign}(J),
$$

as $n\to\infty$, $h_n\to 0$ and $nh_n\to\infty$. ∎

Proof of Theorem 5. As in the proof of Theorem 1, we write:

$$
\bar{T}_t^n=\underbrace{\sqrt{\frac{h_n}{K_2}}\frac{(\hat{\bar{\mu}}_t^n-\mu_{t-}^*)}{\sqrt{\hat{\bar{\sigma}}_t^n}}}_{\bar{\tau}_1}+\underbrace{\sqrt{\frac{h_n}{K_2}}\frac{\mu_{t-}^*}{\sqrt{\hat{\bar{\sigma}}_t^n}}}_{\bar{\tau}_2},
$$

where $\bar{\bar{T}}_2 = O_p(\sqrt{h_n})$ is negligible. We dissect $\hat{\bar{\mu}}_t^n$ into an efficient log-price and noise component:

$$\hat{\bar{\mu}}_t^n = \underbrace{\frac{1}{h_n} \sum_{i=1}^{n-k_n+2} K\left(\frac{t_{i-1}-t}{h_n}\right) \Delta_{i-1}^n \bar{X}}_{M_{X,n}} + \underbrace{\frac{1}{h_n} \sum_{i=1}^{n-k_n+2} K\left(\frac{t_{i-1}-t}{h_n}\right) \Delta_{i-1}^n \bar{\epsilon}}_{M_{\epsilon,n}}.$$

The strategy is again to verify Theorem 2.2.14 in Jacod and Protter (2012). This is more involved now because the summands in the drift estimator are k_n-dependent with $k_n \to \infty$ due to the pre-averaging. We therefore apply a block-splitting technique as in Jacod et al. (2009).

We start with the noise term and write, for an integer $p \geq 2$,

$$M_{\epsilon,n} = M(p)_t^n + M'(p)_t^n + \hat{C}(p)_t^n,$$

where

$$M(p)_t^n = \frac{1}{h_n} \sum_{j=0}^{j_n(p)} \sum_{\ell=\bar{\ell}_j^n(p)}^{\bar{\ell}_j^n(p)+pk_n-1} K\left(\frac{t_{\ell-1}-t}{h_n}\right) \Delta_{\ell-1}^n \bar{\epsilon},$$

$$M'(p)_t^n = \frac{1}{h_n} \sum_{j=0}^{j_n(p)} \sum_{\ell=\bar{\ell}_j^n(p)+pk_n}^{\bar{\ell}_j^n(p)+pk_n+k_n-1} K\left(\frac{t_{\ell-1}-t}{h_n}\right) \Delta_{\ell-1}^n \bar{\epsilon},$$

$$\hat{C}(p)_t^n = \frac{1}{h_n} \sum_{\ell=\bar{\ell}_{j_n(p)+1}^n(p)}^{n-k_n+2} K\left(\frac{t_{\ell-1}-t}{h_n}\right) \Delta_{\ell-1}^n \bar{\epsilon},$$

with $j_n(p) = \left[\frac{(n+1)}{(p+1)k_n}\right] - 1$ and $\bar{\ell}_j^n(p) = j(pk_n - 1) + jk_n + 1$. The term $M(p)_t^n$ is a sum of "big" blocks of dimension pk_n, while the term $M'(p)_t^n$ is a sum of "small" blocks of dimension k_n, which separate the big blocks. $\hat{C}(p)_t^n$ is an end effect. $M(p)_t^n$ is shown to be the dominating term. It can be written as:

$$M(p)_t^n \equiv \sum_{j=0}^{j_n(p)} u_j^n,$$

where $u_j^n = \frac{1}{h_n} \sum_{\ell=\bar{\ell}_j^n(p)}^{\bar{\ell}_j^n(p)+pk_n-1} K\left(\frac{t_{\ell-1}-t}{h_n}\right) \Delta_{\ell-1}^n \bar{\epsilon}$ is by construction independent on $u_{j'}$ when $j' \neq j$ and $k_n > Q + 1$. We can then employ Theorem 2.2.14 in Jacod and Protter (2012) by taking expectations conditional on the discrete-time filtration $\mathscr{G}(p)_j^n = \mathscr{F}_{t_{\bar{\ell}_j^n(p)-1}}$. We immediately get the orthogonality condition (44) and

$$\sum_{j=0}^{j_n(p)} E\left[u_j^n \mid \mathscr{G}(p)_j^n\right] = 0.$$

As for the conditional variance:

$$\sum_{j=0}^{j_n(p)} E\left[(u_j^n)^2 \mid \mathscr{G}(p)_j^n\right] = \frac{1}{h_n^2} \sum_{j=0}^{j_n(p)} E\left[\left(\sum_{\ell=\bar{\ell}_j^n(p)}^{\bar{\ell}_j^n(p)+pk_n-1} K\left(\frac{t_{\ell-1}-t}{h_n}\right) \Delta_{\ell-1}^n \bar{\epsilon}\right)^2 \mid \mathscr{G}(p)_j^n\right]$$

$$= \frac{1}{h_n^2} \sum_{j=0}^{j_n(p)} \sum_{\ell=\bar{\ell}_j^n(p)}^{\bar{\ell}_j^n(p)+pk_n-1} K^2\left(\frac{t_{\ell-1}-t}{h_n}\right) E\left[(\Delta_{\ell-1}^n \bar{\epsilon})^2 \mid \mathscr{G}(p)_j^n\right]$$

$$+ \frac{2}{h_n^2} \sum_{j=0}^{j_n(p)} \sum_{\ell=\bar{\ell}_j^n(p)}^{\bar{\ell}_j^n(p)+pk_n-1} \sum_{\ell'=\ell+1}^{\bar{\ell}_j^n(p)+pk_n-1} K\left(\frac{t_{\ell-1}-t}{h_n}\right) K\left(\frac{t_{\ell'-1}-t}{h_n}\right) E\left[\Delta_{\ell-1}^n \bar{\epsilon} \Delta_{\ell'-1}^n \bar{\epsilon} \mid \mathscr{G}(p)_j^n\right].$$

By the mean value theorem,

$$K\left(\frac{t_{\ell'-1}-t}{h_n}\right) = K\left(\frac{t_{\ell-1}-t}{h_n}\right) + K'\left(\frac{\xi_{t_{\ell-1},t_{\ell'-1}}}{h_n}\right) \frac{t_{\ell'-1}-t_{\ell-1}}{h_n},$$

for a $\xi_{t_{\ell-1},t_{\ell'-1}} \in]t_{\ell'-1} - t, t_{\ell-1} - t[$, so that

$$
\sum_{j=0}^{j_n(p)} E\left[(u_j^n)^2 \mid \mathscr{G}(p)_j^n\right]
$$

$$
= \frac{1}{h_n^2} \sum_{j=0}^{j_n(p)} \sum_{\ell=\bar{\ell}_j^n(p)}^{\bar{\ell}_j^n(p)+pk_n-1} K^2\left(\frac{t_{\ell-1}-t}{h_n}\right) \underbrace{\left(E\left[(\Delta_{\ell-1}^n\bar{\epsilon})^2 \mid \mathscr{G}(p)_j^n\right] + 2 \sum_{\ell'=\ell+1}^{\bar{\ell}_j^n(p)+pk_n-1} E\left[\Delta_{\ell-1}^n\bar{\epsilon}\Delta_{\ell'-1}^n\bar{\epsilon} \mid \mathscr{G}(p)_j^n\right]\right)}_{V_{1,n}}
$$

$$
+ \underbrace{\frac{2}{h_n^2} \sum_{j=0}^{j_n(p)} \sum_{\ell=\bar{\ell}_j^n(p)}^{\bar{\ell}_j^n(p)+pk_n-1} \sum_{\ell'=\ell+1}^{\bar{\ell}_j^n(p)+pk_n-1} K\left(\frac{t_{\ell-1}-t}{h_n}\right) K'\left(\frac{\xi_{t_{\ell-1},t_{\ell'-1}}}{h_n}\right) \frac{t_{\ell'-1}-t_{\ell-1}}{h_n} E\left[\Delta_{\ell-1}^n\bar{\epsilon}\Delta_{\ell'-1}^n\bar{\epsilon} \mid \mathscr{G}(p)_j^n\right]}_{V_{2,n}}.
$$

The term inside the first summation is the long-run variance of a stationary time series that is over-differenced, so $V_{1,n} = 0$ (see e.g. Hassler, 2016, pp. 305–306). The conditional variance is thus dominated by the limit of $V_{2,n}$.

