The drift burst hypothesis

Kim Christensen Roel Oomen Roberto Renò*

August 2018

Abstract

The drift burst hypothesis postulates the existence of short-lived locally explosive trends in the price paths of financial assets. The recent US 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, discuss 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 to 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.

JEL Classification: G10; C58.

Keywords: flash crashes; gradual jumps; volatility bursts; liquidity; nonparametric statistics; microstructure noise

*Christensen: Department of Economics and Business Economics, CREATES, Aarhus University, kim@econ.au.dk. Oomen: Deutsche Bank, London and Department of Statistics, London School of Economics. Renò: Department of Economics, University of Verona, roberto.reno@univr.it. We thank Frederich Hubalek, Aleksey Kolokolov, Nour Meddahi, Per Mykland, Thorsten Rheinlander, and participants at the 9th Annual SoFiE Conference in Hong Kong, XVII–XVIII Workshop in Quantitative Finance in Pisa and Milan, 3rd Empirical Finance Workshop at ESSEC, Paris, 10th CFE Conference in Seville, annual conference on Market Microstructure and High Frequency Data at the Stevanovich Center, U. of Chicago, and at seminars in DCU Dublin, Rady School of Management (UCSD), SAFE Frankfurt, Toulouse, TU Wien, Unicredit, U. of Venice and CREATES for helpful comments and suggestions. Christensen received funding from the Danish Council for Independent Research (DFF – 4182-00050) and was supported by CREATES, which is funded by the Danish National Research Foundation (DNRF78). We thank the CME Group and Deutsche Bank AG for access to the data. The views and opinions rendered in this paper reflect the authors' personal views about the subject and do not necessarily represent the views of Deutsche Bank AG, any part thereof, or any other organisation. This article is necessarily general and is not intended to be comprehensive, nor does it constitute legal or financial advice in relation to any particular situation. MATLAB code to compute the proposed drift burst t-statistic is available at request.

Electronic copy available at: https://ssrn.com/abstract=2842335

1 Introduction

The orderly functioning of financial markets will be viewed by most regulators as their first and foremost objective. It is therefore unsurprising that the recent flash crashes in the US 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; US Treasury, FRB, NY FED, SEC, and CFTC, 2015, and Figure 1 for an illustration). Moreover, 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. Firstly, 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 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, are such events reconcilable within the theory of price formation in the presence of market frictions? This paper addresses all these challenges.

The key feature that distinguishes our approach from the existing literature is that we concentrate on the drift term μ_t in the continuous-time Itô semimartingale decomposition for the log-price X_t of a financial asset:

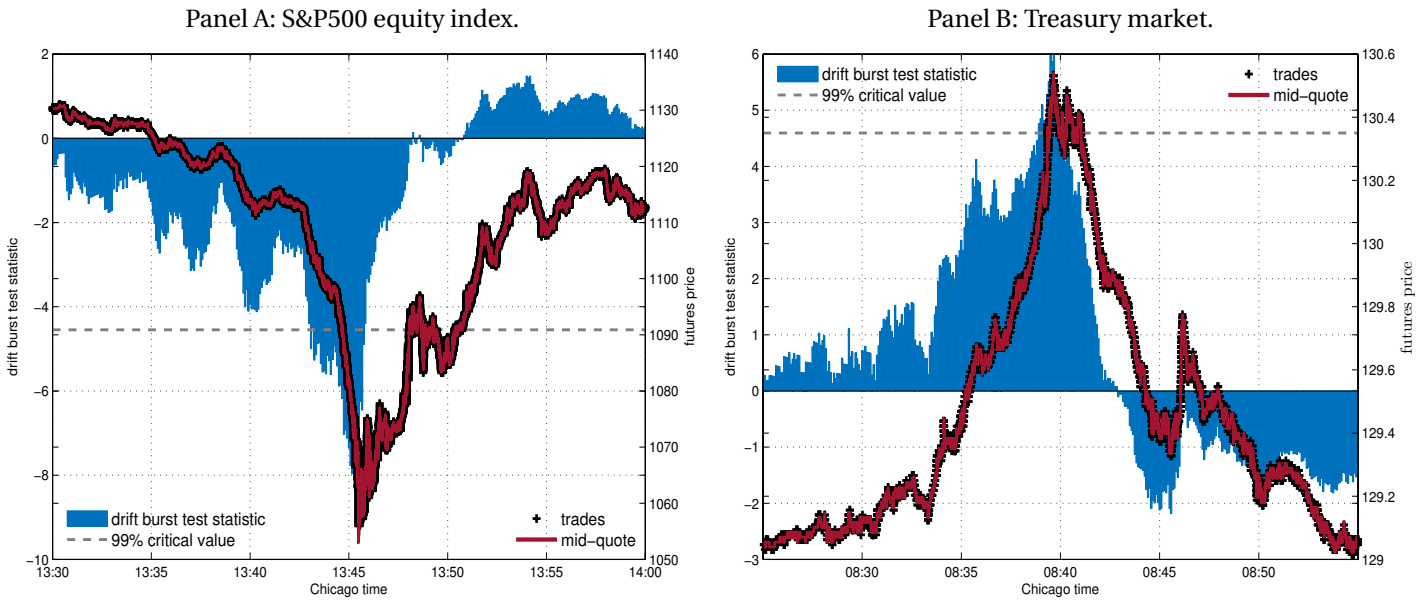
$$\mathrm{d}X_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_t + \mathrm{d}J_t, \tag{1}$$

where 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 $\overline{\Delta} \to 0$, the drift is $O_p(\overline{\Delta})$ (as is the jump term) and swamped by a diffusive component of larger order $O_p(\sqrt{\overline{\Delta}})$. For this reason, 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 subject to considerable measurement error (e.g., Merton, 1980). Hence, the common recommendation is to simply ignore it. Yet, to explain such events as those in Figure 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

[1] In a Financial Times article, Tett (2015) reports on a speech by the CFTC chairman (Massad, 2015) and writes: "Flash crashes affect even commodities markets hitherto considered dull such as corn." Shortly after the first high profile US equity flash crash, a New York Times article by Kaufman and Levin (2011) calls for regulatory action in anticipation of further events. Subsequently, Nanex Research has been reporting hundreds of flash crashes across all major financial markets, see `http://www.nanex.net/NxResearch/`. In a Liberty Street Economics blog of the New York Federal Reserve Bank, Schaumburg and Yang (2015) examine liquidity during flash crashes (see also Golub, Keane, and Poon, 2012, for related work).

Electronic copy available at: https://ssrn.com/abstract=2842335

Figure 1: The US 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.

dynamics. We show how drift bursts can be embedded in the traditional continuous-time model in Eq. (1). Next, we develop a feasible nonparametric identification strategy that enables the on-line detection of drift bursts from high-frequency data. A comprehensive empirical analysis of representative securities from the equity, fixed income, currency, and commodity markets demonstrates that drift bursts are a stylized fact of the price process.

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

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

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 come to 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 impaired (i.e. increased risk aversion γ or fewer active makers M). This prediction fits the 2010 equity flash crash episode in that the extreme price drop appeared to be accompanied by increased risk aversion and a

Electronic copy available at: https://ssrn.com/abstract=2842335

rapid decline in the number of participating market makers: CFTC and SEC (2010) writes "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 Figure 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, Grossman, and Wang (1993) also 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).[2] 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 easily translate into an econometric identification strategy of the impacted sample paths in continuous-time. We deliver a foundation that increases the depth of the empirical work that can be conducted in this area. As an example, we show empirical support for endogenous trading imbalance generated by costly market presence, as postulated in Huang and Wang (2009).

With the drift burst hypothesis in place and the corresponding Itô semimartingale price process specified, we develop an effective identification strategy for the on-line detection of drift burst sample paths from intraday noisy high-frequency data. The method is nonparametric and can be viewed as a type of t-test that aims to establish whether the observed price movement is more likely generated by the drift than be the result of diffusive volatility. Unsurprisingly, the test requires estimation of the local drift and volatility coefficients which is non-trivial for at least two reasons. Firstly, 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 the work by Bandi (2002); Kristensen (2010) for coefficient estimation and Newey and West (1987); Andrews (1991); Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008); Jacod, Li, Mykland, Podolskij, and Vetter (2009); Podolskij and Vetter (2009a,b) 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

[2]There are also a number of practical mechanisms that can amplify, if not cause, violent price drops and surges, including margin calls on leveraged positions (i.e. forced liquidation), dynamic hedging of short-gamma positions, stop-loss orders concentrated around specific price levels, or technical momentum trading strategies.

Electronic copy available at: https://ssrn.com/abstract=2842335

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 well behaved and is capable of identifying drift burst episodes. Interestingly, applying the test to high-frequency data for the days of the US equity and treasury market flash crashes, displayed in Figure 1, we find that they constitute highly significant drift bursts.

Our new mathematical framework, that introduces drift bursts via an exploding drift coefficient, provides an essential ingredient, which helps to to reconcile a number of phenomena observed in financial markets. The first is the already mentioned occurrence of flash crashes, where highly directional and sustained price movements are reversed shortly after. While there is a substantial body of research that focuses on the May 2010 equity market flash crash (a partial list includes Easley, de Prado, and O'Hara, 2011; Madhavan, 2012; Andersen, Bondarenko, Kyle, and Obizhaeva, 2015; Kirilenko, Kyle, Samadi, and Tuzun, 2017; Menkveld and Yueshen, 2018), there has been no attempt thus far to move beyond specific case-studies and analyse 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, Hansen, Lunde, and Shephard (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, Oomen, and Podolskij (2014), who find that the total return variation that can be attributed to the jump component is an order of magnitude smaller than had previously been reported by extensive empirical literature on the topic. 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. Christensen, Oomen, and Podolskij (2014) and Bajgrowicz, Scaillet, and Treccani (2016) show that spurious detection of jumps at low frequency can be explained by an erratic volatility process. However, because a volatility burst merely leads to wider price dispersion, it fails to reconcile the often steady and directional price evolution over such episodes. On the basis of the results presented in this paper, 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.

The empirical analysis we undertake in this paper sets out to determine the prevalence of drift bursts in practice and to characterise 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 period. This systematic assessment provides unparalleled insights into the high resolution price dynamics of an area where hitherto any existing analysis was based on specific case studies of high profile events (e.g. Madhavan, 2012; Kirilenko, Kyle, Samadi, and Tuzun, 2017) or the screening of data based on ad hoc identification rules (e.g. Massad, 2015). Our findings demonstrate that drift bursts are an integral part of the price process across all asset classes considered:

Electronic copy available at: https://ssrn.com/abstract=2842335

over the full sample period, we identify over one thousand significant episodes, 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 more extreme moves between 3% and 8%. We find that roughly two thirds of 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 cited above, we find that trading volume during a drift burst is highly correlated with the subsequent price reversal. The post-drift burst return 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 organised as follows. Section 2 introduces the drift burst hypothesis and describes the mathematical 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 power of the test. The empirical application is found in Section 5, while Section 6 concludes.

2 The hypothesis

Let $X = (X_t)_{t\geq 0}$ denote the log-price of a traded security. We assume the following.

Assumption 1 *X is defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathscr{P})$ and assumed to be an Itô semimartingale 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 and strictly positive (almost surely) 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 model, which represents our frictionless null, is a standard formulation for continuous-time arbitrage-free price processes. 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. It encompasses a wide range of specifications and is compatible with time-varying expected returns, stochastic volatility, leverage effects, and infinite-activity jumps in the log-price and in the volatility. Below, we also add pre-announced jumps and additive noise to the model.

To introduce the drift burst hypothesis, we momentarily enforce that X has continuous sample paths, i.e. $\mathrm{d}J_t = 0$. The jump process is fully reactivated in our theoretical results below. We fix a point τ_{db}. As μ and σ are locally

Electronic copy available at: https://ssrn.com/abstract=2842335

bounded under Assumption 1, it follows that, as $\overline{\Delta} \to 0$:

$$\int_{\tau_{\text{db}}-\overline{\Delta}}^{\tau_{\text{db}}+\overline{\Delta}} |\mu_s| \mathrm{d}s = O_p\left(\overline{\Delta}\right) \qquad \text{and} \qquad \int_{\tau_{\text{db}}-\overline{\Delta}}^{\tau_{\text{db}}+\overline{\Delta}} \sigma_s \mathrm{d}W_s = O_p\left(\sqrt{\overline{\Delta}}\right). \tag{3}$$

Thus, as $\overline{\Delta} \to 0$, the drift is much smaller than the volatility, because $\overline{\Delta} \ll \sqrt{\overline{\Delta}}$. 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 econometric literature to largely neglect the drift.

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

$$\int_{\tau_{\text{db}}-\overline{\Delta}}^{\tau_{\text{db}}+\overline{\Delta}} |\mu_s| \mathrm{d}s = O_p\left(\overline{\Delta}^{\gamma_\mu}\right), \tag{4}$$

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 the continuity of X.

A simple example of an exploding drift leading to a drift burst is:

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

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 have opposite signs (see, e.g., Panel A in Figure 2). It could also accommodate gradual jumps without reversion, e.g. when $a_2 = 0$.[3]

The process X in Eq. (1) with a drift as in Eq. (5) is still a semimartingale.[4] This is necessary—but not sufficient—to exclude arbitrage from the model (e.g., Delbaen and Schachermayer, 1994). To prevent arbitrage, a further condition (imposed by Girsanov's Theorem) is necessary and sufficient for the existence of an equivalent martingale

[3] 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 realised quickly and then may be followed by a gradual continuation as the market interprets and fully incorporates the shock.

[4] While explosive drift does not impede the semimartingale structure of X, 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, Todorov, and Tauchen (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 Figure 1) and the hoax tweet of April 23, 2014 (in Figure 6).

Electronic copy available at: https://ssrn.com/abstract=2842335

measure (e.g., Theorem 4.2 in Karatzas and Shreve, 1998):

$$\int_{\tau_{\text{db}}-\overline{\Delta}}^{\tau_{\text{db}}+\overline{\Delta}} \left(\frac{\mu_s}{\sigma_s}\right)^2 \mathrm{d}s < \infty, \tag{6}$$

which is known as a "structural condition." 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, e.g., 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

$$\int_{\tau_{\text{db}}-\overline{\Delta}}^{\tau_{\text{db}}+\overline{\Delta}} \sigma_s \mathrm{d}W_s = O_p\left(\overline{\Delta}^{\gamma_\sigma}\right), \tag{7}$$

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

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

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

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

so that $\gamma_\sigma = 1/2 - \beta$. Consider the "canonical" alternative model:

$$\mathrm{d}X_t = \mu_t^{\text{db}} \mathrm{d}t + \sigma_t^{\text{vb}} \mathrm{d}W_t, \tag{10}$$

for which $\mu_t^{\text{db}}/\sigma_t^{\text{vb}} \to \infty$ as $t \to \tau_{\text{db}}$ if $\alpha > \beta$. The structural condition in Eq. (6) is readily satisfied when $\alpha - \beta < 1/2$. Thus, this specific example shows that the drift coefficient can explode locally (even after normalising by an exploding volatility) either preserving absence of arbitrage (when $\beta < \alpha < \beta + 1/2$), or allowing local arbitrage opportunities (when $\alpha > \beta + 1/2$).[5]

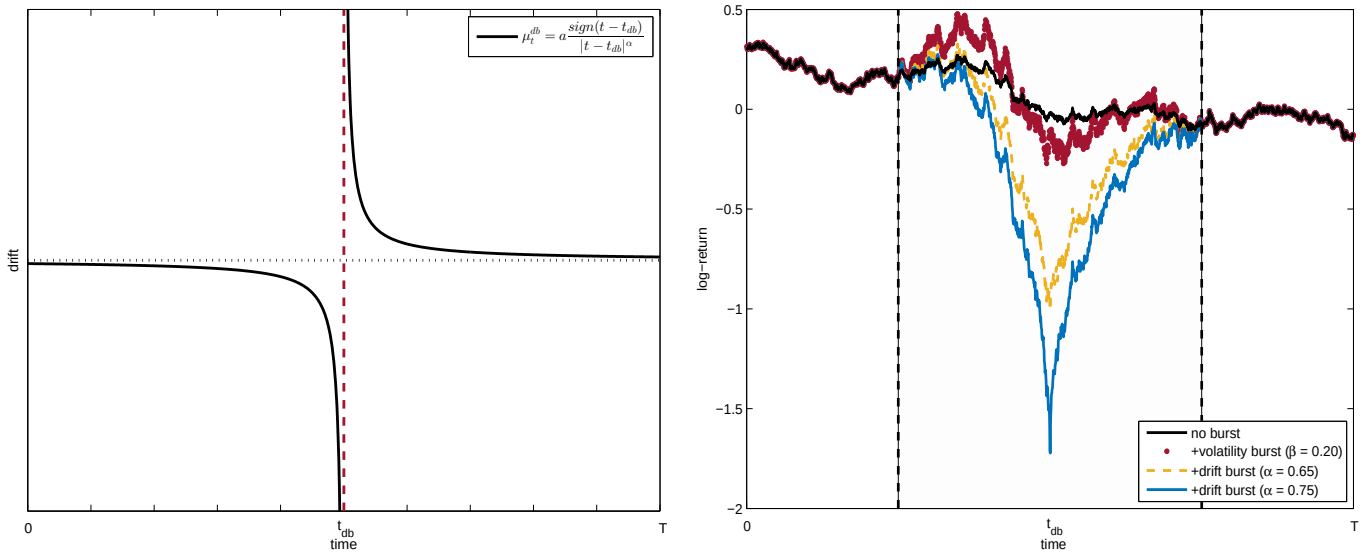
In Figure 2, Panel B, we show simulated sample paths of the associated log-price in this framework. A flash crash is generated from model (10) (see Panel A for the specification of the drift burst component). We set the drift burst rate at $\alpha = 0.65$ and $\alpha = 0.75$ with a volatility burst parameter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

[5] As detailed in Appendix C, we can estimate α and β using a parametric maximum likelihood approach based on model (10). The sample averages across detected events in the empirical high-frequency data analysed in Section 5 are $\bar{\hat{\alpha}}_{\text{ML}} = 0.6250$ and $\bar{\hat{\beta}}_{\text{ML}} = 0.1401$. 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.

