The Drift Burst Hypothesis

Kim Christensen Roel Oomen Roberto Renò*

September 2016

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 such as feedback trading. At a theoretical level, we show how to build drift bursts into the continuous-time Itô semi-martingale model in such a way that the fundamental arbitrage-free property is preserved. We then develop a non-parametric test statistic that allows for the identification of drift bursts from noisy high-frequency data. We apply this methodology to a comprehensive set of tick data and show that drift bursts form an integral part of the price dynamics across equities, fixed income, currencies and commodities. We find that the majority of identified drift bursts are accompanied by strong price reversals and these can therefore be regarded as "flash crashes" that span brief periods of severe market disruption without any material longer term price impacts.

*Christensen: Department of Economics and Business Economics, CREATES, Aarhus University, kim@econ.au.dk. Oomen: Deutsche Bank, London and the Department of Statistics, London School of Economics. Renò: Department of Economics, University of Verona, roberto.reno@.univr.it. We are grateful to Frederich Hubalek, Mark Podolskij, Thorsten Rheinlander, and participants at the SoFiE 9th Annual Conference in Hong Kong, XVII Workshop in Quantitative Finance in Pisa, 3rd Empirical Finance Workshop at ESSEC, Paris, and seminars in TU Wien, DCU Dublin, Unicredit, SAFE Frankfurt, University of Venice for helpful comments and suggestions. Christensen received research funding from the Danish Council for Independent Research (DFF – 4182-00050) and was also 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.

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 of the events). In addition to these two dramatic events, there is growing anecdotal evidence 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 – pose two direct challenges to the academic and financial 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 – as we show in this paper – they are not sufficient. Secondly, how can one identify or test for the presence of such features in the data? This paper addresses both these challenges.

A key feature of our approach – which sets it apart from the existing literature – is that we concentrate on the drift component μ_t in the canonical continuous-time Itô semimartingale decomposition for the log-price X_t of a financial asset:

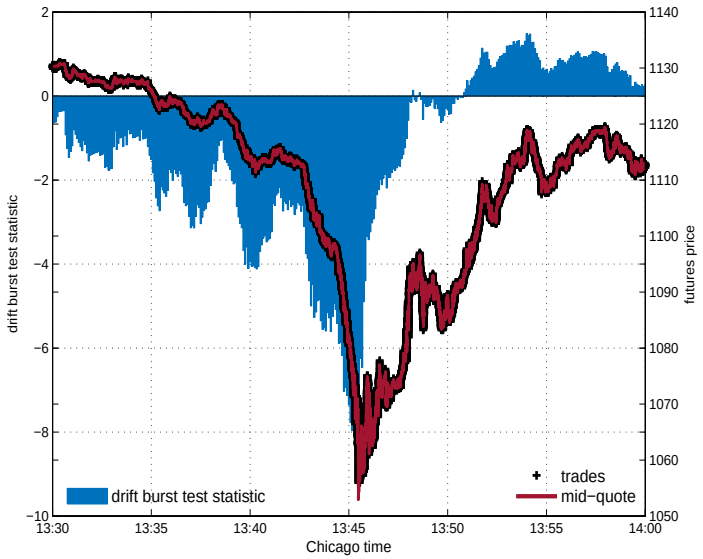
$$\mathrm{d}X_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_t, \tag{1}$$

where σ_t is the volatility and W_t a Brownian motion. In a conventional setup with bounded coefficients, over a vanishing time-interval $\Delta \to 0$, the drift is $O_p(\Delta)$ and swamped by a diffusive component of order $O_p(\sqrt{\Delta})$. For this reason, much of the infill asymptotics is unaffected by the presence of a drift term and the theory therefore invariably neglects it. Also, in many empirical applications, particularly those relying on intra-day data, the drift term is generally small and because estimating it is hard (e.g., Merton, 1980) and adds measurement error into the calculations, the common recommendation is to 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 the *Drift Burst Hypothesis*, which postulates the existence of short-lived locally explosive trends in the price paths of financial assets. We aim to build theoretical and empirical support for the hypothesis and to thereby contribute towards a better understanding of financial market dynamics. We show how drift bursts can be embedded in the traditional continuous-time model in Eq. (1) such that the basic no-arbitrage principle is not violated. A simple economic model is discussed where

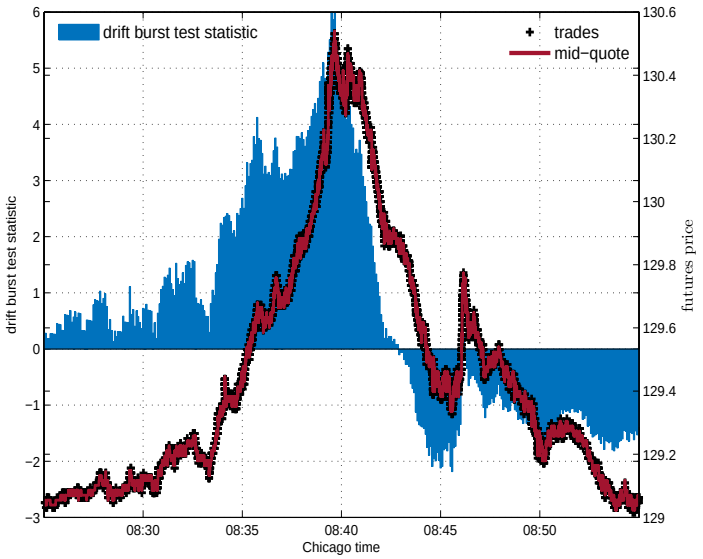
[1] In a Financial Times article, Tett (2015) reports on a speech by the CFTC chairman (Massad, 2015) and writes: "Flash crashes affect even commodities markets hitherto considered dull such as corn". Shortly after the first high profile US equity flash crash, a New York Times article by Kaufman and 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/`. See also a Liberty Street Economics blog of the New York Federal Reserve Bank by Schaumburg and Yang (2015) where liquidity during flash crashes is examined.