We use the multinomial theorem for the fourth conditional moment:

$$
\frac{h_n^4}{k_n^2} \sum_{j=0}^{j_n(p)} E\left[(u_j^n)^4 \mid \mathscr{G}(p)_j^n\right] = \frac{1}{k_n^2} \sum_{j=0}^{j_n(p)} E\left[\left(\sum_{\ell=\bar{\ell}_j^n(p)}^{\bar{\ell}_j^n(p)+pk_n-1} K\left(\frac{t_{\ell-1}-t}{h_n}\right) \Delta_{\ell-1}^n\bar{\epsilon}\right)^4 \mid \mathscr{G}(p)_j^n\right]
$$

$$
= \frac{1}{k_n^2} \sum_{j=0}^{j_n(p)} \left(\sum_{\ell=\bar{\ell}_j^n(p)}^{\bar{\ell}_j^n(p)+pk_n-1} K^4\left(\frac{t_{\ell-1}-t}{h_n}\right) E\left[(\Delta_{\ell-1}^n\bar{\epsilon})^4\right]\right.
$$

$$
+ 4 \sum_{\ell_1\neq\ell_2} K^3\left(\frac{t_{\ell_1-1}-t}{h_n}\right) K\left(\frac{t_{\ell_2-1}-t}{h_n}\right) E\left[(\Delta_{\ell_1-1}^n\bar{\epsilon})^3 \Delta_{\ell_2-1}^n\bar{\epsilon}\right]
$$

$$
+ 6 \sum_{\ell_1\neq\ell_2} K^2\left(\frac{t_{\ell_1-1}-t}{h_n}\right) K^2\left(\frac{t_{\ell_2-1}-t}{h_n}\right) E\left[(\Delta_{\ell_1-1}^n\bar{\epsilon})^2 (\Delta_{\ell_2-1}^n\bar{\epsilon})^2\right]
$$

$$
+ 12 \sum_{\ell_1\neq\ell_2\neq\ell_3} K^2\left(\frac{t_{\ell_1-1}-t}{h_n}\right) K\left(\frac{t_{\ell_2-1}-t}{h_n}\right) K\left(\frac{t_{\ell_3-1}-t}{h_n}\right) E\left[(\Delta_{\ell_1-1}^n\bar{\epsilon})^2 \Delta_{\ell_2-1}^n\bar{\epsilon}\Delta_{\ell_3-1}^n\bar{\epsilon}\right]
$$

$$
+ 24 \sum_{\ell_1\neq\ell_2\neq\ell_3\neq\ell_4} K\left(\frac{t_{\ell_1-1}-t}{h_n}\right) K\left(\frac{t_{\ell_2-1}-t}{h_n}\right) K\left(\frac{t_{\ell_3-1}-t}{h_n}\right) K\left(\frac{t_{\ell_4-1}-t}{h_n}\right) E\left[\Delta_{\ell_1-1}^n\bar{\epsilon}\Delta_{\ell_2-1}^n\bar{\epsilon}\Delta_{\ell_3-1}^n\bar{\epsilon}\Delta_{\ell_4-1}^n\bar{\epsilon}\right]\right).
$$

Now, since $\ell_1, \ell_2, \ell_3, \ell_4$ are no more than $O(pk_n\Delta_n/h_n)$ terms apart, which is going to zero in the limit, we can mean value expand the kernel again to show that

$$
\frac{h_n^4}{k_n^2} \sum_{j=0}^{j_n(p)} E\left[(u_j^n)^4 \mid \mathscr{G}(p)_j^n\right] = Q_{1,n} + Q_{2,n},
$$

where, as for the variance term,

$$
Q_{1,n} = \frac{1}{k_n^2} \sum_{j=0}^{j_n(p)} \left(\sum_{\ell=\bar{\ell}_j^n(p)}^{\bar{\ell}_j^n(p)+pk_n-1} K^4\left(\frac{t_{\ell-1}-t}{h_n}\right) \left(E\left[(\Delta_{\ell-1}^n\bar{\epsilon})^4\right] + 4 \sum_{\ell_1\neq\ell_2} E\left[(\Delta_{\ell_1-1}^n\bar{\epsilon})^3 \Delta_{\ell_2-1}^n\bar{\epsilon}\right]\right.\right.
$$

$$
+ 6 \sum_{\ell_1\neq\ell_2} E\left[(\Delta_{\ell_1-1}^n\bar{\epsilon})^2 (\Delta_{\ell_2-1}^n\bar{\epsilon})^2\right]
$$

$$
\left.\left.+ 12 \sum_{\ell_1\neq\ell_2\neq\ell_3} E\left[(\Delta_{\ell_1-1}^n\bar{\epsilon})^2 \Delta_{\ell_2-1}^n\bar{\epsilon}\Delta_{\ell_3-1}^n\bar{\epsilon}\right] + 24 \sum_{\ell_1\neq\ell_2\neq\ell_3\neq\ell_4} E\left[\Delta_{\ell_1-1}^n\bar{\epsilon}\Delta_{\ell_2-1}^n\bar{\epsilon}\Delta_{\ell_3-1}^n\bar{\epsilon}\Delta_{\ell_4-1}^n\bar{\epsilon}\right]\right)\right) = 0.
$$

With $\ell_1 + \ell_2 + \ell_3 + \ell_4 = 4$, the boundedness of the fourth moment of the noise and the first equation of (5.36) in Jacod et al. (2009), showing each $\Delta_{\ell-1}^n \bar{\epsilon}$ is of order $k_n^{-1/2}$, it follows that

$$E\left[\left(\Delta_{\ell_1-1}^n \bar{\epsilon}\right)^{\ell_1} \left(\Delta_{\ell_2-1}^n \bar{\epsilon}\right)^{\ell_2} \left(\Delta_{\ell_3-1}^n \bar{\epsilon}\right)^{\ell_3} \left(\Delta_{\ell_4-1}^n \bar{\epsilon}\right)^{\ell_4}\right] \leq C k_n^{-2}.$$