Electronic copy available at: https://ssrn.com/abstract=2842335

Figure 2: Illustration of a log-price with a drift burst.

Panel A: drift coefficient. Panel B: simulated log-return.



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 are based on Eq. (5) and (8) with $-a_1 = a_2 = 3$, $b = 0.15$, $\alpha = 0.65$ or 0.75 and $\beta = 0.2$.

is pronounced and qualitatively similar.

The notion that volatility is bursty is uncontroversial, particularly over episodes of market turbulence or dislocation. For instance, Kirilenko, Kyle, Samadi, and Tuzun (2017) and Andersen, Bondarenko, Kyle, and Obizhaeva (2015) report elevated levels of volatility during the equity flash crash (see also Bates, 2018).[6] However, as illustrated by Figure 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 natural and effective.

3 Identification

We now develop a nonparametric approach to detect drift bursts in real data. We propose a t-statistic that exploits the message of Eq. (4), namely if there is a drift burst in X 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 neighbourhood of t. Later, we prove that our "signal-to-noise" measure uncovers drift bursts in X if they are sufficiently strong.

[6] A likelihood ratio test based on model (10), 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.

Electronic copy available at: https://ssrn.com/abstract=2842335

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 $0 = t_0 < t_1 < \ldots < t_n = T$, 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 defined as $\Delta_i^n X = X_{t_i} - X_{t_{i-1}}$. We define:

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

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

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

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 $n \to \infty$:

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

where $\mu_t^* = \mu_t + \int_{\mathbb{R}} \delta(t, x) I_{\{|\delta(t,x)|>1\}} \lambda(\mathrm{d}x)$ and K_2 is a kernel-dependent constant.

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 and 2.

The t-statistic is thus defined as:

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

T_t^n has an intuitive interpretation with the indicator kernel. In that case, it is the ratio of the drift part to the volatility part of the log-return over the interval $[t - h_n, t]$. As $h_n \to 0$, this is valid with any kernel satisfying the stated assumptions.

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

Electronic copy available at: https://ssrn.com/abstract=2842335

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

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

Proof. See Appendix A. ■

Theorem 1 shows that, in the absence of a drift burst, the t-statistic in Eq. (14) has a limiting standard normal distribution.[8] We note that, under the null, the behavior of T_t^n does not depend on the unknown values of μ_{t-}^* and σ_{t-} for large n. Thus, although it is not possible to consistently 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. This is formalized in the drift burst alternative by an exploding μ_t term. We notice that the alternative is broad enough to allow σ_t to co-explode with the drift.

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

$$\mathrm{d}\tilde{X}_t = \mathrm{d}X_t + \tilde{\mu}_t \mathrm{d}t + \tilde{\sigma}_t \mathrm{d}W_t, \tag{16}$$

where $\mathrm{d}X_t$ is the model in Eq. (1) such that the conditions of Theorem 1 hold. Moreover, we assume there exists a stopping time τ_{db} and a $\theta > 0$ such that, when $t \in (\tau_{\mathrm{db}} - \theta, \tau_{\mathrm{db}})$ and for every $\epsilon > 0$, there exists a $M > 0$ such that:

$$\mathscr{P}\left(|\tilde{\mu}_t| > \frac{M}{(\tau_{\mathrm{db}} - t)^\alpha}\right) > 1 - \epsilon \qquad \textit{and} \qquad \mathscr{P}\left(\tilde{\sigma}_t < \frac{M}{(\tau_{\mathrm{db}} - t)^\beta}\right) > 1 - \epsilon, \tag{17}$$

with $0 < \beta < 1/2 < \alpha < 1$ and $\alpha - \beta > 1/2$. Then, as $n \to \infty$, it holds that

$$|T_{\tau_{\mathrm{db}}}^n| = \begin{cases} O_p\left((n h_n)^{1-\alpha}\right) & \textit{if } \frac{\Delta_n^{2(1-\alpha)}}{h_n^{1-2\beta}} \to \infty, \\ O_p\left(h_n^{1/2-\alpha+\beta}\right) & \textit{if } \frac{\Delta_n^{2(1-\alpha)}}{h_n^{1-2\beta}} \to C, \end{cases} \tag{18}$$

where $C \geq 0$ is a constant.

Proof. See Appendix A. ■

[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).

Electronic copy available at: https://ssrn.com/abstract=2842335

Theorem 2 implies that $|T^n_{\tau_{\text{db}}}| \xrightarrow{p} \infty$ under the alternative. The implication is that if the drift explodes fast enough compared to the volatility (i.e, $\alpha - \beta > 1/2$), T^n_t diverges and has asymptotic power converging to one based on a standard decision rule. The condition $\alpha - \beta > 1/2$ is equivalent to require that, in a neighborhood of τ_{db}, the log-return is dominated by drift. This allows, in a frictionless economy, 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 below confirm that the condition $\alpha - \beta > 1/2$ is not needed to achieve power in small samples.

As a corollary of Theorem 2, it can be shown that the t-statistic is robust against an isolated volatility burst with no associated drift burst. In this case, it holds that $T^n_t \xrightarrow{d} N(0,1)$. This theoretical statement is also corroborated by our simulations. Thus, large values of the test statistic cannot be explained by isolated volatility bursts.

3.1 Inference via the maximum statistic

The asymptotic theory asserts that T^n_t 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 $|T^n_t|$ 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^n_t.

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.[9] We compute $\left(T^n_{t^*_i}\right)_{i=1}^m$ at m equispaced time points $t^*_i \in (0,T]$, where T is fixed. We set:

$$T^*_m = \max_{t^*_i} |T^n_{t^*_i}|, \quad i = 1,\dots,m. \tag{19}$$

The crucial point is that, in addition to $T^n_{t^*_i} \xrightarrow{d} N(0,1)$ under the null, the $T^n_{t^*_i}$'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 *The conditions of Theorem 1 hold. Then, if $n \to \infty$, $m \to \infty$ such that $mh_n \to 0$ and $m\sqrt{\log m}\left(\frac{1}{\sqrt{nh_n}} + n^{-\Gamma/2}\right) \to 0$, it holds that:*

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

[9] Bajgrowicz, Scaillet, and Treccani (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, Bollerslev, and Dobrev (2007) propose a Bonferroni correction. The latter was another viable tool to avoid systematic overrejection of the null hypothesis.

Electronic copy available at: https://ssrn.com/abstract=2842335

where

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

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

Proof. See Appendix A. ■

In practice, the Gumbel distribution returns conservative critical values, because of residual dependence in the t-statistics due to small sample effects, microstructure noise and pre-averaging (introduced in Section 3.3). As explained in Appendix B, data-driven critical values can be determined using a simulation-based procedure.

3.2 Robustness to pre-announced jumps

We here study an extension of the model that – on top of the drift, volatility and jump component in Eq. (1) – has a "pre-announced" jump (see, e.g., Jacod, Li, and Zheng, 2017; Dubinsky, Johannes, Kaeck, and Seeger, 2018), where the jump time is fixed across sample paths. We show that these types of jumps do not distort drift burst detection.

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

$$\mathrm{d}\tilde{X}_t = \mathrm{d}X_t + \mathrm{d}J'_t, \tag{22}$$

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

$$T^n_{\tau_J} \xrightarrow{p} \sqrt{\frac{K(0)}{K_2}} \cdot \mathrm{sign}(J). \tag{23}$$

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 adopted in the empirical application has $\sqrt{K(0)/K_2} = 1$, so that $|T^n_{\tau_J}| \xrightarrow{p} 1$. Thus, our proposed t-statistic is – asymptotically – small under the null (standard normal distributed) and pre-announced jump alternative (around one 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

Electronic copy available at: https://ssrn.com/abstract=2842335

test 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 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 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, such that $\gamma_k = 0$ for $k > Q$, where $Q \geq 0$ is an integer (i.e., Q-dependent noise).*

Assumption 2 is a standard additive-noise model in financial econometrics, which allows for autocorrelation in the noise process. 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 slow down the accumulation of noise by pre-averaging Y_{t_i}, as in Jacod, Li, Mykland, Podolskij, and Vetter (2009); Podolskij and Vetter (2009a,b).

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, $g_j^n = g\left(j/k_n\right)$ 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' and such that $g(0) = g(1) = 0$ and $\int_0^1 g^2(s)\mathrm{d}s < \infty$.

Absent a drift burst, it follows from Vetter (2008) 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 (α close to one) dominate the divergence of the asymptotic variance in

Electronic copy available at: https://ssrn.com/abstract=2842335

the drift estimator.

The drift burst t-statistic in Eq. (14) is then redefined with a noise-robust drift and spot volatility estimator computed from the pre-averaged return series:

$$\overline{T}_t^n = \sqrt{\frac{h_n}{K_2}} \frac{\hat{\overline{\mu}}_t^n}{\sqrt{\hat{\overline{\sigma}}_t^n}}, \tag{27}$$

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 (smooth) 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 to include in $\hat{\overline{\sigma}}_t^n$.

The new spot variance estimator $\hat{\overline{\sigma}}_t^n$ is a heteroscedasticity and autocorrelation consistent (HAC)-type statistic (e.g., Newey and West, 1987; Andrews, 1991). The extra complexity is required here to account for any noise dependence and the serial correlation in $(\Delta_i^n \overline{Y})_{i=0}^{n-k_n+1}$ 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), ϵ is as listed in Assumption 2, and Assumption 1, 3, 4, and 5 hold. For every fixed $t \in (0, T]$, as $n \to \infty$, $k_n \to \infty$, $L_n \to \infty$ such that $k_n h_n \to 0$, $k_n h_n' \to 0$, $\frac{k_n}{n h_n} \to 0$, $\frac{k_n}{n h_n'} \to 0$, and $\frac{L_n}{n h_n'} \to 0$, 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 $\frac{k_n}{n h_n} \to 0$ (and the corresponding ones replacing h_n with h_n') call for moderate pre-averaging, so that the number of pre-averaged terms is not too large. The condition $\frac{L_n}{n h_n'} \to 0$ means the lag length also cannot grow too fast when estimating volatility.

Theorem 6 *Set $Y_{t_i} = \tilde{X}_{t_i} + \epsilon_{t_i}$, where $\tilde{X}$ 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 $k_n h_n \to 0$, $k_n h_n' \to 0$, $\frac{k_n}{n h_n} \to 0$, $\frac{k_n}{n h_n'} \to 0$, and $\frac{L_n}{n h_n'} \to 0$, and $k_n h_n^{2(1-\alpha)} \to \infty$, it*

Electronic copy available at: https://ssrn.com/abstract=2842335

holds that:

$$|\overline{T}^{n}_{\tau_{\text{db}}}| = \begin{cases} O_p\left((nh_n)^{1-\alpha}\right) & \text{if } \frac{\Delta_n^{2(1-\alpha)}}{h_n^{1-2\beta}} \to \infty, \\ O_p\left(h_n^{1/2-\alpha+\beta}\right) & \text{if } \frac{\Delta_n^{2(1-\alpha)}}{h_n^{1-2\beta}} \to C, \end{cases} \tag{30}$$

where $C \geq 0$ is constant.

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 with it under the alternative, amid also an exploding volatility and microstructure noise.

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

$$\begin{aligned} \mathrm{d}X_t &= \sigma_t \mathrm{d}W_t, \\ \mathrm{d}\sigma_t^2 &= \kappa\left(\theta - \sigma_t^2\right)\mathrm{d}t + \xi\sigma_t \mathrm{d}B_t, \quad t \in [0,1], \end{aligned} \tag{31}$$

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 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.

Electronic copy available at: https://ssrn.com/abstract=2842335

We create drift and volatility bursts with the parametric model:

$$\mu_t^{\text{db}} = a\frac{\text{sign}(t-\tau_{\text{db}})}{|\tau_{\text{db}}-t|^{\alpha}}, \qquad \sigma_t^{\text{vb}} = b\frac{\theta}{|\tau_{\text{db}}-t|^{\beta}}, \quad \text{for } t\in[0.475,0.525], \tag{32}$$

with $\tau_{\text{db}} = 0.5$ 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.[10] 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.[11] In particular, fixing $a = 3$ we induce a cumulative return $\int_0^{\tau_{\text{db}}} \mu_t^{\text{db}} \text{d}t$ 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.

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). γ is the noise-to-volatility ratio. We set $\gamma = 0.5$, which amounts to a medium contamination level (e.g., Christensen, Oomen, and Podolskij, 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 $g(x) = \min(x, 1-x)$.[12,13] $\hat{\bar{\mu}}_t^n$ and $\hat{\bar{\sigma}}_t^n$ are constructed from $(\Delta_i^n \overline{Y})_{i=0}^{n-2k_n+1}$ based on Eq. (28) and (29) with a left-sided exponential kernel $K(x) = \exp(-|x|)$, for $x \leq 0$.

A Parzen kernel is selected for w:

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

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

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

[12] 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.

[13] With equidistant data, it follows that if k_n is even and $g(x) = \min(x, 1-x)$ the pre-averaged return in (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 simple averaging of the rescaled noisy log-price series, $(Y_{i/n})_{i=0}^n$, in a neighbourhood of i/n, thus making the use of the term pre-averaging and the associated notation transparent.

Electronic copy available at: https://ssrn.com/abstract=2842335

This choice has some profound advantages in our framework. Firstly, the Parzen kernel ensures that $\hat{\overline{\sigma}}_t^n$ 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); Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009). The slight loss of efficiency brings the distinct merit that $\hat{\overline{\sigma}}_t^n$ 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 often large, and 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 $\hat{\overline{\sigma}}_t^n$. 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, Hansen, Lunde, and Shephard, 2009).[14] 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 $\hat{\overline{\sigma}}_t^n$ to better capture persistence in volatility and estimate the microstructure-induced return variation.

A new value of $\overline{T}_t^n$ is recorded at every 60th transaction update.[15] We extract $T_m^* = \max_{i=1,\ldots,m} |\overline{T}_{t_i^*}^n|$ based on the resulting $m = 341$ tests run in each sample. The simulation-based approach explained in Appendix B is adopted to find a critical value of T_m^*.

Figure 3 reports Q-Q plots of the distribution of $\overline{T}_t^n$ under the null hypothesis of no drift burst. In Panel A, we show the results from the pure Heston (1993)-type SV model. As readily seen, the Gaussian curve is an accurate description of the sampling variation of $\overline{T}_t^n$, although the t-statistic is slightly thin-tailed with $h_n = 120$. This is caused by a modest correlation between the numerator and denominator in the test statistic, due to $\hat{\bar{\mu}}_t^n$ and $\hat{\overline{\sigma}}_t^n$ being computed from overlapping data; an effect that is more pronounced for small bandwidths. In Panel B, the outcome of the process featuring a large volatility burst ($\beta = 0.4$) is plotted.[16] While the volatility burst puts some mass further into the tails of the distribution of $\overline{T}_t^n$ this is hardly noticeable, and the normal continues to be a good approximation also in this setting.