We deduce that:

$$
\begin{aligned}
|Q_{2,n}| \leq C \frac{p k_n \Delta_n}{h_n} \frac{1}{k_n^4} \sum_{j=0}^{j_n(p)} &\left(\sum_{\ell_1 \neq \ell_2} K^3\left(\frac{t_{\ell_1-1}-t}{h_n}\right) K'\left(\frac{t_{\ell_2-1}-t}{h_n}\right) + \sum_{\ell_1 \neq \ell_2} K^2\left(\frac{t_{\ell_1-1}-t}{h_n}\right) (K^2)'\left(\frac{t_{\ell_2-1}-t}{h_n}\right) \right. \\
&+ \sum_{\ell_1 \neq \ell_2 \neq \ell_3} K^2\left(\frac{t_{\ell_1-1}-t}{h_n}\right) \left(K'\left(\frac{t_{\ell_2-1}-t}{h_n}\right) K\left(\frac{t_{\ell_3-1}-t}{h_n}\right) + K\left(\frac{t_{\ell_2-1}-t}{h_n}\right) K'\left(\frac{t_{\ell_3-1}-t}{h_n}\right) + o\left(\frac{p k_n \Delta_n}{h_n}\right)\right) \\
&+ \sum_{\ell_1 \neq \ell_2 \neq \ell_3 \neq \ell_4} K\left(\frac{t_{\ell_1-1}-t}{h_n}\right) \left(K'\left(\frac{t_{\ell_2-1}-t}{h_n}\right) K\left(\frac{t_{\ell_3-1}-t}{h_n}\right) K\left(\frac{t_{\ell_4-1}-t}{h_n}\right) \right. \\
&+ K\left(\frac{t_{\ell_2-1}-t}{h_n}\right) K'\left(\frac{t_{\ell_3-1}-t}{h_n}\right) K\left(\frac{t_{\ell_4-1}-t}{h_n}\right) \\
&\left.\left. + K\left(\frac{t_{\ell_2-1}-t}{h_n}\right) K\left(\frac{t_{\ell_3-1}-t}{h_n}\right) K'\left(\frac{t_{\ell_4-1}-t}{h_n}\right) + o\left(\frac{p k_n \Delta_n}{h_n}\right)\right)\right) = O_p\left(\frac{1}{k_n^3}\right).
\end{aligned}
$$

$M'(p)_t^n$ can be bounded by Doob's inequality:

$$
\begin{aligned}
E\left[\sup_{s \leq t} \left|M'(p)_s^n\right|^2\right] &\leq 4 \frac{1}{h_n^2} \sum_{j=0}^{j_n(p)} E\left[\left(\sum_{\ell=\bar{\ell}_j^n(p)+pk_n}^{\bar{\ell}_j^n(p)+pk_n+k_n-1} K\left(\frac{t_{\ell-1}-t}{h_n}\right) \Delta_{\ell-1}^n \bar{\epsilon}\right)^2\right] \\
&= \frac{1}{h_n^2} \sum_{j=0}^{j_n(p)} \sum_{\ell} \sum_{\ell'=\ell+1} K\left(\frac{t_{\ell-1}-t}{h_n}\right) K'\left(\frac{\xi_{t_{\ell-1},t_{\ell'-1}}}{h_n}\right) \frac{t_{\ell'-1}-t_{\ell-1}}{h_n} E\left[\Delta_{\ell-1}^n \bar{\epsilon} \cdot \Delta_{\ell'-1}^n \bar{\epsilon}\right]
\end{aligned}
$$

which is negligible compared to $V_{2,n}$, as it asymptotically is an identical term summed over a small block.

Finally, the end-effect term can also be neglected, as

$$\left|\frac{1}{h_n} \sum_{\ell=\bar{\ell}_{j_n(p)+1}^n(p)}^{n-k_n+2} K\left(\frac{t_{\ell-1}-t}{h_n}\right) \Delta_{\ell-1}^n \bar{\epsilon}\right| \leq \frac{1}{h_n} \sum_{\ell=\bar{\ell}_{j_n(p)+1}^n(p)}^{n-k_n+2} K\left(\frac{t_{\ell-1}-t}{h_n}\right) |\Delta_{\ell-1}^n \bar{\epsilon}| \leq C \frac{k_n^{1/2}}{h_n},$$

since $\Delta_{\ell-1}^n \bar{\epsilon} = O_p\left(k_n^{-1/2}\right)$.

We next analyze the $M_{X,n}$ term and show this is negligible too. We write $M_{X,n} = \tilde{M}(p)_t^n + \tilde{M}'(p)_t^n + \widehat{\tilde{C}}(p)_t^n$ with an identical decomposition as for the $M_{\epsilon,n}$ term. Arguing as above, the dominating term is $\tilde{M}(p)_t^n$. We decompose $\tilde{M}(p)_t^n \equiv \sum_{j=0}^{j_n(p)} \tilde{u}_j^n$, where $\tilde{u}_j^n = \frac{1}{h_n} \sum_{\ell=\bar{\ell}_j^n(p)}^{\bar{\ell}_j^n(p)+pk_n-1} K\left(\frac{t_{\ell-1}-t}{h_n}\right) \Delta_{\ell-1}^n \bar{X}$ and then compute:

$$
\begin{aligned}
\sum_{j=0}^{j_n(p)} E\left[(\tilde{u}_j^n)^2 \mid \mathscr{G}(p)_j^n\right] &= \frac{1}{h_n^2} \sum_{j=0}^{j_n(p)} E\left[\left(\sum_{\ell=\bar{\ell}_j^n(p)}^{\bar{\ell}_j^n(p)+pk_n-1} K\left(\frac{t_{\ell-1}-t}{h_n}\right) \Delta_{\ell-1}^n \bar{X}\right)^2 \mid \mathscr{G}(p)_j^n\right] \\
&= \frac{1}{h_n^2} \sum_{j=0}^{j_n(p)} \sum_{\ell=\bar{\ell}_j^n(p)}^{\bar{\ell}_j^n(p)+pk_n-1} K^2\left(\frac{t_{\ell-1}-t}{h_n}\right) E\left[(\Delta_{\ell-1}^n \bar{X})^2 \mid \mathscr{G}(p)_j^n\right] \\
&+ \frac{2}{h_n^2} \sum_{j=0}^{j_n(p)} \sum_{\ell=\bar{\ell}_j^n(p)}^{\bar{\ell}_j^n(p)+pk_n-1} \sum_{\ell'=\ell+1}^{\bar{\ell}_j^n(p)+pk_n-1} K\left(\frac{t_{\ell-1}-t}{h_n}\right) K\left(\frac{t_{\ell'-1}-t}{h_n}\right) E\left[\Delta_{\ell-1}^n \bar{X} \Delta_{\ell'-1}^n \bar{X} \mid \mathscr{G}(p)_j^n\right] \\
&= \frac{2}{h_n^2} \sum_{j=0}^{j_n(p)} \sum_{\ell=\bar{\ell}_j^n(p)}^{\bar{\ell}_j^n(p)+pk_n-1} \sum_{\ell'=\ell+1}^{\bar{\ell}_j^n(p)+pk_n-1} K\left(\frac{t_{\ell-1}-t}{h_n}\right) K'\left(\frac{\xi_{t_{\ell-1},t_{\ell'-1}}}{h_n}\right) \frac{t_{\ell'-1}-t_{\ell-1}}{h_n} \\
&\times E\left[\Delta_{\ell-1}^n \bar{X} \Delta_{\ell'-1}^n \bar{X} \mid \mathscr{G}(p)_j^n\right],
\end{aligned}
$$

where

$$E\left[\Delta_{\ell-1}^n \bar{X} \Delta_{\ell'-1}^n \bar{X} \mid \mathscr{G}(p)_j^n\right] = E\left[\left(\sum_{j=0}^{k_n-1} H_j^n \Delta_i^n X_{\ell+j-1}\right)\left(\sum_{j=0}^{k_n-1} H_j^n \Delta_i^n X_{\ell'+j-1}\right) \mid \mathscr{G}(p)_j^n\right] = \sum_{j=0}^{k_n-\ell'+\ell-1} H_j^n H_{j+\ell'-\ell}^n \int_{t_{j-1}}^{t_j} \sigma_s^2 ds.$$

We conclude that $\tilde{M}(p)_t^n = O_p\left(\frac{\sqrt{\Delta_n k_n}}{h_n}\right)$ is of smaller order than $M_{\epsilon,n}$.

Next, we investigate the estimator of the long-run variance of $\hat{\bar{\mu}}_t^n$, i.e. $\hat{\bar{\sigma}}_t^n$. We only analyze the market microstructure component, which is the leading term:

$$E\left[\hat{\bar{\sigma}}_t^n\right] = \frac{1}{h_n}\sum_{i=1}^{n-k_n+2} K^2\left(\frac{t_{i-1}-t}{h_n}\right) E\left[\left(\Delta_{i-1}^n \bar{Y}\right)^2\right] + \frac{2}{h_n}\sum_{L=1}^{L_n} w\left(\frac{L}{L_n}\right) \sum_{i=1}^{n-k_n-L+2} K\left(\frac{t_{i-1}-t}{h_n}\right) K\left(\frac{t_{i+L-1}-t}{h_n}\right)$$

$$\times E\left[\Delta_{i-1}^n \bar{Y} \Delta_{i-1+L}^n \bar{Y}\right]$$

$$= \frac{1}{h_n}\sum_{i=1}^{n-k_n+2} K^2\left(\frac{t_{i-1}-t}{h_n}\right) E\left[\left(\Delta_{i-1}^n \bar{\epsilon}\right)^2\right]$$

$$+ \frac{2}{h_n}\sum_{L=1}^{L_n} w\left(\frac{L}{L_n}\right) \sum_{i=1}^{n-k_n-L+2} K\left(\frac{t_{i-1}-t}{h_n}\right) K\left(\frac{t_{i+L-1}-t}{h_n}\right) E\left[\Delta_{i-1}^n \bar{\epsilon} \Delta_{i-1+L}^n \bar{\epsilon}\right] + R_n''$$

$$\equiv \sigma_{0,n} + R_n'',$$

where R_n'' is asymptotically negligible, following the calculation of the limiting variance of the drift estimator. Assuming $L_n\Delta_n/h_n \to 0$, we write $K\left(\frac{t_{i+L-1}-t}{h_n}\right) = K\left(\frac{t_{i-1}-t}{h_n}\right) + K'\left(\frac{\xi_{i-1,i+L-1}-t}{h_n}\right)\frac{L\Delta_n}{h_n}$ and decompose $\sigma_{0,n}$ into:

$$\sigma_{0,n} = \underbrace{\frac{1}{h_n}\sum_{i=1}^{n-k_n+2} K^2\left(\frac{t_{i-1}-t}{h_n}\right) E\left[\left(\Delta_{i-1}^n \bar{\epsilon}\right)^2\right] + \frac{2}{h_n}\sum_{L=1}^{L_n}\sum_{i=1}^{n-k_n-L+2} K^2\left(\frac{t_{i-1}-t}{h_n}\right) E\left[\Delta_{i-1}^n \bar{\epsilon}\Delta_{i-1+L}^n \bar{\epsilon}\right]}_{\sigma_{1,n}}$$

$$+ \underbrace{\frac{2}{h_n}\sum_{L=1}^{L_n} w\left(\frac{L}{L_n}\right)\sum_{i=1}^{n-k_n-L+2} K\left(\frac{t_{i-1}-t}{h_n}\right) K'\left(\frac{\xi_{i-1,i+L-1}}{h_n}\right) L\frac{\Delta_n}{h_n} E\left[\Delta_{i-1}^n \bar{\epsilon}\Delta_{i-1+L}^n \bar{\epsilon}\right]}_{\sigma_{2,n}},$$

where $\sigma_{1,n} = 0$. Since $w(L/L_n) = 1 + O\left(\frac{L}{L_n}\right)$ after the proper rescaling the limit of $\sigma_{2,n}$ and $V_{2,n}$ are identical. ∎

Proof of Theorem 6. As in the proof of Theorem 2, we set $\tau_{db} = 1$ and further $c_{1,t} = c_{2,t} = 1$. We also write $\widetilde{X}_t = X_t + D_t + V_t$. Based on Theorem 5, $\frac{1}{h_n}\sum_{i=1}^{n-k_n+1} K\left(\frac{t_{i-1}-t}{h_n}\right)\left(\Delta_{i-1}^n \bar{X} + \Delta_{i-1}^n \bar{\epsilon}\right) = O_p\left(\frac{\sqrt{k_n}}{h_n}\right)$. Next, we note that for suitable $\xi_j \in [t_j, t_{j+1}]$ and invoking again the mean-value theorem and ignoring end-effects:

$$\frac{1}{h_n}\sum_{i=1}^{n-k_n+2} K\left(\frac{t_{i-1}-1}{h_n}\right)\Delta_{i-1}^n \bar{D} = \frac{1}{h_n}\sum_{i=1}^{n-k_n+2} K\left(\frac{t_{i-1}-1}{h_n}\right)\sum_{j=1}^{k_n-1} g_j^n(D_{t_{i+j-1}} - D_{t_{i+j-2}})$$

$$= \frac{1}{h_n}\sum_{i=1}^{n-k_n+2} K\left(\frac{t_{i-1}-1}{h_n}\right)\sum_{j=1}^{k_n-1} g_j^n \frac{(1-t_{i+j})^{1-\alpha} - (1-t_{i+j-1})^{1-\alpha}}{1-\alpha}$$

$$= \frac{1}{h_n}\sum_{i=1}^{n-k_n+2} K\left(\frac{t_{i-1}-1}{h_n}\right)\sum_{j=1}^{k_n-1} g_j^n(1-\xi_{i+j})^{-\alpha}\Delta_{i+j,n}$$

$$= \frac{1}{h_n}\sum_{i=1}^{n-k_n+2} K\left(\frac{t_{i-1}-1}{h_n}\right)(1-t_{i-1}+O(k_n\Delta_n))^{-\alpha}\Delta_{i,n}\sum_{j=1}^{k_n-1} g_j^n.$$

As $\frac{1}{k_n}\sum_{j=1}^{k_n-1} g_j^n \to \psi_0 = \int_0^1 g(s)ds$, the above has asymptotic rate $k_n h_n^{-\alpha}$ (as for the term A_n in the proof of Theorem 2) if $k_n\Delta_n \to 0$, which is assured by assumption. This dominates the noise term if $\frac{k_n h_n^{-\alpha}}{\frac{\sqrt{k_n}}{h_n}} \to \infty$, i.e. $\sqrt{k_n}h_n^{1-\alpha} \to \infty$.

The pre-averaged volatility burst term can be written as

$$
\frac{1}{h_n} \sum_{i=1}^{n-k_n+2} K\left(\frac{t_{i-1}-1}{h_n}\right) \Delta_{i-1}^n \bar{V} = \frac{1}{h_n} \sum_{i=1}^{n-k_n+2} K\left(\frac{t_{i-1}-1}{h_n}\right) \sum_{j=1}^{k_n-1} g_j^n \int_{t_{i+j-1}}^{t_{i+j}} (1-s)^{-\beta} dW_s
$$

$$
= \frac{1}{h_n} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} (1-s)^{-\beta} dW_s \sum_{j=1}^{k_n-1 \wedge i-1} g_j^n K\left(\frac{t_{i-j-1}-1}{h_n}\right),
$$

whose variance is

$$
\frac{1}{h_n^2} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} (1-s)^{-2\beta} ds \left(\sum_{j=1}^{k_n-1 \wedge i-1} g_j^n K\left(\frac{t_{i-j-1}-1}{h_n}\right) \right)^2
$$

$$
= \frac{1}{h_n^2} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} (1-s)^{-2\beta} ds \, K^2\left(\frac{t_{i-1}-1}{h_n} + O\left(\frac{\Delta_n k_n}{h_n}\right)\right) \Delta_{i,n} \left(\sum_{j=1}^{k_n-1 \wedge i-1} g_j^n \right)^2 = O_p\left(k_n^2 h_n^{-2\beta-1}\right),
$$