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 first at the columns with $\mu_t^{\text{db}} \equiv 0$, 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 Figure 3. As β increases, T_m^* is mildly inflated yielding a tiny size distortion, but this is benign and only present for the largest β

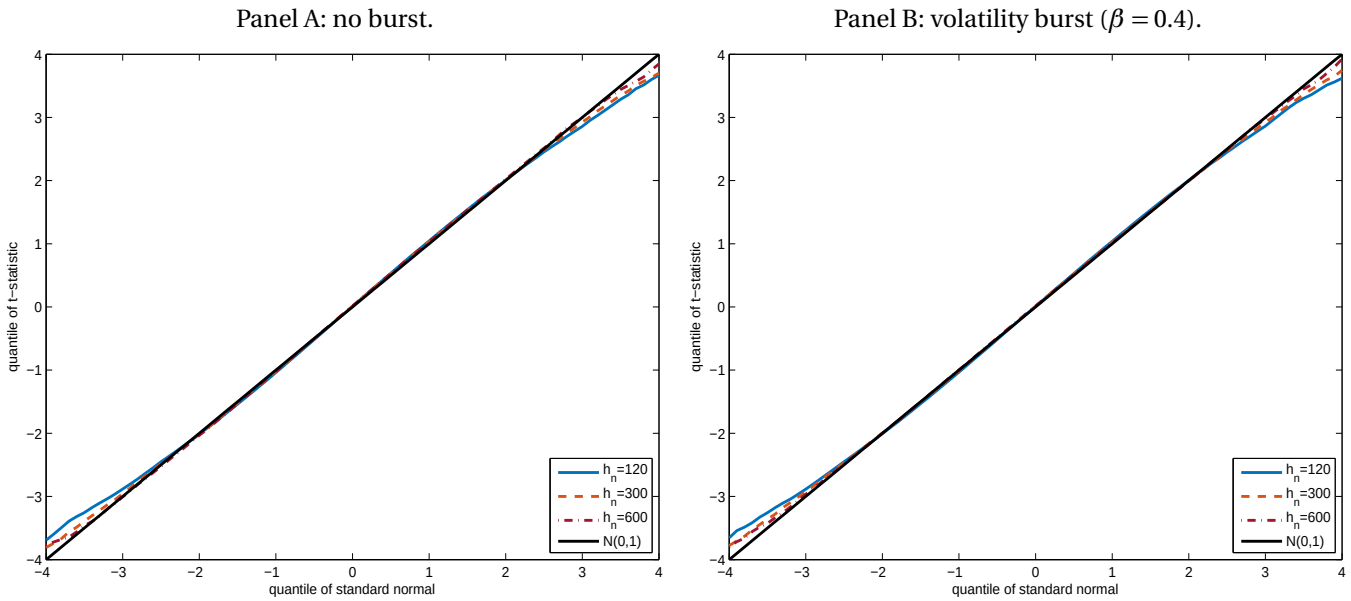
[14] In our simulations, the average value of Q^* is 11.9, while its interquartile range is 8 – 17.

[15] We set a burn-in period of a full volatility bandwidth of trading time to allow for a sufficient number of observations in the construction of $\overline{T}_t^n$.

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

Electronic copy available at: https://ssrn.com/abstract=2842335

Figure 3: Q–Q plot of $\overline{T}_t^n$ without drift burst.



Note. We present a Q-Q plot of $\overline{T}_t^n$ under the null hypothesis of no drift burst. Panel A is from the pure Heston (1993)-type SV model with no burst in neither drift nor volatility, while Panel B adds a volatility burst using the parametric model in Eq. (32) with $b = 0.15$ and $\beta = 0.4$.

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.

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 $h_n = 300$ 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.

Electronic copy available at: https://ssrn.com/abstract=2842335

Table 1: Size and power of drift burst t-statistic T_m^*.

		$\Pr(T_m^* > q_{0.950})$				$\Pr(T_m^* > q_{0.990})$				$\Pr(T_m^* > q_{0.995})$			
		vb (size)	db (power)			vb (size)	db (power)			vb (size)	db (power)		
		$\mu_t^{\text{db}} \equiv 0$	$\alpha =$	0.55	0.65	0.75	$\mu_t^{\text{db}} \equiv 0$	$\alpha =$	0.55	0.65	0.75	$\mu_t^{\text{db}} \equiv 0$	$\alpha =$	0.55	0.65	0.75
Panel A: $h_n = 120$																
$\beta =$	0.0*	1.1		47.4	86.6	98.5	0.1		30.8	68.6	93.1	0.0		26.0	59.9	89.0
	0.1	1.1		31.1	76.6	97.7	0.2		17.1	56.0	90.4	0.1		13.2	47.7	83.7
	0.2	1.2		20.8	66.9	96.3	0.1		9.8	46.9	87.1	0.1		6.3	36.1	79.2
	0.3	1.0		11.3	49.8	94.5	0.0		2.5	26.6	79.0	0.0		1.2	19.8	68.3
	0.4	1.2		5.2	25.5	82.1	0.2		0.6	9.5	56.4	0.0		0.4	5.3	44.5
Panel B: $h_n = 300$																
$\beta =$	0.0*	2.6		58.0	92.4	99.8	0.2		45.5	84.6	98.9	0.2		41.8	80.9	98.2
	0.1	2.6		47.8	88.5	99.5	0.3		32.2	77.8	98.2	0.3		28.5	72.7	97.6
	0.2	2.6		40.6	84.2	99.2	0.2		24.8	71.3	97.7	0.2		20.9	64.9	96.2
	0.3	3.1		28.4	75.0	98.9	0.4		15.7	59.6	95.9	0.2		11.9	53.7	93.1
	0.4	3.6		19.4	55.6	96.3	0.6		7.0	38.3	88.8	0.3		5.1	31.4	83.0
Panel C: $h_n = 600$																
$\beta =$	0.0*	4.7		55.8	89.3	99.5	0.5		43.3	80.1	98.6	0.2		39.1	77.6	97.5
	0.1	4.6		49.1	85.6	99.4	0.7		35.7	77.2	97.8	0.4		30.1	73.3	96.8
	0.2	4.5		44.1	82.6	99.3	0.8		29.7	73.0	97.4	0.2		24.6	68.3	95.8
	0.3	4.8		36.5	77.0	98.8	1.0		21.1	64.3	96.4	0.5		17.1	59.4	94.3
	0.4	5.2		26.5	64.0	97.3	1.4		14.2	48.7	92.0	0.8		10.0	43.4	88.3

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 pure Heston (1993)-type SV model with no volatility burst. h_n is the bandwidth of $\hat{\mu}_t^n$ (measured as effective sample size), while the bandwidth of $\hat{\bar{\sigma}}_t^n$ is $5h_n$. $\gamma = 0.5$ is the level of noise-to-volatility (per increment) and $k_n = 3$ is the pre-averaging horizon.

5 Drift bursts in financial markets

We now apply the drift burst test statistic developed above to a large set of intraday tick data, covering a broad range of financial assets. The aim here is to establish that drift bursts are present empirically and – as such – to illustrate some of their basic properties. Our analysis opens an opportunity to examine whether some of the predictions made in work about liquidity provision, i.e. Huang and Wang (2009), are supported by the data.

5.1 Data

We use a comprehensive set of tick data – 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

Electronic copy available at: https://ssrn.com/abstract=2842335

Table 2: CME futures data summary statistics.

			volume		# quote	inside	sub-sample retained	
code	name	# days	# contracts	notional ($bn)	updates	spread (bps)	by volume	by quotes
6E	Euro FX	1,536	209,281	31.9	62,357	0.69	92.87%	87.25%
CL	Crude oil	1,545	299,821	18.8	69,580	1.70	95.58%	92.28%
ES	E-mini S&P500	2,053	1,763,100	145.5	27,513	1.54	97.33%	91.94%
GC	Gold	1,544	162,056	22.0	56,562	0.97	88.14%	84.59%
ZC	Corn	1,528	112,612	2.6	4,870	5.89	87.23%	71.18%
ZN	10-Year T-Note	1,542	1,121,122	112.1	6,372	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.

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 US cash equity market covering over 5,000 stocks traded across more than ten different exchanges.[17] 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 US 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 $k_n = 3$.[18] 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 $x \leq 0$. In practice, a backward-looking kernel is more powerful for testing the drift burst hypothesis. The intuition

[17] See https://batstrading.com/market_summary/ for daily US equity market volume statistics.

[18] 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.

Electronic copy available at: https://ssrn.com/abstract=2842335

Table 3: Drift burst test summary statistics.

		empirical distribution			# of identified drift bursts				
code	name	σ	σ_q	kurtosis	$\lvert T\rvert > 4.0$	>4.5	>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.

is that after the full development of a drift burst, a high level of volatility (which is both theoretically imposed by absence of arbitrage and empirically forceful, as shown in Appendix C), combined with volatility persistence, delivers a low value of the test statistic after the peak, eroding its power.

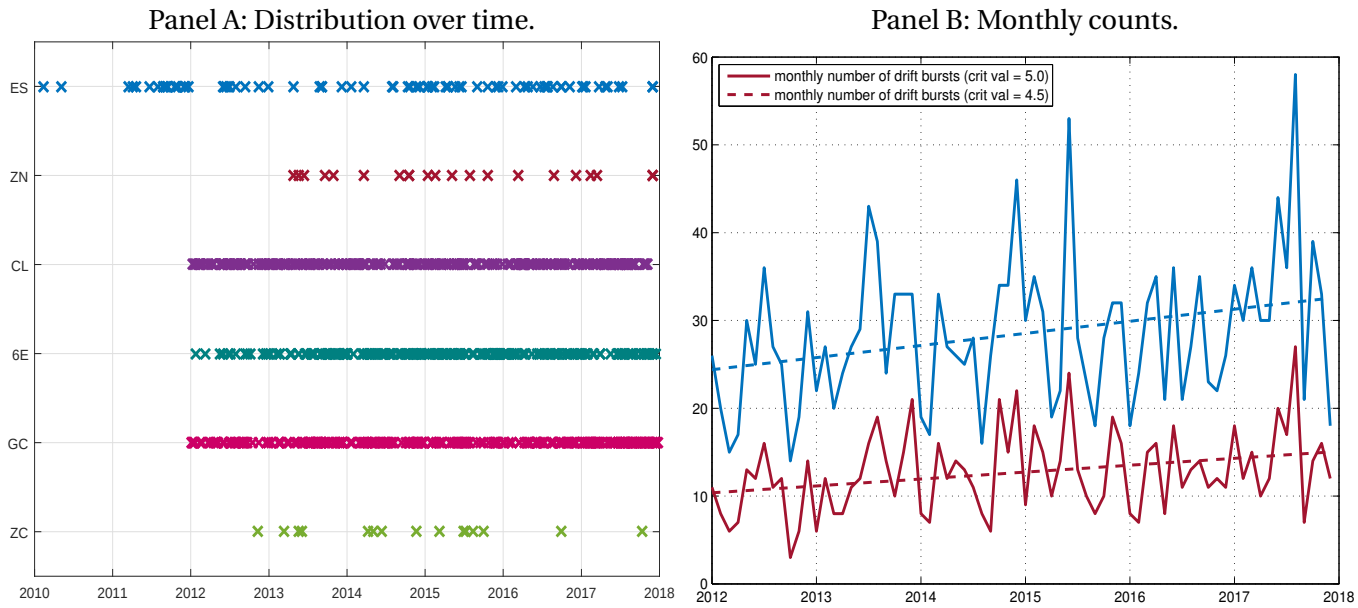
Table 3 reports selected descriptive measures of the calculated drift burst t-statistics over the full sample. Judging by the estimates of standard deviation and kurtosis, the distribution of the test statistic is close to standard normal, as consistent with the asymptotic theory under the null hypothesis 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 in the data.[19] At a critical value of 4.5, for instance, we identify 202 drift bursts in the E-mini S&P500 futures, or about one every two weeks. The number of expected false positives (computed as described in Appendix B) is practically zero. Drift bursts are more prevalent in the Euro FX, Gold, and Oil contracts but less frequent in the Treasury and Corn futures.

Panel A of Figure 4 indicates the location of the identified drift bursts for the various securities, while Panel B reports the pooled monthly event counts. These results lend support to the perception that flash crashes are an increasingly observed phenomenon. The time trend coefficients are positive for each security but lack statistical significance for some due to the limited number of observations. The pooled results, however, indicate a modest and statistically significant time trend of about 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.

[19] 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.

Electronic copy available at: https://ssrn.com/abstract=2842335

Figure 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.

Figure 5 shows some examples of single-asset drift bursts identified by the test. A multi-asset drift burst is presented in Figure 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 instead a plausible alternative to model the data.

5.3 Reversion after the drift burst

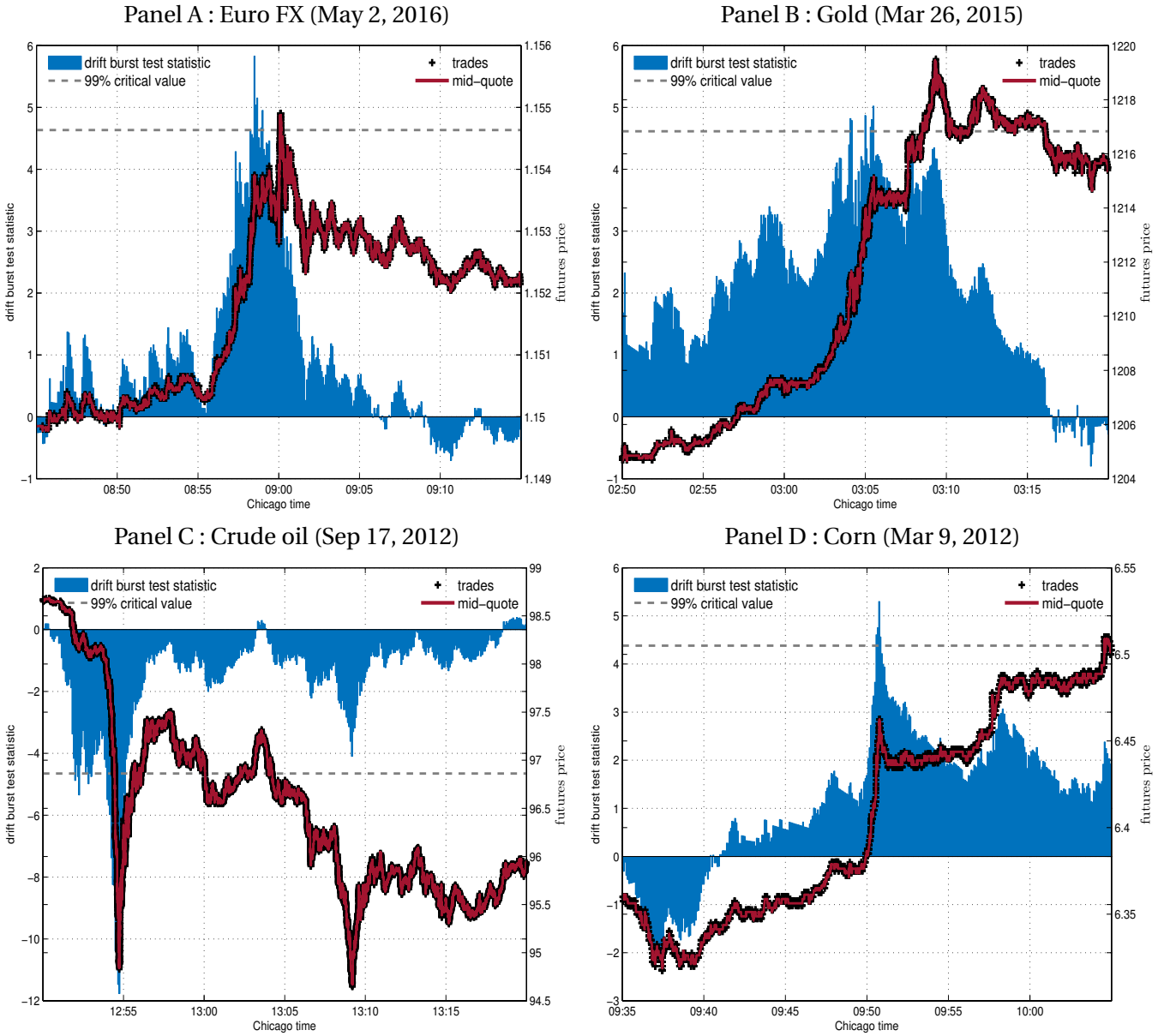
We now investigate in more detail whether the mean reversion experienced during the flash crashes in Figure 1 and Figure 6 is more broadly associated with the price dynamics of a drift burst. Let $\{t_j\}$ denote the set of time points, where drift bursts are identified. The start of a drift burst is set to five minutes before t_j.[20] We sample the mid-quote process at this frequency pre- and post-drift burst and calculate:

$$R^-_{t_j} = X_{t_j} - X_{t_j - 5\text{m}} \qquad \text{and} \qquad R^+_{t_j} = X_{t_j + 5\text{m}} - X_{t_j}, \tag{35}$$

[20]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.