so

$$
\frac{1}{h_n} \sum_{i=1}^{n-k_n+2} K\left(\frac{t_{i-1}-1}{h_n}\right) \Delta_{i-1}^n \bar{V} = O_p\left(k_n h_n^{-\beta-1/2}\right),
$$

which is negligible with respect to the previous term, since $\alpha - \beta > 1/2$. Thus, $\hat{\bar{\mu}}_t^n$ diverges as $k_n h_n^{-\alpha}$ when $\sqrt{k_n} h_n^{1-\alpha} \to \infty$. The leading orders in the denominator are:

$$
\frac{1}{h_n} \sum_{i=1}^{n-k_n+2} K^2\left(\frac{t_{i-1}-t}{h_n}\right) \left(\Delta_{i-1}^n \bar{D}\right)^2 + \frac{2}{h_n} \sum_{L=1}^{L_n} w\left(\frac{L}{L_n}\right) \sum_{i=1}^{n-k_n-L+2} K\left(\frac{t_{i-1}-t}{h_n}\right) K\left(\frac{t_{i+L-1}-t}{h_n}\right) \Delta_{i-1}^n \bar{D} \Delta_{i+L-1}^n \bar{D}
$$

$$
\overset{p}{\sim} \frac{1}{h_n} \sum_{i=1}^{n-k_n+2} K^2\left(\frac{t_{i-1}-t}{h_n}\right) \left(\sum_{j=1}^{k_n-1} g_j^n \frac{(1-t_{i-1}+O(k_n \Delta_n))^{-\alpha}}{1-\alpha} \right)^2 = O_p\left(k_n^2 \frac{\Delta_n^{2(1-\alpha)}}{h_n}\right),
$$

and

$$
\frac{1}{h_n} \sum_{i=1}^{n-k_n+2} K^2\left(\frac{t_{i-1}-t}{h_n}\right) \left(\Delta_{i-1}^n \bar{V}\right)^2 + \frac{2}{h_n} \sum_{L=1}^{L_n} w\left(\frac{L}{L_n}\right) \sum_{i=1}^{n-k_n-L+2} K\left(\frac{t_{i-1}-t}{h_n}\right) K\left(\frac{t_{i+L-1}-t}{h_n}\right) \Delta_{i-1}^n \bar{V} \Delta_{i+L-1}^n \bar{V}
$$

$$
= O_p\left(k_n^2 h_n^{-2\beta}\right).
$$

Rationalizing as in the proof of Theorem 2, we arrive at the claim of this theorem. ∎

Appendix B. Critical value of drift burst t-statistic

In the paper, we show that $(T_{t_i^*}^n)_{i=1}^m$ is asymptotically a sequence of independent standard normal random variables, if m does not increase too fast. A standard extreme value theory can then be applied. In practice, however, the route we follow with frequent sampling of our t-statistic leads $(T_{t_i^*}^n)_{i=1}^m$ to be constructed from overlapping data. The extent of this depends on the interplay between the sampling frequency n, the grid points $(t_i^*)_{i=1}^m$, the kernel K, and the bandwidth h_n. In our implementation $(T_{t_i^*}^n)_{i=1}^m$ exhibits a strong serial correlation, so that m severely overstates the effective number of "independent" copies in a given sample. This implies our test is too conservative when evaluated against the Gumbel distribution.

With the left-sided exponential kernel advocated in the paper, the autocorrelation function (ACF) of $(T_{t_i^*}^n)_{i=1}^m$ turns out to decay close to that of a covariance-stationary AR(1) process (see Panel A in Fig. 8):

$$
Z_i = \rho Z_{i-1} + \epsilon_i, \quad i = 1, \ldots, m, \tag{46}
$$

where $|\rho| < 1$ and $\epsilon_i \overset{\text{i.i.d.}}{\sim} N\left(0, 1-\rho^2\right)$. In this model, $Z_i \sim N(0, 1)$ as consistent with the limit distribution of T_t^n, while the ACF is $\text{cor}(Z_i, Z_j) = \rho^{|i-j|}$.

To account for dependence in $(T_{t_i^*}^n)_{i=1}^m$ and get better size and power properties of our test, we simulate the above AR(1) model. We input a value of ρ that is found by conditional maximum likelihood estimation of Eq. (46) from each individual series of t-statistics (i.e. OLS of Eq. (46) based on $(T_{t_i^*}^n)_{i=1}^m$). We then generate a total of 100,000,000 Monte Carlo replica of the resulting process with a burn-in time of 10,000 observations that are discarded. In each simulation,
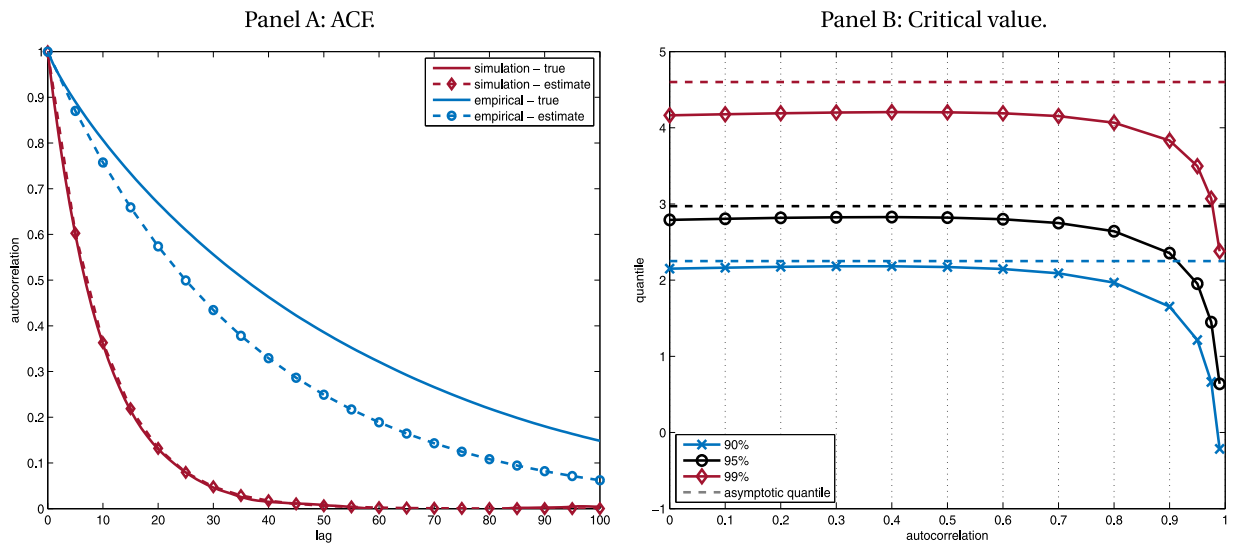
Panel A: ACF.   Panel B: Critical value.

Fig. 8. Autocorrelation function and critical value of *t*-statistic. *Note.* In Panel A, we plot the ACF of $T_{t_i^*}^n$ from the simulation section (averaged across Monte Carlo replica) and the empirical application (averaged across asset markets and over time). The associated dashed curve is that implied by maximum likelihood estimation of the AR(1) approximation in Eq. (46) based on the whole sequence $(T_{t_i^*}^n)_{i=1}^m$ in our sample. The *t*-statistic is constructed as advocated in the main text. In Panel B, we plot the finite sample quantile found via simulation of the AR(1) model in Eq. (46) by inputting the estimated autoregressive coefficient. This figure is based on $m = 2500$ and shows the effect of varying the autocorrelation coefficient ρ and confidence level $1 - \alpha$.

we record the extreme value $Z_m^* = \max_{i=1,\ldots,m} |Z_i|$. We tabulate the quantile function of the raw and normalized Z_m^* series from Eqs. (19)–(20) across the entire universe of simulations and use this table to draw inference.[29]

In Fig. 8, we provide an illustration of this approach. In Panel A, we show the ACF of our *t*-statistic for the stochastic volatility model considered in Section 4 and the empirical high-frequency data analyzed in Section 5. We also plot the curve fitted using the above AR(1) approximation. The estimated ACF is close to the observed one, although there is a slight attenuation bias for the empirical estimates. In Panel B, we report the simulated critical value, as a function of ρ and α with m fixed. These are compared to the ones from the Gumbel distribution. We note a pronounced gap between the finite sample and associated asymptotic quantile, which starts to grow noticeably wider in the region, where ρ exceeds about 0.7–0.8. Apart from that, the extreme value theory offers a decent description of the finite sample distribution for low confidence levels, if the degree of autocorrelation is small, while it gets materially worse, as we go farther into the tails. The latter is explained in part by the fact that even if the underlying sample is uncorrelated, and hence independent in our setting, convergence in law of the maximum term to the Gumbel is known to be exceedingly slow for Gaussian processes (e.g. Hall, 1979).

Appendix C. A parametric test for drift bursts

In this section, as a robustness check, we propose an alternative drift burst test, which is based on a local parametric model. We assume that, in the window $[0, T]$, the log-price X follows the dynamics:

$$\mathrm{d}X_t = \mu(T - t)^{-\alpha}\mathrm{d}t + \sigma(T - t)^{-\beta}\mathrm{d}W_t, \quad t \in [0, T], \tag{47}$$

where $\alpha \in [0, 1)$, $\beta \in [0, 1/2)$, $\mu \in \mathbb{R}$ and $\sigma > 0$ are constant.

The discretely sampled log-return is distributed as $\Delta_i^n X \overset{\mathrm{d}}{\sim} N(\mu_{i,\mathrm{db}}, \sigma_{i,\mathrm{db}}^2)$ with:

$$\mu_{i,\mathrm{db}} = \int_{t_{i-1}}^{t_i} \mu(T - s)^{-\alpha}\mathrm{d}s = \frac{\mu}{1 - \alpha}\Big[(T - t_{i-1})^{1-\alpha} - (T - t_i)^{1-\alpha}\Big], \tag{48}$$

and:

$$\sigma_{i,\mathrm{db}}^2 = \int_{t_{i-1}}^{t_i} \sigma^2(T - s)^{-2\beta}\mathrm{d}s = \frac{\sigma^2}{1 - 2\beta}\Big[(T - t_{i-1})^{1-2\beta} - (T - t_i)^{1-2\beta}\Big]. \tag{49}$$

---

[29] To speed this up for practical work, we prepared in advance a file with the quantile function of the raw and normalized Z_m^* based on the above setup for several choices of m, ρ and selected levels of significance α. This file, along with an interpolation routine to find critical values for other m and ρ, can be retrieved from the authors at request.

Panel A: Average log-return by direction.

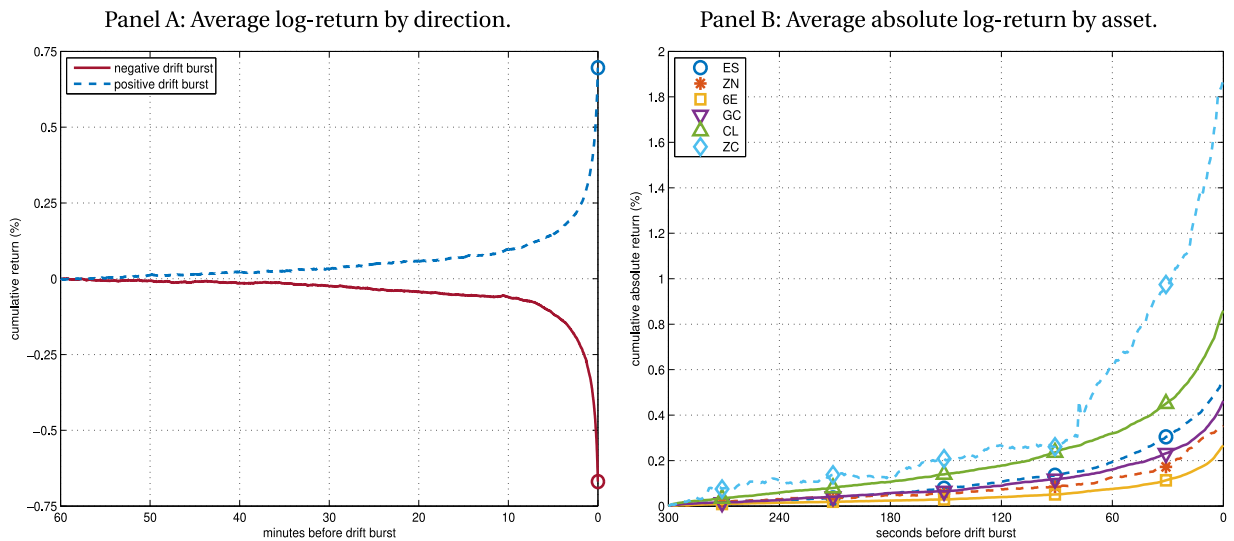Panel B: Average absolute log-return by asset.

Fig. 9. Average drift burst log-return. *Note.* In Panel A, we show the cumulative log-return for positive and negative drift bursts, averaged across assets, in a one-hour window prior to the event. In Panel B, we extend this analysis by plotting the average absolute cumulative log-return by asset in a five-minute window prior to the event.

The log-likelihood for the model is:

$$\log \mathscr{L}(\Theta; X) = \sum_{j=1}^{n} \log \phi(\mu_{j,\mathrm{db}}, \sigma_{j,\mathrm{db}}^2; \Delta_j^n X) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{j=1}^{n} \log(\sigma_{j,\mathrm{db}}^2) - \frac{1}{2} \sum_{j=1}^{n} \frac{(\Delta_j^n X - \mu_{j,\mathrm{db}})^2}{\sigma_{j,\mathrm{db}}^2}, \tag{50}$$

where $\phi(\mu_{j,\mathrm{db}}, \sigma_{j,\mathrm{db}}^2; \Delta_j^n X)$ is the Gaussian density function.

The parameter vector $\Theta = (\mu, \sigma, \alpha, \beta)$ is estimated by maximum likelihood:

$$\hat{\Theta}_{\mathrm{ML}} = \arg \max_{\Theta} \log \mathscr{L}(\Theta; X). \tag{51}$$

We test the drift burst hypothesis:

$$\mathscr{H}_0 : \alpha = 0 \quad \text{against} \quad \mathscr{H}_1 : \alpha > 0 \tag{52}$$

with a likelihood ratio statistic $\Lambda = -2(\log \mathscr{L}(\hat{\Theta}_{\mathrm{ML}}; X) - \log \mathscr{L}(\Theta_{\mathscr{H}_0}; X)) \overset{a}{\sim} \chi_1^2$ (a volatility burst is allowed, since β is unrestricted). In the Online Appendix, based on simulations from the parametric model in Eq. (47), the test is found to perform reasonably in finite samples.

We implement the parametric test on two samples. The first, detected by the nonparametric test with a maximum t-statistic of at least five (in absolute value) and labeled the "drift burst sample", consists of second-by-second prices sampled in a one-hour run-up window before the drift burst. The sample has 933 events. In Fig. 9, we illustrate a typical drift burst in these data. In Panel A, we plot the average cumulative log-return of negative and positive drift bursts in the event window. In Panel B, we show the average absolute cumulative log-return for different asset classes in the last five minutes. The second "no signal" control sample has the corresponding data (when it is available) a week after each drift burst, which is 855 events.