Electronic copy available at: https://ssrn.com/abstract=2842335

Figure 5: Drift burst examples.

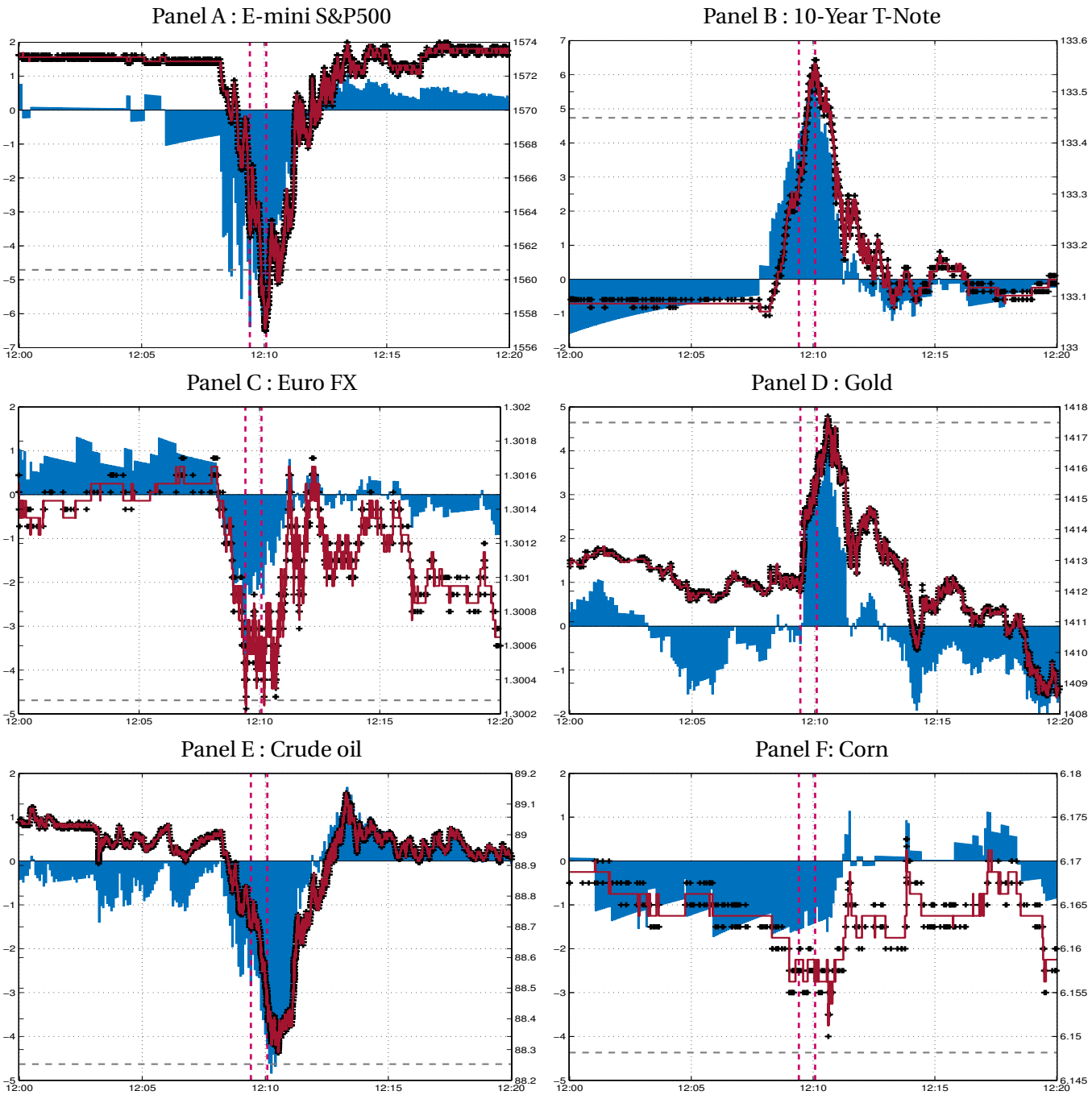


Note. This figure draws, for some identified drift bursts, the sample path of the mid-quote and traded price (right axis) together with the t-statistic (left axis) over a 30-minute window that includes the peak of the drift burst.

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 Figure 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.

Electronic copy available at: https://ssrn.com/abstract=2842335

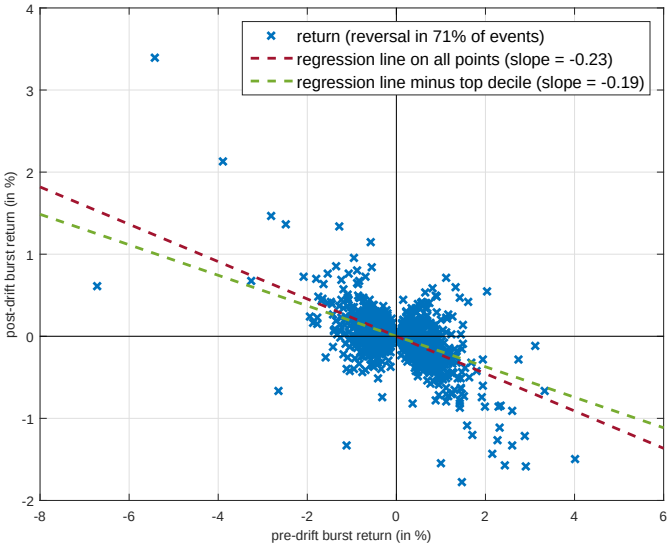
Figure 6: The Twitter hoax of April 23, 2013: A synchronous flash event.



Note. This figure draws the sample path of 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 by the event.

Electronic copy available at: https://ssrn.com/abstract=2842335

Figure 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.

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

$$R^+_{t_j} = a + b R^-_{t_j} + \epsilon_{t_j}. \tag{36}$$

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 indicates there is a substantial predictive power. 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 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.

Electronic copy available at: https://ssrn.com/abstract=2842335

Table 4: Reversion in the drift burst.

	all				without top decile			
sym	#	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), the associated 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.

5.4 Asymmetric reversals, 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 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 can arise endogenously due to costly market participation. Such a mechanism always leads to selling pressure, which tends to attenuate rallies and exacerbate sell-offs, resulting in market crashes absent big news on fundamentals (Result 1 – 3 in their paper). As in Grossman and

Electronic copy available at: https://ssrn.com/abstract=2842335

Miller (1988), the model induces negative serial correlation in observed returns (Result 4a), but it is asymmetric in that negative returns exhibit stronger 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). In this section, we conduct an empirical assessment to check whether our sample of pre- and post-drift burst returns are consistent with these testable implications. To conserve space, 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. Then, for each interval j we compute:

- $R^-_{t_j}$ and $V^-_{t_j}$: the 5-minute log-return and traded volume,[21]
- $R^+_{t_j}$: the subsequent 5-minute log-return,
- ρ: correlation between $R^-_{t_j}$ and $R^+_{t_j}$,
- $\sigma_p = \sqrt{(R^-_{t_j})^2 + (R^+_{t_j})^2}$: volatility measure.

Huang and Wang (2009) report $R^-_{t_j}$, $R^+_{t_j}$, $V^-_{t_j}$, ρ and σ_p based on *simulated* data.[22] 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 here), 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 now 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 move of −37.39bps compared to +36.41bps for a positive one.[23] The subsequent 5-minute log-return is +8.01 and −5.44bps, on average. The serial correlation is −0.7958 (−0.2573) for negative (positive) drift

[21] We construct a normalised 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.

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

[23] These price changes are far out in the tails of the return distribution. A −37.39bps 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.

Electronic copy available at: https://ssrn.com/abstract=2842335

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 sample 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).

bursts. Looking further into 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 the positive drift bursts. At last, volatility is markedly higher for tail drift bursts (in either direction) and those accompanied by 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 reversion is stronger if the price change or volume is large. The results are again more pronounced for negative drift bursts.

Electronic copy available at: https://ssrn.com/abstract=2842335

Table 6: ES post-drift burst return conditional on pre-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 $R^+_{t_j}$ for each subgroup after double-sorting first by the pre-drift burst return $R^-_{t_j}$ and second by the normalized gross trading volume $V^-_{t_j}$. As in Huang and Wang (2009), the sample is divided into three buckets, where "low" is defined as the 1st quartile, "medium" is the inter-quartile range, and "high" the 4th quartile. The numbers are expressed in basis points (bps).

As an alternative way to describe the double-conditioning of Table 6, we estimate the forecasting equation of Campbell, Grossman, and Wang (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 notice 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, Table 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 likely symptom of endogenous demand for liquidity due to costly market presence. 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.

Electronic copy available at: https://ssrn.com/abstract=2842335

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, Grossman, and Wang (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 for positive and negative drift bursts and by conditioning on $R^-_{t_j}$ being below and above its average value $E[R^-_{t_j}]$ in each of the two subsamples.

6 Conclusion

The drift burst hypothesis is proposed as a theoretical framework for the modelling 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 provide 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 made in the market microstructure literature of liquidity provision.

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? Other areas of the literature that may be impacted include, for instance, the pricing of barrier options, the properties of stop-loss and take-profit orders commonly used by retail investors, intra-day value at risk calculations, and more generally the microstructure of liquidity provision.

Electronic copy available at: https://ssrn.com/abstract=2842335

A Mathematical Appendix

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

Assumption 3 *Using the notation of Assumption 1:*

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 $\mathbb{R}$, while $\delta : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ is predictable and 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 nonnegative Γ_n with $min(|\delta(t,x)|,1) \leq \Gamma_n(x)$ and $\int_{\mathbb{R}} \Gamma_n(x)^2 \lambda(dx) < \infty$ for all (t,x) and $n \geq 1$.

ii) Fix $t \in (0,T]$ and let $B_\epsilon(t) = [t-\epsilon, t]$ with $\epsilon > 0$ fixed. We assume there exists a $\Gamma > 0$ and a sequence of $\mathscr{F}_t$-stopping times $\tau_m \to \infty$ and constants $C_t^{(m)}$ such that for all m, $(\omega, s) \in \Omega \times B_\epsilon(t) \cap [0, \tau_m(\omega)[$, and $u \in B_\epsilon(t)$,

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

where $E_t[\cdot] = E[\cdot|\mathscr{F}_t]$.

Remark 1 *Under Assumptions 1 and 3, the localization procedure in Jacod and Protter (2012, Section 4.4.1) implies that we can and shall 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 $\int_{\mathbb{R}} \overline{\Gamma}(x)^2 \lambda(\mathrm{d}x) < \infty$.*

Assumption 4 *$(t_i)_{i=0}^n$ is a deterministic sequence with $\max_{i=1,\ldots,n}\{\Delta_{i,n}\} = O(\Delta_n)$, where $\Delta_n = T/n$ is the equidistant spacing. Moreover, denoting the "quadratic variation of time up to t" as $H(t) = \lim_{n\to\infty} H_n(t)$, where $H_n(t) = \frac{1}{\Delta_n} \sum_{t_i \leq t} \left(\Delta_{i,n}\right)^2$, we assume $H(t)$ exists and is Lebesgue-almost surely differentiable in $(0,T)$ with derivative H' such that: $\left|H'(t_i) - \Delta_{i,n}/\Delta_n\right| \leq C\Delta_{i,n}$, for some $C \geq 0$ (not depending on i).*

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;

Electronic copy available at: https://ssrn.com/abstract=2842335

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

(K3) It holds that for every positive sequence $g_n \to \infty$*,* $\int_{-\infty}^{-g_n} K(x)\mathrm{d}x \leq C g_n^{-\beta}$ *for some* $\beta > 0$ *and* $C > 0$ *(i.e., K has a fast vanishing tail);*

(K4) $m_K(\alpha) = \int_{-\infty}^{0} K(x)|x|^{\alpha}\mathrm{d}x < \infty$*, for all* $\alpha \geq -\Gamma$*.*

Remark 2 *Condition (K0) is without loss of generality and can be relaxed to allow for a two-sided kernel without changing any of the theoretical results with minor modifications in the proofs. We impose it just to make the mathematical exposition less cumbersome and aligned to 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 Assumption 1, 4 and 5 hold. Then, for every fixed* $t \in (0, T]$ *as* $n \to \infty$ *and* $h_n \to 0$*, it holds that:*

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

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

This also applies if μ_t *is replaced by* σ_t*.*

Proof. Write:

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

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:

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

The proof for the term B_n follows along the same line. ∎

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

$$B_n = \int_0^T \frac{1}{h_n}K\left(\frac{s-t}{h_n}\right)\mu_s\mathrm{d}s - \mu_{t-} = O_p\left(h_n^{\Gamma/2} + h_n^B\right).$$

Electronic copy available at: https://ssrn.com/abstract=2842335

This also applies if μ_t is replaced by σ_t.

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

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

so we can write:

$$|B_n| = \left|\int_0^T \frac{1}{h_n}K\left(\frac{s-t}{h_n}\right)(\mu_s-\mu_{t-})\mathrm{d}s + \mu_{t-}\int_{-\infty}^{-t/h_n} K(x)\mathrm{d}x\right| \leq \int_0^T \frac{1}{h_n}K\left(\frac{s-t}{h_n}\right)|\mu_s-\mu_{t-}|\mathrm{d}s + Ch_n^B,$$

where (K3) is applied. Then, by Jensen's inequality and Eq. (39):

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

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

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

This concludes the proof. ■

Lemma 3 *Assume that the conditions of Assumption 1, 3, 4, and 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 $X'_t = \int_0^t \mu_s^* \mathrm{d}s + \int_0^t \sigma_s \mathrm{d}W_s$, where $\mu_t^* = \mu_t + \int_{\mathbb{R}} \delta(t,x) I_{\{|\delta(t,x)|>1\}}\lambda(\mathrm{d}x)$ is bounded, and $X''_t = X_t - X'_t = \int_{\mathbb{R}} \delta(s,x)(\nu(\mathrm{d}s,\mathrm{d}x) - \tilde{\nu}(\mathrm{d}s,\mathrm{d}x))$. Now, Mancini, Mattiussi, and Renò (2015) established the convergence in probability:

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

Thus, it is enough to show that:

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

Electronic copy available at: https://ssrn.com/abstract=2842335

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

$$\Delta_i^n X_1''(\kappa) = \int_{\mathbb{R}} \delta(s,x) I_{\{|\delta(s,x)|\le\kappa\}} (\nu(\mathrm{d}s,\mathrm{d}x) - \tilde{\nu}(\mathrm{d}s,\mathrm{d}x)) \quad \text{and} \quad \Delta_i^n X_2''(\kappa) = \int_{\mathbb{R}} \delta(s,x) I_{\{|\delta(s,x)|>\kappa\}} (\nu(\mathrm{d}s,\mathrm{d}x) - \tilde{\nu}(\mathrm{d}s,\mathrm{d}x)).$$

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

$$\begin{aligned} |R_n^{\hat{\sigma}}| &\le \frac{1}{h_n} \sum_{i=1}^n K\left(\frac{t_{i-1}-t}{h_n}\right) \left| \left(\Delta_i^n X\right)^2 - \left(\Delta_i^n X'\right)^2 \right| \\ &\le \frac{1}{h_n} \sum_{i=1}^n K\left(\frac{t_{i-1}-t}{h_n}\right) \left(\epsilon \left(\Delta_i^n X'(\kappa)\right)^2 + \frac{C}{\epsilon} \left(\left(\Delta_i^n X_1''(\kappa)\right)^2 + \left(\Delta_i^n X_2''(\kappa)\right)^2 \right) \right), \end{aligned}$$

for $0 < \epsilon \le 1$.

Next, define $\Omega_n(\psi,\kappa) \subseteq \Omega$ such that the associated Poisson process has no jumps of size greater than κ in the interval $(t - h_n^{\psi}, t]$, for $0 < \psi < 1$. Note that $\Omega_n(\psi,\kappa) \to \Omega$, as $n \to \infty$ and $h_n^{\psi} \to 0$. On $\Omega_n(\psi,\kappa)$, we find that:

$$\begin{aligned} E\left[\frac{1}{h_n} \sum_{i=1}^n K\left(\frac{t_{i-1}-t}{h_n}\right) \frac{C}{\epsilon} \left(\Delta_i^n X_2''(\kappa)\right)^2 | \, \Omega_n(\psi,\kappa) \right] &= \frac{1}{h_n} \sum_{t_{i-1} \le t - h_n^{\psi}} K\left(\frac{t_{i-1}-t}{h_n}\right) \frac{C}{\epsilon} E\left[\left(\Delta_i^n X_2''(\kappa)\right)^2 \right] \\ &\le \frac{1}{h_n} \sum_{t_{i-1} \le t - h_n^{\psi}} K\left(\frac{t_{i-1}-t}{h_n}\right) \frac{C}{\epsilon} \Delta_{i,n} \\ &\le \frac{C}{\epsilon} h_n^{B(1-\psi)} \end{aligned}$$

by a Riemann approximation and (K3). Moreover,

$$E\left[\left(\Delta_i^n X_1''\right)^2 \right] \le C \Delta_{i,n} \int_{\{x:\bar{\Gamma}(x)\le\kappa\}} \bar{\Gamma}(x)^2 \lambda(\mathrm{d}x),$$

so that

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

Now, writing $\mathscr{P}\left(|R_n^{\hat{\sigma}}| > c\right) = \mathscr{P}\left(|R_n^{\hat{\sigma}}| > c \mid \Omega_n^{\complement}(\psi,\kappa)\right) + \mathscr{P}\left(|R_n^{\hat{\sigma}}| > c \mid \Omega_n(\psi,\kappa)\right) \le \mathscr{P}\left(\Omega_n^{\complement}(\psi,\kappa)\right) + E\left[|R_n^{\hat{\sigma}}| \, | \, \Omega_n(\psi,\kappa)\right]/c$ (by Markov's inequality), it follows that:

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

Electronic copy available at: https://ssrn.com/abstract=2842335

Setting $\epsilon = \left(\int_{\{x:\bar{\Gamma}(x)\leq\kappa\}} \bar{\Gamma}(x)^2\lambda(\mathrm{d}x) + h_n^{B(1-\psi)}\right)^{\psi'}$ with $0<\psi'<1$ and noticing that $\left(\int_{\{x:\bar{\Gamma}(x)\leq\kappa\}} \bar{\Gamma}(x)^2\lambda(\mathrm{d}x) + h_n^{B(1-\psi)}\right) \to 0$ 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:

$$T_t^n = \underbrace{\sqrt{\frac{h_n}{\mathbf{K}_2}}\frac{(\hat{\mu}_t^n - \mu_{t-}^*)}{\hat{\sigma}_t^n}}_{T_1} + \underbrace{\sqrt{\frac{h_n}{\mathbf{K}_2}}\frac{\mu_{t-}^*}{\hat{\sigma}_t^n}}_{T_2},$$

where μ_t^* is the 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 Lemma 1 – 2 we can write:

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

Hence,

$$\begin{aligned}\sqrt{h_n}\left(\hat{\mu}_t^n - \mu_{t-}^*\right) = &\underbrace{\frac{1}{\sqrt{h_n}}\sum_{i=1}^n K\left(\frac{t_{i-1}-t}{h_n}\right)\int_{t_{i-1}}^{t_i}\sigma_s\mathrm{d}W_s}_{G_n} \\ &+ \underbrace{\frac{1}{\sqrt{h_n}}\sum_{i=1}^n K\left(\frac{t_{i-1}-t}{h_n}\right)\int_{t_{i-1}}^{t_i}\int_{\mathbb{R}}\delta(s,x)(\nu(\mathrm{d}s,\mathrm{d}x)-\tilde{\nu}(\mathrm{d}s,\mathrm{d}x))}_{G_n'} + O_p\left(\frac{\sqrt{h_n}}{nh_n} + h_n^{\Gamma/2+1/2} + h_n^{B+1/2}\right).\end{aligned}$$

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

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

Electronic copy available at: https://ssrn.com/abstract=2842335

and since $\mathscr{P}\left(|G'_{n,2}|>c\right)\leq\mathscr{P}\left(\Omega_n^{\complement}(\kappa,\psi)\right)+E\Big[|G'_{n,2}|\,|\,\Omega_n(\kappa,\psi)\Big]/c$, 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 $G'_{n,2}\xrightarrow{p}0$ as $n\to\infty$. Next, we notice that

$$E\Big[(G'_{n,1})^2\Big]=\frac{1}{h_n}\sum_{i=1}^{n}K^2\left(\frac{t_{i-1}-t}{h_n}\right)E\Big[\left(\Delta_i^n X_2''(\kappa)\right)^2\Big]\leq\frac{C}{h_n}\sum_{i=1}^{n}K^2\left(\frac{t_{i-1}-1}{h_n}\right)\Delta_{i,n}\int_{\{x:\overline{\Gamma}(x)\leq\kappa\}}\overline{\Gamma}(x)^2\lambda(\mathrm{d}x).$$

The bound converges to $C\mathbf{K}_2\int_{\{x:\overline{\Gamma}(x)\leq\kappa\}}\overline{\Gamma}(x)^2\lambda(\mathrm{d}x)$, which can be made arbitrarily small by letting $\kappa\to0$. We conclude that $G'_{n,1}\xrightarrow{p}0$ and, therefore, $G'_n=o_p(1)$.

G_n is the leading term, which we write as $G_n=\sum_{i=1}^n\Delta_i^n u$ with $\Delta_i^n u=\frac{1}{\sqrt{h_n}}K\left(\frac{t_{i-1}-t}{h_n}\right)\int_{t_{i-1}}^{t_i}\sigma_s\mathrm{d}W_s$. The aim is to prove that G_n – and hence $\sqrt{h_n}\left(\hat{\mu}_t^n-\mu_{t-}^*\right)$ – converges stably in law to $N\left(0,\mathbf{K}_2\sigma_{t-}^2\right)$. We exploit Theorem 2.2.14 in Jacod and Protter (2012), which lists four sufficient conditions for this to hold:

$$\sum_{i=1}^{n}E_{t_{i-1}}\Big[\Delta_i^n u\Big]\xrightarrow{p}0, \tag{40}$$

$$\sum_{i=1}^{n}E_{t_{i-1}}\Big[(\Delta_i^n u)^2\Big]\xrightarrow{p}\mathbf{K}_2\sigma_{t-}^2, \tag{41}$$

$$\sum_{i=1}^{n}E_{t_{i-1}}\Big[(\Delta_i^n u)^4\Big]\xrightarrow{p}0, \tag{42}$$

$$\sum_{i=1}^{n}E_{t_{i-1}}\Big[\Delta_i^n u\Delta_i^n Z\Big]\xrightarrow{p}0, \tag{43}$$

where either $Z_t=W_t$ or $Z_t=W'_t$ with W'_t being orthogonal to W_t. The condition in Eq. (40) is immediate. Next, from Itô's Lemma, we deduce that:

$$\left(\int_{t_{i-1}}^{t_i}\sigma_s\mathrm{d}W_s\right)^2=\int_{t_{i-1}}^{t_i}\sigma_s^2\mathrm{d}s+2\int_{t_{i-1}}^{t_i}\sigma_s\left(\int_{t_{i-1}}^{s}\sigma_u\mathrm{d}W_u\right)\mathrm{d}W_s,$$

so that

$$\sum_{i=1}^{n}E_{t_{i-1}}\Big[(\Delta_i^n u)^2\Big]=\sum_{i=1}^{n}\frac{1}{h_n}K^2\left(\frac{t_{i-1}-t}{h_n}\right)E_{t_{i-1}}\left[\left(\int_{t_{i-1}}^{t_i}\sigma_s\mathrm{d}W_s\right)^2\right]$$

Electronic copy available at: https://ssrn.com/abstract=2842335

$$= \sum_{i=1}^{n} \frac{1}{h_n} K^2\left(\frac{t_{i-1}-t}{h_n}\right) E_{t_{i-1}}\left[\int_{t_{i-1}}^{t_i} \sigma_s^2 \mathrm{d}s\right]$$
$$= \sum_{i=1}^{n} \frac{1}{h_n} K^2\left(\frac{t_{i-1}-t}{h_n}\right)\left(\sigma_{t_{i-1}}^2 \Delta_{i,n} + E_{t_{i-1}}\left[\int_{t_{i-1}}^{t_i} \left(\sigma_s^2 - \sigma_{t_{i-1}}^2\right)\mathrm{d}s\right]\right).$$

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

$$\sum_{i=1}^{n} \frac{1}{h_n} K^2\left(\frac{t_{i-1}-t}{h_n}\right) E_{t_{i-1}}\left[\int_{t_{i-1}}^{t_i} \left(\sigma_s^2 - \sigma_{t_{i-1}}^2\right)\mathrm{d}s\right] \leq \sum_{i=1}^{n} \frac{1}{h_n} K^2\left(\frac{t_{i-1}-t}{h_n}\right) \Delta_{i,n}\Delta_{i,n}^{\Gamma} = O_p\left(\Delta_n^{\Gamma}\right). \tag{44}$$

To deal with the third condition in Eq. (42), we notice that by the Burkholder-Davis-Gundy inequality and from the boundedness of σ_t, it holds that:

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

which leads to

$$\sum_{i=1}^{n} E_{t_{i-1}}\left[(\Delta_i^n u)^4\right] = \sum_{i=1}^{n} \frac{1}{h_n^2} K^4\left(\frac{t_{i-1}-t}{h_n}\right) E_{t_{i-1}}\left[\left(\int_{t_{i-1}}^{t_i} \sigma_s \mathrm{d}W_s\right)^4\right]$$
$$\leq C \sum_{i=1}^{n} \frac{1}{h_n^2} K^4\left(\frac{t_{i-1}-t}{h_n}\right)(\Delta_{i,n})^2 = O\left(\frac{\Delta_n}{h_n}\right).$$

To deal with Eq. (43), we first set $Z_t = W_t$. Then, using the Cauchy-Schwartz inequality:

$$E_{t_{i-1}}\left[\Delta_i^n W \int_{t_{i-1}}^{t_i} \sigma_s \mathrm{d}W_s\right] \leq \sqrt{E_{t_{i-1}}\left[(\Delta_i^n W)^2\right]}\sqrt{E_{t_{i-1}}\left[\left(\int_{t_{i-1}}^{t_i} \sigma_s \mathrm{d}W_s\right)^2\right]}$$
$$= \sqrt{\Delta_{i,n}}\sqrt{E_{t_{i-1}}\left[\int_{t_{i-1}}^{t_i} \sigma_s^2 \mathrm{d}s\right]} = O_p(\Delta_n),$$

and therefore

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

Electronic copy available at: https://ssrn.com/abstract=2842335

If $Z_t = W'_t$, the process $W'_t \int_0^t \sigma_s \mathrm{d}W_s$ is a martingale by orthogonality, so that:

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

This verifies that $\sqrt{h_n}\left(\hat{\mu}_t^n - \mu_{t-}^*\right) \xrightarrow{d} N\left(0, \mathbf{K}_2 \sigma_{t-}^2\right)$, where the convergence is in law stably. Combined with Lemma 3, this yields $T_t^n \xrightarrow{d} N(0,1)$. ■

Proof of Theorem 2. Without loss of generality, we set $\tau_{\mathrm{db}} = T = 1$ and conservatively assume that $\tilde{\mu}_t = (1-t)^{-\alpha}$ and $\tilde{\sigma}_t = (1-t)^{-\beta}$. We write $\tilde{X}_t = X_t + D_t + V_t$, where $D_t = \int_0^t (1-s)^{-\alpha}\mathrm{d}s$ and $V_t = \int_0^t (1-s)^{-\beta}\mathrm{d}W_s$. In Theorem 1, we already showed that $\frac{1}{h_n}\sum_{i=1}^n K\left(\frac{t_{i-1}-1}{h_n}\right)\Delta_i^n X = O_p\left(\frac{1}{\sqrt{h_n}}\right)$. Further,

$$\frac{1}{h_n}\sum_{i=1}^n K\left(\frac{t_{i-1}-1}{h_n}\right)\Delta_i^n D = \frac{1}{h_n}\sum_{i=1}^n K\left(\frac{t_{i-1}-1}{h_n}\right)\Delta_{i,n}(1-\xi_{t_{i-1},t_i})^{-\alpha},$$

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 $h_n^{-\alpha} m_K(-\alpha)$, where $m_K(-\alpha)$ is the constant in (K4).

The term $\frac{1}{h_n}\sum_{i=1}^n K\left(\frac{t_{i-1}-1}{h_n}\right)\Delta_i^n V$ has mean zero and variance:

$$\frac{1}{h_n^2}\sum_{i=1}^n K^2\left(\frac{t_{i-1}-1}{h_n}\right)\int_{t_{i-1}}^{t_i}(1-s)^{-2\beta}\mathrm{d}s.$$

As before, this is asymptotically equal to $h_n^{-(2\beta+1)} m_K(-2\beta)$. Thus, we conclude that $\frac{1}{h_n}\sum_{i=1}^n K\left(\frac{t_{i-1}-1}{h_n}\right)\Delta_i^n V = O_p\left(h_n^{-(\beta+1/2)}\right)$ and $\hat{\mu}_t^n = O_p(h_n^{-\alpha})$, since $\alpha - \beta > 1/2$.

We then 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

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

Now,

$$\begin{aligned}\frac{1}{h_n}\sum_{i=1}^n K\left(\frac{t_{i-1}-1}{h_n}\right)(\Delta_i^n D)^2 &= \frac{1}{h_n}\sum_{i=1}^n K\left(\frac{t_{i-1}-1}{h_n}\right)\left(\int_{t_{i-1}}^{t_i}(1-s)^{-\alpha}\mathrm{d}s\right)^2 \\ &= \frac{1}{(1-\alpha)^2}\frac{1}{h_n}\sum_{i=1}^n K\left(\frac{t_{i-1}-1}{h_n}\right)\left((1-t_{i-1}-\Delta_{i,n})^{1-\alpha} - (1-t_{i-1})^{1-\alpha}\right)^2\end{aligned}$$

Electronic copy available at: https://ssrn.com/abstract=2842335

$$= \frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-1}{h_n}\right)(1-\xi_{t_{i-1},t_i})^{-2\alpha}\Delta_{i,n}^2 = O\left(\frac{\Delta_n^{2(1-\alpha)}}{h_n}\right),$$

for some $t_{i-1} \leq \xi_{t_{i-1},t_i} \leq t_i$. The final order is due to:

$$\sum_{i=1}^{n} K\left(\frac{t_{i-1}-1}{h_n}\right)(1-\xi_{t_{i-1},t_i})^{-2\alpha}\Delta_{i,n}^2 \leq C\Delta_n^2\sum_{i=1}^{n}(\xi_{t_{i-1},t_i})^{-2\alpha} \sim \Delta_n^{2(1-\alpha)}\sum_{i=1}^{n}\frac{1}{i^{2\alpha}},$$

where the sum is convergent (to a strictly positive number), because $1/2 < \alpha < 1$. Moreover,

$$E\left[\frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-1}{h_n}\right)(\Delta_i^n\ V)^2\right] = \frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-1}{h_n}\right)\int_{t_{i-1}}^{t_i}(1-s)^{-2\beta}\,\mathrm{d}s,$$

which is asymptotically equivalent to $m_K(-2\beta)h_n^{-2\beta}$. Finally, we exploit that for all $\epsilon > 0$ and a, b and c real: $(a+b+c)^2 - a^2 - b^2 \leq \epsilon(a^2+b^2) + \frac{1+\epsilon}{\epsilon}c^2$, so that