In Panel A of Fig. 10, we report the distribution of the likelihood ratio statistic in these two samples. As seen, Λ is close to chi-square distributed in the no signal sample, while it is strongly skewed to the right in the drift burst sample.[30] In Panel B, we show a histogram of $\hat{\alpha}_{\mathrm{ML}}$ and $\hat{\beta}_{\mathrm{ML}}$ based on the 597 events in the drift burst sample, where the likelihood ratio test statistic is significant at a 5% level. The sample averages are $\bar{\hat{\alpha}}_{\mathrm{ML}} = 0.6250$ and $\bar{\hat{\beta}}_{\mathrm{ML}} = 0.1401$. $\hat{\beta}_{\mathrm{ML}}$ is typically much smaller than $\hat{\alpha}_{\mathrm{ML}}$ (otherwise it is hard for the nonparametric test to detect an event). This aligns with the theory in Section 2. The average difference $\bar{\hat{\alpha}}_{\mathrm{ML}} - \bar{\hat{\beta}}_{\mathrm{ML}} = 0.4849$ is borderline with the no-arbitrage boundary of 0.5 (the full histogram is in Panel C). A high correlation of 51.94% between $\hat{\alpha}_{\mathrm{ML}}$ and $\hat{\beta}_{\mathrm{ML}}$ is noticed, as also evident by the scatter plot in Panel D. It indicates that a stronger drift burst is accompanied by a stronger volatility burst, as consistent with the theory.

---

[30] As a robustness check, we applied three implementation windows: one hour, thirty minutes and ten minutes. In the no signal sample, and at a 5% significance level, the likelihood ratio statistic rejects $\mathscr{H}_0$ in 1.87%, 1.87% and 1.17% of the events, while the comparable numbers in the drift burst sample are 63.99%, 54.88% and 42.66%. Thus, drift bursts identified by the nonparametric statistic are also more likely to be identified by the parametric test as more data are included. This is consistent with the simulation analysis in Section 4.
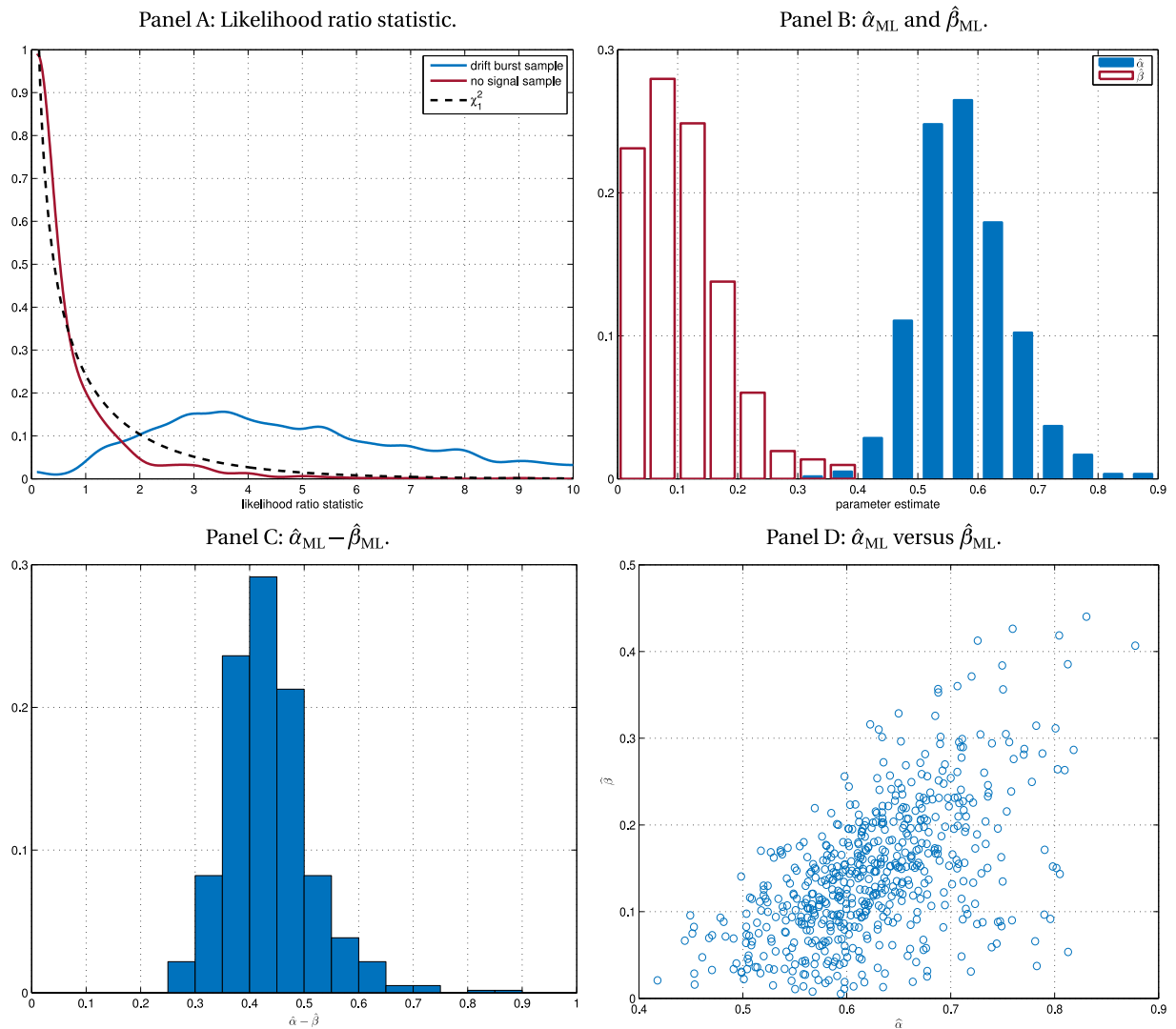
Panel A: Likelihood ratio statistic.

Panel B: $\hat{\alpha}_{\mathrm{ML}}$ and $\hat{\beta}_{\mathrm{ML}}$.

Panel C: $\hat{\alpha}_{\mathrm{ML}} - \hat{\beta}_{\mathrm{ML}}$.

Panel D: $\hat{\alpha}_{\mathrm{ML}}$ versus $\hat{\beta}_{\mathrm{ML}}$.

Fig. 10. Empirical analysis of the parametric drift burst testing procedure. *Note.* The figure reports on the outcome of the empirical analysis based on the parametric test for drift bursts. In Panel A, we show a kernel estimate of the distribution of the likelihood ratio statistic in the drift burst and no signal sample, which are compared to the χ_1^2 distribution. In Panel B–C, we plot a histogram of the maximum likelihood estimates $\hat{\alpha}_{\mathrm{ML}}$, $\hat{\beta}_{\mathrm{ML}}$, and $\hat{\alpha}_{\mathrm{ML}} - \hat{\beta}_{\mathrm{ML}}$. In Panel D, a scatter plot of $\hat{\alpha}_{\mathrm{ML}}$ against $\hat{\beta}_{\mathrm{ML}}$ is shown.

The parametric model can also be used to test a volatility burst:

$$\mathscr{H}_0' : \beta = 0 \quad \text{against} \quad \mathscr{H}_1' : \beta > 0 \tag{53}$$

with $\Lambda' = -2(\log \mathscr{L}(\hat{\Theta}_{\mathrm{ML}}; X) - \log \mathscr{L}(\Theta_{\mathscr{H}_0'}; X)) \overset{a}{\sim} \chi_1^2$. If we apply this to the drift burst sample, $\mathscr{H}_0'$ is discarded at the 5% level for 93.68% of the events, showing that volatility is almost always co-exploding during a drift burst.