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

so letting $\epsilon \sim h_n^{\beta}$, we can make R_n' negligible. Thus, $(\hat{\sigma}_t^n)^2 = O_p\left(\frac{\Delta_n^{2(1-\alpha)}}{h_n} + h_n^{-2\beta}\right)$. This implies that the rate of divergence is determined by the speed at which $h_n \to 0$. Write $h_n \sim \Delta_n^{\xi}$ (with $0 < \xi < 1$ to ensure $nh_n \to \infty$). The rate depends on whether $\xi > \frac{2(1-\alpha)}{1-2\beta}$. If the condition is true, which is possible only if $\alpha - \beta > 1/2$, then $|T_{\tau_{\mathrm{db}}}^n| = O_p\left((nh_n)^{1-\alpha}\right) \to \infty$. If it is false, $|T_{\tau_{\mathrm{db}}}^n| = O_p\left(h_n^{1/2-\alpha+\beta}\right)$, which diverges if and only if $\alpha - \beta > 1/2$. ■

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

$$T_t^n = \sqrt{\frac{h_n}{\mathbf{K}_2}}\frac{(\hat{\mu}_t^n - \mu_{t-}^*)}{\hat{\sigma}_t^n} + \sqrt{\frac{h_n}{\mathbf{K}_2}}\frac{\mu_{t-}^*}{\hat{\sigma}_t^n} = \sqrt{\frac{h_n}{\mathbf{K}_2}}\frac{(\hat{\mu}_t^n - \mu_{t-}^*)}{\sigma_{t-}} + \underbrace{\sqrt{\frac{h_n}{\mathbf{K}_2}}\frac{(\hat{\mu}_t^n - \mu_{t-}^*)}{\sigma_{t-}}\left(\frac{\sigma_{t-}}{\hat{\sigma}_t^n} - 1\right)}_{T_{t,1}^n} + \underbrace{\sqrt{\frac{h_n}{\mathbf{K}_2}}\frac{\mu_{t-}^*}{\hat{\sigma}_t^n}}_{O_p(\sqrt{h_n})}, \tag{45}$$

where the last order is uniform in t. From the proof of Theorem 1 (and with that notation):

$$\sqrt{\frac{h_n}{\mathbf{K}_2}}\frac{(\hat{\mu}_t^n - \mu_{t-}^*)}{\sigma_{t-}} = \sqrt{\frac{1}{\mathbf{K}_2}}\frac{G_n + G_n' + O_p\left(\frac{\sqrt{h_n}}{nh_n} + h_n^{\Gamma/2+1/2} + h_n^{B+1/2}\right)}{\sigma_{t-}},$$

Electronic copy available at: https://ssrn.com/abstract=2842335

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

We decompose G_n as:

$$\frac{G_n}{\sigma_{t-}}=\underbrace{\frac{1}{\sqrt{h_n}}\sum_{i=1}^{n}K\left(\frac{t_{i-1}-t}{h_n}\right)I_{\{-h_n\le t_{i-1}-t\le 0\}}\left(W_{t_{i-1}}-W_{t_i}\right)}_{G^n_{t,1}}+\underbrace{\frac{1}{\sigma_{t-}\sqrt{h_n}}\sum_{i=1}^{n}K\left(\frac{t_{i-1}-t}{h_n}\right)I_{\{t_{i-1}-t\le -h_n\}}\int_{t_{i-1}}^{t_i}\sigma_s\mathrm{d}W_s}_{G^n_{t,2}}$$
$$+\underbrace{\frac{1}{\sigma_{t-}\sqrt{h_n}}\sum_{i=1}^{n}K\left(\frac{t_{i-1}-t}{h_n}\right)I_{\{-h_n\le t_{i-1}-t\le 0\}}\int_{t_{i-1}}^{t_i}(\sigma_s-\sigma_{t-})\mathrm{d}W_s}_{G^n_{t,3}}.$$

Thus, since $nh_n\to 0$ and $mh_n\to 0$, $G^n_{t_i,1}\perp\!\!\!\perp G^n_{t_j,1}$ for $i\neq j$ so that:

$$a_m\left(\max_{i=1,\dots,m}G^n_{t_i,1}-b_m\right)\xrightarrow{d}\xi.$$

The fast vanishing tails of the kernel in (K3) further imply that:

$$\max_{i=1,\dots,m}G^n_{t_i,2}\le\sum_{i=1}^{m}\left|G^n_{t_i,2}\right|=O_p\left(m\frac{1}{\sqrt{nh_n}}h_n^B\right),$$

while we have:

$$\max_{i=1,\dots,m}G^n_{t_i,3}\le\sum_{i=1}^{m}\left|G^n_{t_i,3}\right|=O_p\left(mn^{-\Gamma/2}\right).$$

where the order comes from Eq. (39) and Burkholder-Davis-Gundy inequality, using the same strategy leading to Eq. (44). The middle term in Eq. (45) has

$$\max_{i=1,\dots,m}T^n_{t_i,1}\le\sum_{i=1}^{m}\left|T^n_{t_i,1}\right|=O_p\left(m\frac{1}{\sqrt{nh_n}}\right).$$

Thus, if $a_m m\left(\frac{1}{\sqrt{nh_n}}+n^{-\Gamma/2}\right)\to 0$, we conclude that $a_m\left(T^*_m-b_m\right)\xrightarrow{d}\xi$. ■

Proof of Theorem 4. As before we set $\tau_{\mathrm{db}}=T=1$. Then,

$$T^n_{\tau_J}=\sqrt{\frac{h_n}{\mathbf{K}_2}}\frac{\hat{\mu}^n_t}{\hat{\sigma}^n_t}=\sqrt{\frac{h_n}{\mathbf{K}_2}}\frac{\frac{1}{h_n}\sum_{i=1}^{n}K\left(\frac{t_{i-1}-1}{h_n}\right)\Delta^n_iX+\frac{1}{h_n}K\left(\frac{\Delta_{n,n}}{h_n}\right)J}{\left(\frac{1}{h_n}\sum_{i=1}^{n}K\left(\frac{t_{i-1}-1}{h_n}\right)(\Delta^n_iX)^2+\frac{2}{h_n}K\left(\frac{\Delta_{n,n}}{h_n}\right)J\Delta^n_nX+\frac{1}{h_n}K\left(\frac{\Delta_{n,n}}{h_n}\right)J^2\right)^{1/2}}$$

Electronic copy available at: https://ssrn.com/abstract=2842335

$$= \frac{N\left(0, \sigma_{t-}^2\right) + \sqrt{\frac{1}{\mathbf{K}_2 h_n}} K\left(\frac{\Delta_{n,n}}{h_n}\right) J + o_p(1)}{\left(\sigma_{t-}^2 + \frac{2}{h_n} K\left(\frac{\Delta_{n,n}}{h_n}\right) J O_p\left(\sqrt{\Delta_n}\right) + \frac{1}{h_n} K\left(\frac{\Delta_{n,n}}{h_n}\right) J^2 + o_p(1)\right)^{1/2}} \xrightarrow{p} \sqrt{\frac{K(0)}{\mathbf{K}_2}} \cdot \operatorname{sign}(J),$$

as $n \to \infty$, $h_n \to 0$ and $n h_n \to \infty$. ■

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

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

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

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

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 (Jacod, Li, Mykland, Podolskij, and Vetter, 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

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

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

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

Electronic copy available at: https://ssrn.com/abstract=2842335

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.

$M(p)_t^n$ 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 use Theorem 2.2.14 in Jacod and Protter (2012) by taking conditional expectations with respect to the discrete-time filtration $\mathcal{G}(p)_j^n = \mathcal{F}_{t_{\bar{\ell}_j^n(p)-1}}$. We immediately get the orthogonality condition (43), and

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

As for the conditional variance:

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

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

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

Electronic copy available at: https://ssrn.com/abstract=2842335

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

To deal with the term $V_{1,n}$ we write, for k_n sufficiently large, and adopting the convention $H_j^n = g_{j+1}^n - g_j^n$ when $0 \le j \le k_n - 1$ and $H_j^n = 0$ when $j < 0$ or $j \ge k_n$,

$$\begin{aligned}E\left[\Delta_{\ell-1}^n\bar{\epsilon}\cdot\Delta_{\ell'-1}^n\bar{\epsilon}\mid\mathscr{G}(p)_j^n\right] &= E\left[\left(\sum_{j=0}^{k_n-1}H_j^n\epsilon_{\ell+j-1}\right)\left(\sum_{j=0}^{k_n-1}H_j^n\epsilon_{\ell'+j-1}\right)\right]\\ &= \gamma(0)\sum_{j=0}^{k_n-\ell'+\ell-1}H_j^nH_{j+\ell'-\ell}^n+\sum_{q=1}^{Q}\gamma(q)\sum_{j=0}^{k_n-1}H_j^n\left(H_{j-q+\ell'-\ell}^n+H_{j+q+\ell'-\ell}^n\right).\end{aligned}$$

This implies that, for pk_n large enough to include all non-zero autocovariances,

$$\begin{aligned}E\left[(\Delta_{\ell-1}^n\bar{\epsilon})^2\mid\mathscr{G}(p)_j^n\right]+2\sum_{\ell'=\ell+1}^{\bar{\ell}_j^n(p)+pk_n-1}E\left[\Delta_{\ell-1}^n\bar{\epsilon}\cdot\Delta_{\ell'-1}^n\bar{\epsilon}\mid\mathscr{G}(p)_j^n\right] &= \gamma(0)\sum_{j=0}^{k_n-1}(H_j^n)^2+\sum_{q=1}^{Q}\gamma(q)\sum_{j=0}^{k_n-1}H_j^n\left(H_{j-q}^n+H_{j+q}^n\right)\\ &\quad+2\left(\sum_{L=1}^{k_n+Q-1}\gamma(0)\sum_{j=0}^{k_n-L-1}H_j^nH_{j+L}^n+\sum_{q=1}^{Q}\gamma(q)\sum_{j=0}^{k_n-1}H_j^n\left(H_{j-q+L}^n+H_{j+q+\right.\right.\\ &= (Q+1)\left(\gamma(0)+2\sum_{q=1}^{Q}\gamma(q)\right)\left(\sum_{j=1-Q}^{k_n-1+Q}H_j^n\right)^2=0,\end{aligned}$$

so that $V_{1,n}=0$. Next, $V_{2,n}$ can be rewritten:

$$\begin{aligned}V_{2,n} &= \frac{2}{h_n^2}\sum_{j=0}^{j_n(p)}\sum_{\ell=\bar{\ell}_j^n(p)}^{\bar{\ell}_j^n(p)+pk_n-1}K\left(\frac{t_{\ell-1}-t}{h_n}\right)\sum_{\ell'=\ell+1}^{\bar{\ell}_j^n(p)+pk_n-1}K'\left(\frac{\xi_{t_{\ell-1},t_{\ell'-1}}}{h_n}\right)\frac{t_{\ell'-1}-t_{\ell-1}}{h_n}E\left[\Delta_{\ell-1}^n\bar{\epsilon}\cdot\Delta_{\ell'-1}^n\bar{\epsilon}\mid\mathscr{G}(p)_j^n\right]\\ &= \frac{2}{h_n^3}\sum_{j=0}^{j_n(p)}\sum_{\ell=\bar{\ell}_j^n(p)}^{\bar{\ell}_j^n(p)+pk_n-1}K\left(\frac{t_{\ell-1}-t}{h_n}\right)K'\left(\frac{t_{\ell-1}-t}{h_n}+O\left(p\frac{k_n\Delta_n}{h_n}\right)\right)\sum_{\ell'=\ell+1}^{\bar{\ell}_j^n(p)+pk_n-1}(t_{\ell'-1}-t_{\ell-1})E\left[\Delta_{\ell-1}^n\bar{\epsilon}\cdot\Delta_{\ell'-1}^n\bar{\epsilon}\mid\mathscr{G}(p)_j^n\right]\\ &= \frac{2}{h_n^3}\sum_{j=0}^{j_n(p)}\sum_{\ell=\bar{\ell}_j^n(p)}^{\bar{\ell}_j^n(p)+pk_n-1}K\left(\frac{t_{\ell-1}-t}{h_n}\right)K'\left(\frac{t_{\ell-1}-t}{h_n}+O\left(p\frac{k_n\Delta_n}{h_n}\right)\right)\sum_{\ell'=\ell+1}^{\bar{\ell}_j^n(p)+pk_n-1}(t_{\ell'-1}-t_{\ell-1})\phi_{\ell,\ell',n},\end{aligned}$$

Electronic copy available at: https://ssrn.com/abstract=2842335

where

$$\phi_{\ell,\ell',n} = \gamma(0)\sum_{j=0}^{k_n-1} H_j^n H_{j+\ell'-\ell}^n + \sum_{q=1}^{Q}\gamma(q)\sum_{j=0}^{k_n-1} H_j^n\left(H_{j-q+\ell'-\ell}^n + H_{j+q+\ell'-\ell}^n\right).$$

Notice that:

$$\sum_{\ell'=\ell+1}^{\bar{\ell}_j^n(p)+pk_n-1}\left[(t_{\ell'-1}-t_{\ell-1})-(\ell'-\ell)\Delta_n\right]\phi_{\ell,\ell',n} = \sum_{\ell'=\ell+1}^{\bar{\ell}_j^n(p)+pk_n-1}\left[\sum_{k=\ell}^{\ell'-1}(\Delta_{i,n}-\Delta_n)\right]\phi_{\ell,\ell',n} = O(\Delta_n)\sum_{\ell'=\ell+1}^{\bar{\ell}_j^n(p)+pk_n-1}\phi_{\ell,\ell',n}.$$

We now have, for a fixed integer L (see, e.g., Eq. (5.36) in Jacod, Li, Mykland, Podolskij, and Vetter, 2009),

$$\sum_{j=0}^{k_n-1} H_j^n H_{j+L}^n = \frac{1}{k_n}\phi_1\left(\frac{L}{k_n}\right) + O_p\left(p\Delta_n + \Delta_n^{\frac{\Gamma+1}{2}}\right),$$

where $\phi_1(s) = \int_s^1 g'(u)g'(u-s)\mathrm{d}u$ when $0 \le s \le 1$ and $\phi_1(s) = 0$ otherwise. It therefore follows that, for $\ell' \ge \ell$, and because $Q/k_n \to 0$,

$$\sum_{\ell'=\ell+1}^{\bar{\ell}_j^n(p)+pk_n-1}\phi_{\ell,\ell',n} \to LRV_\epsilon\int_0^1\phi_1(s)\mathrm{d}s,$$

where $LRV_\epsilon = \gamma_0 + 2\sum_{q=1}^{Q}\gamma(q)$. Hence, $V_{2,n}$ is asymptotically equivalent to

$$\frac{2}{h_n^3}\sum_{j=0}^{j_n(p)}\sum_{\ell=j(p+1)k_n}^{\bar{\ell}_j^n(p)} K\left(\frac{t_{\ell-1}-t}{h_n}\right)K'\left(\frac{t_{\ell-1}-t}{h_n} + O\left(p\frac{k_n\Delta_n}{h_n}\right)\right)\sum_{\ell'=\ell+1}^{\bar{\ell}_j^n(p)}(\ell'-\ell)\Delta_n\phi_{\ell,\ell',n}.$$

As $p \to \infty$ such that $pk_n\Delta_n/h_n \to 0$, a standard Riemann approximation (see Lemma 4) and the fact that the big blocks dominate the small blocks as $p \to \infty$ shows that:

$$\frac{h_n^2}{k_n}V_{2,n} \xrightarrow{p} 2LRV_\epsilon\mathbf{K}_3\int_0^1 s\phi_1(s)\mathrm{d}s.$$

We use the multinomial theorem for the fourth conditional moment:

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

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

Electronic copy available at: https://ssrn.com/abstract=2842335

$$
\begin{aligned}
&+4\sum_{\ell_1\neq\ell_2} K^3\left(\frac{t_{\ell_1-1}-t}{h_n}\right)K\left(\frac{t_{\ell_2-1}-t}{h_n}\right)E\left[\left(\Delta^n_{\ell_1-1}\bar{\epsilon}\right)^3\Delta^n_{\ell_2-1}\bar{\epsilon}\right]\\
&+6\sum_{\ell_1\neq\ell_2} K^2\left(\frac{t_{\ell_1-1}-t}{h_n}\right)K^2\left(\frac{t_{\ell_2-1}-t}{h_n}\right)E\left[\left(\Delta^n_{\ell_1-1}\bar{\epsilon}\right)^2\left(\Delta^n_{\ell_2-1}\bar{\epsilon}\right)^2\right]\\
&+12\sum_{\ell_1\neq\ell_2\neq\ell_3} K^2\left(\frac{t_{\ell_1-1}-t}{h_n}\right)K\left(\frac{t_{\ell_2-1}-t}{h_n}\right)K\left(\frac{t_{\ell_3-1}-t}{h_n}\right)E\left[\left(\Delta^n_{\ell_1-1}\bar{\epsilon}\right)^2\Delta^n_{\ell_2-1}\bar{\epsilon}\Delta^n_{\ell_3-1}\bar{\epsilon}\right]\\
&+24\sum_{\ell_1\neq\ell_2\neq\ell_3\neq\ell_4} K\left(\frac{t_{\ell_1-1}-t}{h_n}\right)K\left(\frac{t_{\ell_2-1}-t}{h_n}\right)K\left(\frac{t_{\ell_3-1}-t}{h_n}\right)K\left(\frac{t_{\ell_4-1}-t}{h_n}\right)E\left[\Delta^n_{\ell_1-1}\bar{\epsilon}\Delta^n_{\ell_2-1}\bar{\epsilon}\Delta^n_{\ell_3-1}\bar{\epsilon}\Delta^n_{\ell_4-1}\bar{\epsilon}\right]
\end{aligned}
$$

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 find that

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

where, as for the variance term,

$$
\begin{aligned}
Q_{1,n} = \frac{1}{k_n^2}\sum_{j=0}^{j_n(p)}\Bigg(\sum_{\ell=\bar{\ell}^n_j(p)}^{\bar{\ell}^n_j(p)+pk_n-1} K^4\left(\frac{t_{\ell-1}-t}{h_n}\right)\Bigg(E\left[\left(\Delta^n_{\ell-1}\bar{\epsilon}\right)^4\right]+4\sum_{\ell_1\neq\ell_2}E\left[\left(\Delta^n_{\ell_1-1}\bar{\epsilon}\right)^3\Delta^n_{\ell_2-1}\bar{\epsilon}\right]+6\sum_{\ell_1\neq\ell_2}E\left[\left(\Delta^n_{\ell_1-1}\bar{\epsilon}\right)^2\left(\Delta^n_{\ell_2-1}\bar{\epsilon}\right)^2\right]\\
+12\sum_{\ell_1\neq\ell_2\neq\ell_3}E\left[\left(\Delta^n_{\ell_1-1}\bar{\epsilon}\right)^2\Delta^n_{\ell_2-1}\bar{\epsilon}\Delta^n_{\ell_3-1}\bar{\epsilon}\right]+24\sum_{\ell_1\neq\ell_2\neq\ell_3\neq\ell_4}E\left[\Delta^n_{\ell_1-1}\bar{\epsilon}\Delta^n_{\ell_2-1}\bar{\epsilon}\Delta^n_{\ell_3-1}\bar{\epsilon}\Delta^n_{\ell_4-1}\bar{\epsilon}\right]\Bigg)\Bigg)=0,
\end{aligned}
$$

while, when $\ell_1+\ell_2+\ell_3+\ell_4=4$, the boundedness of the fourth moment of the noise means

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

We deduce that:

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

Electronic copy available at: https://ssrn.com/abstract=2842335

$$+K\left(\frac{t_{\ell_2-1}-t}{h_n}\right)K\left(\frac{t_{\ell_3-1}-t}{h_n}\right)K'\left(\frac{t_{\ell_4-1}-t}{h_n}\right)+O\left(\frac{pk_n\Delta_n}{h_n}\right)\Bigg)\Bigg)=O_p\left(\frac{1}{k_n}\right).$$

Hence, we conclude that $\frac{h_n}{\sqrt{k_n}}M(p)^n_t \to N\left(0, 2\cdot LRV_\epsilon \mathbf{K}_3 \int_0^1 s\phi_1(s)\mathrm{d}s\right)$.

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

$$\begin{aligned}
E\left[\sup_{s\le t}\left|M'(p)^n_s\right|^2\right] &\le 4\frac{1}{h_n^2}\sum_{j=0}^{j_n(p)} E\left[\left(\sum_{\ell=\bar\ell^n_j(p)+pk_n}^{\bar\ell^n_j(p)+pk_n+k_n-1} K\left(\frac{t_{\ell-1}-t}{h_n}\right)\Delta^n_{\ell-1}\bar\epsilon\right)^2\right] \\
&= \frac{1}{h_n^2}\sum_{j=0}^{j_n(p)}\sum_{\ell}\sum_{\ell'=\ell+1} K\left(\frac{t_{\ell-1}-t}{h_n}\right)K'\left(\frac{\xi_{t_{\ell-1},t_{\ell'-1}}}{h_n}\right)\frac{t_{\ell'-1}-t_{\ell-1}}{h_n}E\left[\Delta^n_{\ell-1}\bar\epsilon\cdot\Delta^n_{\ell'-1}\bar\epsilon\right] \\
&= \frac{1}{h_n^2}\sum_{j=0}^{j_n(p)} K\left(\frac{t_{\bar\ell^n_j(p)+pk_n-1}-t}{h_n}+O\left(\frac{k_n\Delta_n}{h_n}\right)\right)K'\left(\frac{t_{\bar\ell^n_j(p)+pk_n-1}-t}{h_n}+O\left(\frac{k_n\Delta_n}{h_n}\right)\right)\sum_{\ell}\sum_{\ell'=\ell+1}\frac{t_{\ell'-1}-t_{\ell-1}}{h_n}\phi_{\ell,\ell',n}.
\end{aligned}$$

Lemma 4 *Assume that the conditions of Assumption 4 and 5 hold. Examine the decomposition:*

$$\frac{1}{h_n}\sum_{\ell=0}^{\bar\ell^n_{j_n(p)}(p)+pk_n+k_n-1} K\left(\frac{t_{\ell-1}-t}{h_n}\right)\Delta_{i,n} = \frac{1}{h_n}\sum_{j=0}^{j_n(p)}\sum_{\ell=\bar\ell^n_j(p)}^{\bar\ell^n_j(p)+pk_n-1} K\left(\frac{t_{\ell-1}-t}{h_n}\right)\Delta_{i,n} + \frac{1}{h_n}\sum_{j=0}^{j_n(p)}\sum_{\ell=\bar\ell^n_j(p)+pk_n}^{\bar\ell^n_j(p)+pk_n+k_n-1} K\left(\frac{t_{\ell-1}-t}{h_n}\right)\Delta_{i,n}.$$

If $p\to\infty$ such that $pk_n\Delta_n/h_n\to 0$:

$$\frac{1}{h_n}\sum_{j=0}^{j_n(p)}\sum_{\ell=\bar\ell^n_j(p)}^{\bar\ell^n_j(p)+pk_n-1} K\left(\frac{t_{\ell-1}-t}{h_n}\right)\Delta_{i,n} = 1+O\left(\frac{pk_n\Delta_n}{h_n}\right) \qquad \text{and} \qquad \frac{1}{h_n}\sum_{j=0}^{j_n(p)}\sum_{\ell=\bar\ell^n_j(p)+pk_n}^{\bar\ell^n_j(p)+pk_n+k_n-1} K\left(\frac{t_{\ell-1}-t}{h_n}\right)\Delta_{i,n} = O\left(\frac{pk_n\Delta_n}{h_n}\right).$$

This also applies to any integrable function replacing K.

Proof. Write

$$\begin{aligned}
&\frac{1}{h_n}\sum_{j=0}^{j_n(p)}\sum_{\ell=\bar\ell^n_j(p)}^{\bar\ell^n_j(p)+pk_n-1} K\left(\frac{t_{\ell-1}-t}{h_n}\right)\Delta_{i,n} \\
&= \underbrace{\frac{1}{h_n}\sum_{j=0}^{j_n(p)} K\left(\frac{t_{\bar\ell^n_j(p)-1}-t}{h_n}\right)\sum_{\ell=\bar\ell^n_j(p)}^{\bar\ell^n_j(p)+pk_n-1}\Delta_{i,n}}_{=1+O\left(\frac{pk_n\Delta_n}{h_n}\right)} + \underbrace{\frac{1}{h_n}\sum_{j=0}^{j_n(p)}\sum_{\ell=\bar\ell^n_j(p)}^{\bar\ell^n_j(p)+pk_n-1} K'\left(\frac{\xi_{j,\ell}-t}{h_n}\right)\left(t_{\ell-1}-t_{\bar\ell^n_j(p)-1}\right)\Delta_{i,n}}_{=O\left(\frac{pk_n\Delta_n}{h_n}\right)}.
\end{aligned}$$

The first approximation follows from a Riemann argument and the second from the boundedness of K'.

Electronic copy available at: https://ssrn.com/abstract=2842335

Now, using Lemma 4:

$$E\left[\sup_{s\le t}\left|M'(p)_s^n\right|^2\right]=O_p\left(\frac{k_n}{h_n^2}\frac{pk_n\Delta_n}{h_n}\right),$$

which is negligible in comparison to $M(p)_s^n$. Finally, the end-effect term can also be neglected since:

$$\left|\frac{1}{h_n}\sum_{\ell=\bar{\ell}^n_{j_n(p)+1}(p)}^{n-k_n+2}K\left(\frac{t_{\ell-1}-t}{h_n}\right)\Delta^n_{\ell-1}\overline{\epsilon}\right|\le\frac{1}{h_n}\sum_{\ell=\bar{\ell}^n_{j_n(p)+1}(p)}^{n-k_n+2}K\left(\frac{t_{\ell-1}-t}{h_n}\right)\left|\Delta^n_{\ell-1}\overline{\epsilon}\right|\le C\frac{k_n^2}{h_n}.$$

We next analyze the $M_{X,n}$ term and write $M_{X,n}=\tilde{M}(p)_t^n+\tilde{M}'(p)_t^n+\widehat{\widehat{C}}(p)_t^n$ 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 $\tilde{u}_j^n=\frac{1}{h_n}\sum_{\ell=\bar{\ell}^n_j(p)}^{\bar{\ell}^n_j(p)+pk_n-1}K\left(\frac{t_{\ell-1}-t}{h_n}\right)\Delta^n_{\ell-1}\overline{X}$ and then compute:

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

where

$$E\left[\Delta^n_{\ell-1}\overline{X}\Delta^n_{\ell'-1}\overline{X}\mid\mathscr{G}(p)_j^n\right]=E\left[\left(\sum_{j=0}^{k_n-1}H_j^n\Delta_i^nX_{\ell+j-1}\right)\left(\sum_{j=0}^{k_n-1}H_j^n\Delta_i^nX_{\ell'+j-1}\right)\mid\mathscr{G}(p)_j^n\right]=\sum_{j=0}^{k_n-\ell'+\ell-1}H_j^nH^n_{j+\ell'-\ell}\int_{t_{j-1}}^{t_j}\sigma_s^2\mathrm{d}s.$$

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

At last we look at the spot variance estimator:

Electronic copy available at: https://ssrn.com/abstract=2842335

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

$$= \underbrace{\frac{1}{h_n}\sum_{i=1}^{n-k_n+2} K^2\left(\frac{t_{i-1}-t}{h_n}\right) E\left[\left(\Delta_{i-1}^n \bar{\epsilon}\right)^2\right] + \frac{2}{h_n}\sum_{L=1}^{L_n} w\left(\frac{L}{L_n}\right) \sum_{i=1}^{n-k_n-L+2} K\left(\frac{t_{i-1}-t}{h_n}\right) K\left(\frac{t_{i+L-1}-t}{h_n}\right) E\left[\Delta_{i-1}^n \bar{\epsilon} \Delta_{i-1+L}^n \bar{\epsilon}\right]}_{\sigma_{0,n}} + R_n'',$$

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

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

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

$$+ \underbrace{\frac{2}{h_n}\sum_{L=1}^{L_n} w\left(\frac{L}{L_n}\right) \sum_{i=1}^{n-k_n-L+2} K\left(\frac{t_{i-1}-t}{h_n}\right) K'\left(\frac{\xi_{i-1,i+L-1}}{h_n}\right) L\frac{\Delta_n}{h_n} E\left[\Delta_{i-1}^n \bar{\epsilon} \Delta_{i-1+L}^n \bar{\epsilon}\right]}_{\sigma_{3,n}}.$$

As before, it holds that $\sigma_{1,n} = 0$ and, as $L_n \to \infty$,

$$\sigma_{2,n} \overset{p}{\sim} \frac{k_n}{h_n^2} 2LRV_\epsilon \mathbf{K}_3 \int_0^1 s\phi_1(s)\mathrm{d}s,$$

and, applying that $w(L/L_n) = 1 + O\left(\frac{L}{L_n}\right)$, $\sigma_{3,n} = O_p\left(\frac{1}{L_n}\frac{k_n}{h_n^2} 2LRV_\epsilon \mathbf{K}_2 \int_0^1 s\phi_1(s)\mathrm{d}s\right)$, which is negligible as $L_n \to \infty$. This concludes the proof. ■

Proof of Theorem 6. As in the proof of Theorem 2, we set $\tau_{\mathrm{db}} = T = 1$, $\tilde{\mu}_t = (1-t)^{-\alpha}$, $\tilde{\sigma}_t = (1-t)^{-\beta}$ and write: $\tilde{X}_t = X_t + D_t + V_t$. From Theorem 5, $\frac{1}{h_n}\sum_{i=1}^{n-k_n+1} K\left(\frac{t_{i-1}-t}{h_n}\right)\left(\Delta_{i-1}^n \overline{X} + \Delta_{i-1}^n \bar{\epsilon}\right) = O_p\left(\frac{\sqrt{k_n}}{h_n}\right)$. Next, we notice that for suitable $\xi_j \in [t_j, t_{j+1}]$ and using again the mean-value theorem:

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

Electronic copy available at: https://ssrn.com/abstract=2842335

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

As $\frac{1}{k_n}\sum_{j=1}^{k_n-1} g_j^n \to \psi_0 = \int_0^1 g(s)\mathrm{d}s$, the above is $O\left(k_n h_n^{-\alpha}\right)$ 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. $\sqrt{k_n} h_n^{1-\alpha} \to \infty$. For the pre-averaged volatility burst term, we have

$$\frac{1}{h_n} \sum_{i=1}^{n-k_n+2} K\left(\frac{t_{i-1}-1}{h_n}\right) \Delta_{i-1}^n \overline{V} = \frac{1}{h_n} \sum_{i=1}^{n-k_n+2} K\left(\frac{t_{i-1}-1}{h_n}\right) \sum_{j=1}^{k_n-1} g_j^n \int_{t_{i+j-1}}^{t_{i+j}} (1-s)^{-\beta} \mathrm{d}W_s$$
$$= \frac{1}{h_n} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} (1-s)^{-\beta} \mathrm{d}W_s \sum_{j=1}^{k_n-1\wedge i-1} g_j^n K\left(\frac{t_{i-j-1}-1}{h_n}\right),$$

whose variance is

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

which implies

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

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

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

Electronic copy available at: https://ssrn.com/abstract=2842335

and

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

The reasoning in the proof of Theorem 2 then leads to the desired result. ■

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 $\left(t_i^*\right)_{i=1}^m$, 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 Figure 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, 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 Eq. (19) – (20) across the entire universe of simulations and use this table to draw inference.[24]

In Figure 8, we provide an illustration of this approach. In Panel A, we show the ACF of our t-statistic for the

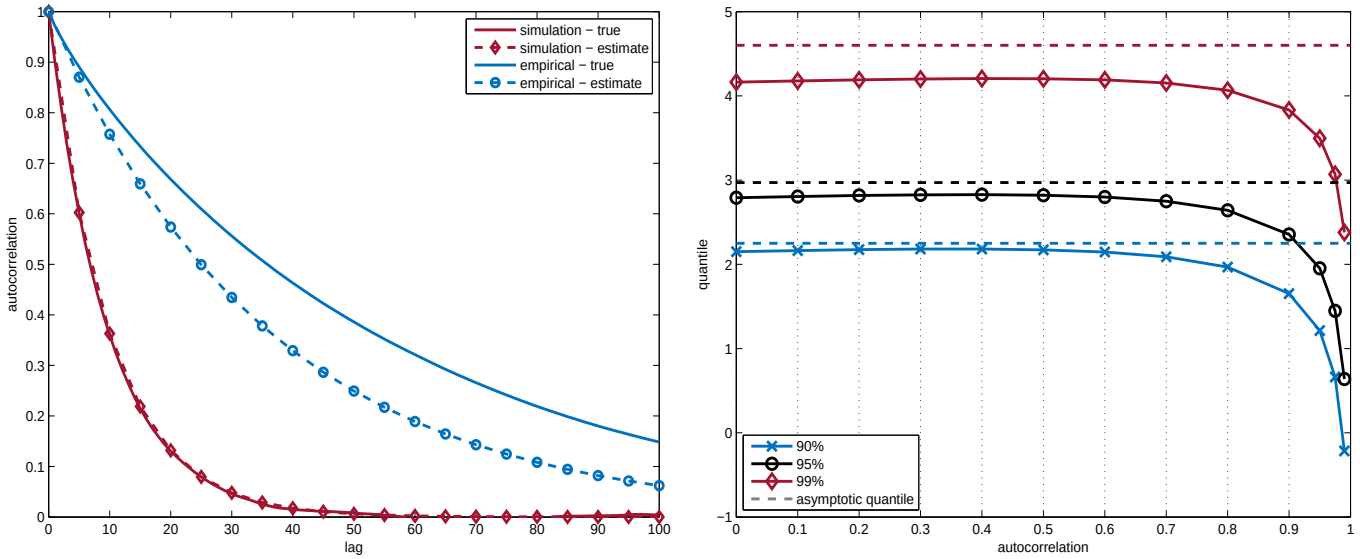
[24] To speed this up for practical work, we prepared in advance a file with the quantile function of the raw and normalized Z_m^* based on the above setup for several choices of m, ρ and selected levels of significance α. This file, along with an interpolation routine to find critical values for other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

Electronic copy available at: https://ssrn.com/abstract=2842335

Figure 8: Autocorrelation function and critical value of t-statistic.

Panel A: ACF

Panel B: Critical value



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 $m = 2,500$ and shows the effect of varying the autocorrelation coefficient ρ and confidence level $1-\alpha$.

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).

Electronic copy available at: https://ssrn.com/abstract=2842335

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

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

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

The log-likelihood for the model is:

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

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

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

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

We test the drift burst hypothesis:

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

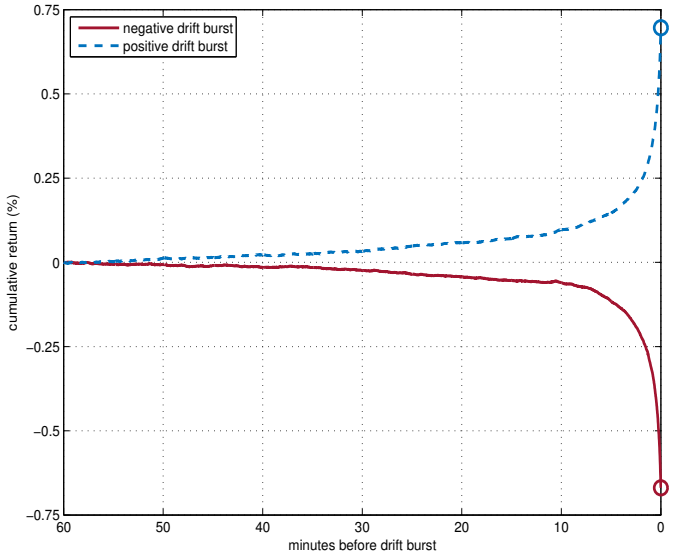
with a likelihood ratio statistic $\Lambda = -2(\ln\mathscr{L}(\hat{\Theta}_{\mathrm{ML}};X)-\ln\mathscr{L}(\Theta_{\mathscr{H}_0};X)) \overset{a}{\sim} \chi^2_1$ (a volatility burst is allowed, since β is unrestricted). In the Online Appendix, based on simulations from the parametric model in Eq. (47), 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 labelled 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 Figure 9, we il-

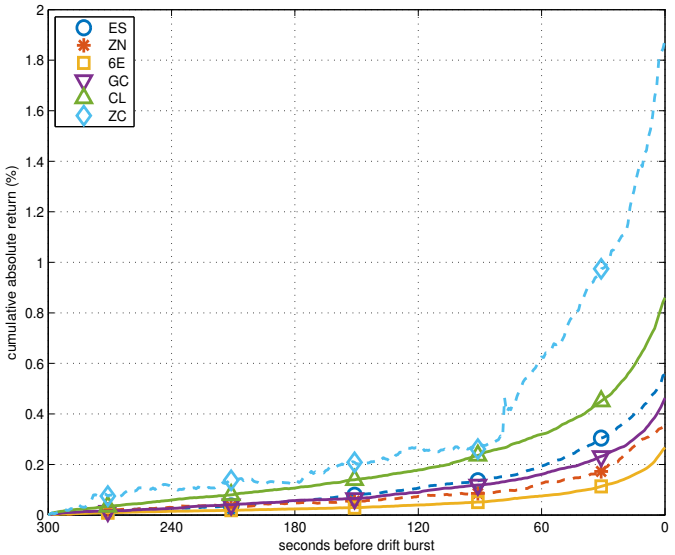
Electronic copy available at: https://ssrn.com/abstract=2842335

Figure 9: Average drift burst log-return.

Panel A: Average log-return by direction



Panel B: Average absolute log-return by asset



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.

lustrate 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 one hour 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.

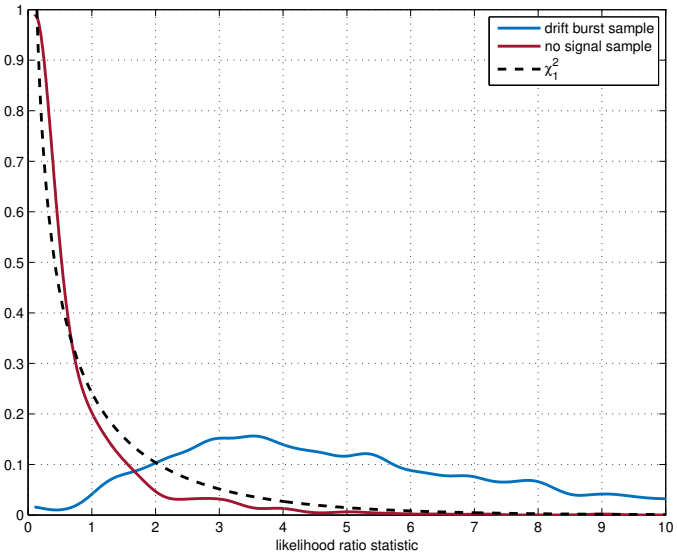
In Panel A of Figure 10, we report the distribution of the likelihood ratio statistic in these two samples. As seen, Λ is close to chi-square distributed in the no signal sample, while it is strongly skewed to the right in the drift burst sample.[25] In Panel B, we show a histogram of $\hat{\alpha}_{\mathrm{ML}}$ 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 border-line 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.

[25] 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.

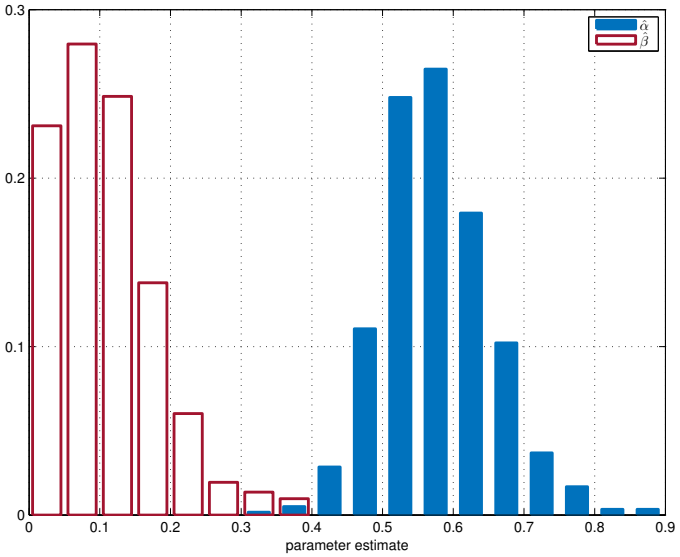
Electronic copy available at: https://ssrn.com/abstract=2842335

Figure 10: Empirical analysis of the parametric drift burst testing procedure.

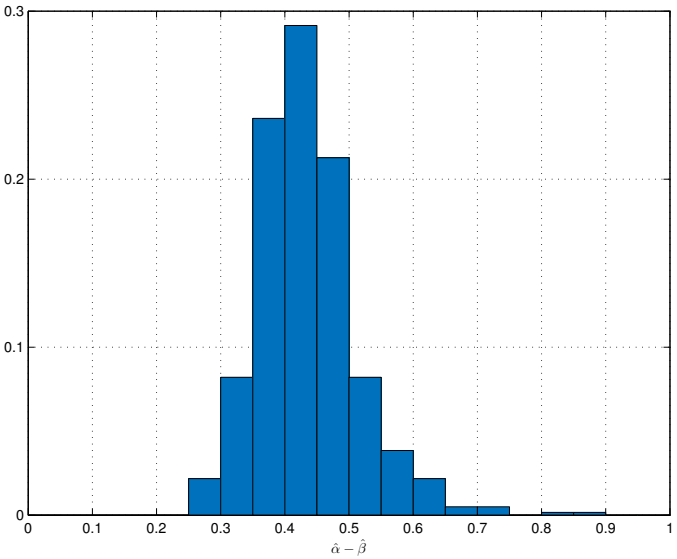
Panel A: Likelihood ratio statistic.



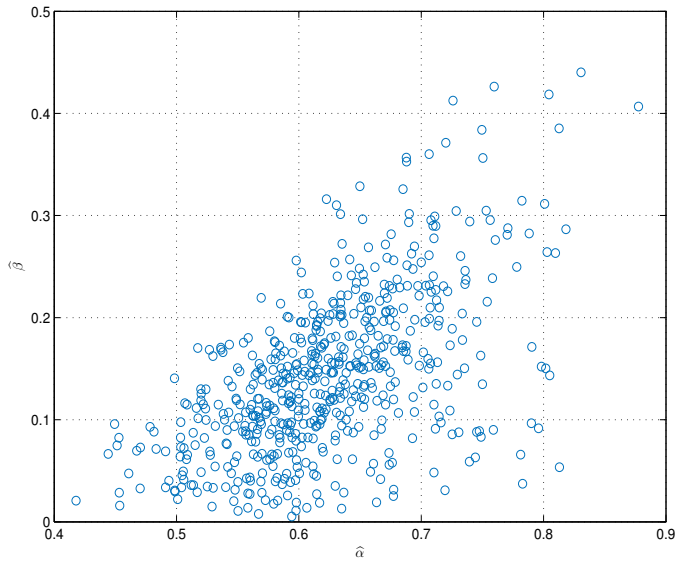
Panel B: $\hat{\alpha}_{\mathrm{ML}}$ and $\hat{\beta}_{\mathrm{ML}}$.



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



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



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^2_1 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 $\Lambda' = -2(\ln \mathscr{L}(\hat{\Theta}_{\mathrm{ML}}; X) - \ln \mathscr{L}(\Theta_{\mathscr{H}_0'}; X)) \overset{a}{\sim} \chi^2_1$. If we apply this to the drift burst sample, $\mathscr{H}_0'$ is discarded at the 5%

Electronic copy available at: https://ssrn.com/abstract=2842335

level for 93.68% of the events, showing that volatility is almost always co-exploding during a drift burst.

References

Aït-Sahalia, Y., and R. Kimmel, 2007, "Maximum likelihood estimation of stochastic volatility models," *Journal of Financial Economics*, 83(2), 413–452.

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

Andersen, T. G., O. Bondarenko, A. S. Kyle, and A. A. Obizhaeva, 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.

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

Bandi, F. M., 2002, "Short-term interest rate dynamics: A spatial approach," *Journal of Financial Economics*, 65(1), 73–110.

Bandi, F. M., and J. R. Russell, 2006, "Separating microstructure noise from volatility," *Journal of Financial Economics*, 79(3), 655–692.

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

———, 2009, "Realized kernels in practice: trades and quotes," *Econometrics Journal*, 12(3), 1–32.

Bates, D. S., 2018, "How crashes develop: Intradaily volatility and crash evolution," *Journal of Finance*, (Forthcoming).

Björk, T., 2003, *Arbitrage Theory in Continuous Time*. Oxford University Press, 2nd edn.

Black, F., 1986, "Noise," *Journal of Finance*, 41(3), 529–543.

Brunnermeier, M. K., and L. H. Pedersen, 2005, "Predatory trading," *Journal of Finance*, 60(4), 1825–1863.

———, 2009, "Market liquidity and funding liquidity," *Review of Financial Studies*, 22(6), 2201–2238.

Campbell, J. Y., S. J. Grossman, and J. Wang, 1993, "Trading volume and serial correlation in stock returns," *Quarterly Journal of Economics*, 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.

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

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

Electronic copy available at: https://ssrn.com/abstract=2842335

Dubinsky, A., M. Johannes, A. Kaeck, and N. J. Seeger, 2018, "Option pricing of earnings announcement risks," *Review of Financial Studies*, (Forthcoming).

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

Golub, A., J. Keane, and S.-H. Poon, 2012, "High frequency trading and mini flash crashes," Working paper, University of Mancester.

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

Hall, P., 1979, "On the rate of convergence of normal extremes," *Journal of Applied Probability*, 16(2), 433–439.

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

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

Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter, 2009, "Microstructure noise in the continuous case: The pre-averaging approach," *Stochastic Processes and their Applications*, 119(7), 2249–2276.

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

Jacod, J., and P. E. Protter, 2012, *Discretization of Processes*. Springer-Verlag, Berlin, 2nd edn.

Johnson, T. C., 2016, "Rethinking reversals," *Journal of Financial Economics*, 120(2), 211–228.

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

Karatzas, I., and S. E. Shreve, 1998, *Methods of Mathematical Finance*. Springer-Verlag, Berlin, 1st edn.

Kaufman, E. E., and C. M. Levin, 2011, "Preventing the next flash crash," *New York Times*, May 5, available at http://www.nytimes.com/2011/05/06/opinion/06kaufman.html.

Kirilenko, A., A. S. Kyle, M. Samadi, and T. Tuzun, 2017, "The Flash Crash: High frequency trading in an electronic market," *Journal of 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., and P. A. Mykland, 2008, "Jumps in financial markets: A new nonparametric test and jump dynamics," *Review of Financial Studies*, 21(6), 2535–2563.

Li, J., V. Todorov, and G. Tauchen, 2015, "Robust jump regressions," *Journal of the American Statistical Association*, 112(517), 332–341.

Madhavan, A. N., 2012, "Exchange-traded funds, market structure and the Flash Crash," *Financial Analysts Journal*, 68(4), 20–35.

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

Massad, T., 2015, "Remarks of Chairman Timothy Massad before the Conference on the Evolving Structure of the U.S. Treasury Market," CFTC Speeches & Testimony, October 21, available at http://www.cftc.gov/PressRoom/SpeechesTestimony/opamassad-30.

Electronic copy available at: https://ssrn.com/abstract=2842335

Menkveld, A. J., and B. Z. Yueshen, 2018, "The Flash Crash: A cautionary tale about highly fragmented markets," *Management Science*, Forthcoming.

Merton, R. C., 1980, "On estimating the expected return on the market: An exploratory investigation," *Journal of Financial Economics*, 8(4), 323–361.

Nagel, S., 2012, "Evaporating liquidity," *Review of Financial Studies*, 25(7), 2005–2039.

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

———, 1994, "Automatic lag selection in covariance matrix estimation," *Review of Economic Studies*, 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," *Journal of Business and Economic Statistics*, 24(2), 195–202.

Podolskij, M., and M. Vetter, 2009a, "Bipower-type estimation in a noisy diffusion setting," *Stochastic Processes and their Applications*, 119(9), 2803–2831.

———, 2009b, "Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps," *Bernoulli*, 15(3), 634–658.

Schaumburg, E., and R. Yang, 2015, "Liquidity during flash events," Liberty Street Economics, Federal Reserve Bank of New York, August 18, available at http://libertystreeteconomics.newyorkfed.org/2015/08/liquidity-during-flash-events.html.

Stoll, H. R., 2000, "Friction," *Journal of Finance*, 55(4), 1479–1514.

Tett, G., 2015, "How humans can wrest control of the markets back from computers," *Financial Times*, October 22, available at https://next.ft.com/content/c5a3c5bc-77fb-11e5-a95a-27d368e1ddf7.

US Treasury, FRB, NY FED, 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.

Vetter, M., 2008, "Estimation methods in noisy diffusion models," Ph.D. thesis, Ruhr-Universität Bochum.

Electronic copy available at: https://ssrn.com/abstract=2842335