Appendix D. Supporting information

The Online Appendix of this article is available from the corresponding author at request.

References

Aït-Sahalia, Y., Jacod, J., Li, J., 2012. Testing for jumps in noisy high frequency data. J. Econometrics 168 (2), 207–222.

Aït-Sahalia, Y., Kimmel, R., 2007. Maximum likelihood estimation of stochastic volatility models. J. Financ. Econom. 83 (2), 413–452.

Aït-Sahalia, Y., Xiu, D., 2015. Increased correlation among asset classes: Are volatility or jumps to blame, or both? J. Econometrics 194 (2), 205–219.

Andersen, T.G., Bollerslev, T., Dobrev, D., 2007. No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications. J. Econometrics 138 (1), 125–180.

Andersen, T.G., Bondarenko, O., Kyle, A.S., Obizhaeva, A.A., 2015. Intraday Trading Invariance in the E-Mini S&P 500 Futures Market. Working paper, Northwestern University.

Andrews, D.W.K., 1991. Heteroscedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59 (3), 817–858.

Baeumer, B., Meerschaert, M.M., 2010. Tempered stable Lévy motion and transient super-diffusion. J. Comput. Appl. Math. 233 (10), 2438–2448.

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

Bandi, F.M., 2002. Short-term interest rate dynamics: A spatial approach. J. Financ. Econom. 65 (1), 73–110.

Bandi, F.M., Russell, J.R., 2006. Separating microstructure noise from volatility. J. Financ. Econom. 79 (3), 655–692.

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), 1–32.

Bates, D.S., 2018. How crashes develop: Intradaily volatility and crash evolution. J. Finance 74 (1), 193–238.

Björk, T., 2003. Arbitrage Theory in Continuous Time, second ed. Oxford University Press.

Black, F., 1986. Noise. J. Finance 41 (3), 529–543.

Brunnermeier, M.K., Pedersen, L.H., 2005. Predatory trading. J. Finance 60 (4), 1825–1863.

Brunnermeier, M.K., Pedersen, L.H., 2009. Market liquidity and funding liquidity. Rev. Financ. Stud. 22 (6), 2201–2238.

Campbell, J.Y., Grossman, S.J., Wang, J., 1993. Trading volume and serial correlation in stock returns. Q. J. Econ. 108 (4), 905–939.

CFTC and SEC, 2010. Findings regarding the market events of may 6, 2010. report of the staffs of the CFTC and SEC to the joint advisory committee on emerging regulatory issues, available at http://www.sec.gov/news/studies/2010/marketevents-report.pdf.

CFTC and SEC, 2011. Recommendations regarding regulatory responses to the market events of may 6, 2010. report of the Joint CFTC-SEC Advisory Committee on Emerging Regulatory Issues, available at http://www.sec.gov/spotlight/sec-cftcjointcommittee/021811-report.pdf.

Christensen, K., Oomen, R.C.A., Podolskij, M., 2014. Fact or friction: Jumps at ultra high frequency. J. Financ. Econom. 114 (3), 576–599.

Delbaen, F., Schachermayer, W., 1994. A general version of the fundamental theorem of asset pricing. Math. Ann. 300 (1), 463–520.

Dubinsky, A., Johannes, M., Kaeck, A., Seeger, N.J., 2018. Option pricing of earnings announcement risks. Rev. Financ. Stud. 32 (2), 646–687.

Dzhaparidze, K., van Zanten, J.H., 2001. On Bernstein-type inequalities for martingales. Stochastic Process. Appl. 93 (1), 109–117.

Easley, D., de Prado, M.M.L., O'Hara, M., 2011. The microstructure of the "flash crash": Flow toxicity, liquidity crashes and the probability of informed trading. J. Portf. Manag. 37 (2), 118–128.

Golub, A., Keane, J., Poon, S.-H., 2012. High Frequency Trading and Mini Flash Crashes. Working paper, University of Mancester.

Grossman, S.J., Miller, M.H., 1988. Liquidity and market structure. J. Finance 43 (3), 617–633.

Hall, P., 1979. On the rate of convergence of normal extremes. J. Appl. Probab. 16 (2), 433–439.

Hassler, U., 2016. Stochastic Processes and Calculus, first ed. Springer-Verlag, Berlin.

Hengartner, N.W., Linton, O.B., 1996. Nonparametric regression estimation at design poles and zeros. Canad. J. Statist. 24 (4), 583–591.

Heston, S.L., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev. Financ. Stud. 6 (2), 327–343.

Huang, J., Wang, J., 2009. Liquidity and market crashes. Rev. Financ. Stud. 22 (7), 2607–2643.

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., Protter, P.E., 2012. Discretization of Processes, second ed. Springer-Verlag, Berlin.

Johnson, T.C., 2016. Rethinking reversals. J. Financ. Econom. 120 (2), 211–228.

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

Karatzas, I., Shreve, S.E., 1998. Methods of Mathematical Finance, first ed. Springer-Verlag, Berlin.

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.

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

Lee, S.S., Mykland, P.A., 2008. Jumps in financial markets: A new nonparametric test and jump dynamics. Rev. Financ. Stud. 21 (6), 2535–2563.

Li, Z.M., Linton, O.B., 2020. A ReMeDi for Microstructure Noise. Working paper, University of Cambridge.

Li, J., Todorov, V., Tauchen, G., 2015. Robust jump regressions. J. Amer. Statist. Assoc. 112 (517), 332–341.

Madhavan, A.N., 2012. Exchange-traded funds, market structure and the flash crash. Financ. Anal. J. 68 (4), 20–35.

Mancini, C., Mattiussi, V., Renò, R., 2015. Spot volatility estimation using delta sequences. Finance Stoch. 19 (2), 261–293.

Menkveld, A.J., Yueshen, B.Z., 2018. The flash crash: A cautionary tale about highly fragmented markets. Manage. Sci. 65 (10), 4470–4488.

Merton, R.C., 1980. On estimating the expected return on the market: An exploratory investigation. J. Financ. Econom. 8 (4), 323–361.

Nagel, S., 2012. Evaporating liquidity. Rev. Financ. Stud. 25 (7), 2005–2039.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–708.

Newey, W.K., West, K.D., 1994. Automatic lag selection in covariance matrix estimation. Rev. Econom. Stud. 61 (4), 631–653.

Oomen, R.C.A., 2006. Comment on 2005 JBES invited address "Realized variance and market microstructure noise" by Peter R. Hansen and Asger Lunde. J. Bus. Econom. Statist. 24 (2), 195–202.

Phillips, P.C.B., Yu, J., 2011. Dating the timeline of financial bubbles during the subprime crisis. Quant. Econ. 2 (3), 455–491.

Stoll, H.R., 2000. Friction. J. Finance 55 (4), 1479–1514.

Todorov, V., Tauchen, G., 2014. Limit theorems for the empirical distribution function of scaled increments of Itō semimartingales at high frequencies. Ann. Appl. Probab. 24 (5), 1850–1888.

Todorov, V., Tauchen, G., Grynkiv, I., 2014. Volatility activity: Specification and estimation. J. Econometrics 178 (1), 180–193.

U.S. Treasury, FRB, NYFED, SEC, and CFTC, 2015. The U.S. treasury market on october 15, 2014. joint staff report, available at https://www.treasury.gov/press-center/press-releases/Documents/Joint_Staff_Report_Treasury_10-15-2015.pdf.