Figure 1: The US S&P 500 equity index and treasury market flash crash.

Panel A: S&P 500 equity index.



Panel B: Treasury market.



Note. This figure draws the midquote and traded price (right axis) of the E-mini S&P 500 (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 non-parametric drift burst t-statistic (left axis) proposed in this paper.

drift bursts arise endogenously under mild assumptions via – what can be interpreted as – the interaction of traders which hold different information sets and the presence of a feedback mechanism (Cont and Wagalath, 2013, 2014). Next, we develop a feasible non-parametric 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 is a number of theoretical models of price formation that provide backing for the idea. For instance, Grossman and Miller (1988) consider a crowd 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 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) show that as liquidity demand increases (as measured by traded volume) the price reaction and subsequent reversal grows in magnitude. Alternative mechanisms that can generate price dynamics of this kind include trading frictions as in Huang and Wang (2009), forced liquidation and predatory trading as in Brunnermeier and Pedersen (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 trading volume or return serial correlation. 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. This limits the depth of the empirical work that can be conducted in this framework.

With the drift burst hypothesis in place and the corresponding arbitrage-free 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 non-parametric 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) for the robustification to microstructure noise, we formulate a non-parametric 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 comes to dominate over a burst. When calculated sequentially and using potentially overlapping data, the critical values of the test are determined on the basis of extreme value theory, similar to Lee and Mykland (2008). A simulation study confirms that the test is well

[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.

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 is pivotal 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, 2016), 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 be attributed to the discontinuous 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, 2016) 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: 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%. Quite remarkably, we find that the majority of drift bursts are followed by very strong price reversion. This is consistent with the predictions of the theoretical literature cited above and means that many of the identified episodes resemble (mini) flash crashes that are symptomatic of liquidity shocks or 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 discusses the economic mechanisms that can give rise to drift bursts. Section 4 develops the identification strategy on the basis of noisy high-frequency data. Section 5 includes an extensive simulation study that demonstrates the power of the test. The empirical application is found in Section 6 and Section 7 concludes.

2 The hypothesis

Let $X = (X_t)_{t\geq 0}$ be the logarithmic price of a tradable security. X is defined on a given filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$, satisfying the "usual conditions" of right-continuity and $\mathbb{P}$-completeness. As customary in the literature, we assume X is an Itô diffusion as given by the dynamics in Eq. (1), where $\mu = (\mu_t)_{t\geq 0}$ is a predictable drift, $\sigma = (\sigma_t)_{t\geq 0}$ is an adapted, càdlàg and strictly positive (almost surely) volatility, while $W = (W_t)_{t\geq 0}$ is a standard Brownian motion. As such, the price process features continuous sample paths and the mild regularity conditions imposed on the driving terms μ_t and σ_t allow it to encompass a wide range of dynamic specifications. In particular, the model is compatible with the notion of stochastic volatility, leverage, and possibly jumps in drift and volatility (e.g., Todorov and Tauchen, 2011; Bandi and Renò, 2016). The main exclusion is that of a jump component in X, which is made for ease of exposition only (Theorem 4 in Section 4 shows that our framework is also robust to the presence of jumps in price).

The log-return over time window $[t-\Delta, t]$ is defined as:

$$\Delta X_t = X_t - X_{t-\Delta} \equiv \int_{t-\Delta}^{t} \mu_s \mathrm{d}s + \int_{t-\Delta}^{t} \sigma_s \mathrm{d}W_s. \tag{3}$$

When μ and σ are locally bounded, it follows that

$$\int_{t-\Delta}^{t} \mu_s \mathrm{d}s = O_p(\Delta) \qquad \text{and} \qquad \int_{t-\Delta}^{t} \sigma_s \mathrm{d}W_s = O_p(\sqrt{\Delta}). \tag{4}$$

Thus, as $\Delta \to 0$, it holds that for a continuous price change over a small time-interval Δ, the drift component of ΔX_t

is an order of magnitude smaller than volatility, because $\Delta \ll \sqrt{\Delta}$. The basic version of our model, in which coefficients are bounded, is therefore consistent with the notion that, over short time-intervals, the main contributor to return innovation is volatility. It is this feature that has led the literature to largely neglect the drift.

However, the drift can prevail if it is allowed to diverge in a way that:

$$\int_{t-\Delta}^{t} \mu_s \mathrm{d}s = O_p\big(\Delta^{\gamma_\mu}\big), \tag{5}$$

with $0 < \gamma_\mu < 1/2$. A point in time where μ_t explodes is referred to as the *drift burst time* and is denoted by t_{db}. Comparing Eq. (4) with the assumption in Eq. (5), we see that if σ_t is bounded in a neighborhood of t_{db}, the condition $0 < \gamma_\mu < 1/2$ implies that the drift term prevails over volatility on a vanishing interval of length Δ, so that ΔX_t is now dominated by the drift, not the volatility. The condition $\gamma_\mu > 0$ ensures the continuity of X.

A canonical example of a drift exhibiting a burst is:

$$\mu_t = \frac{1}{|t - t_{\mathrm{db}}|^\alpha}, \tag{6}$$

with $1/2 < \alpha < 1$. Setting $\gamma_\mu = 1 - \alpha$, this formulation is consistent with Eq. (5).

Allowing μ_t to diverge is unconventional. Nevertheless, we can show that the model in Eq. (1), but with a drift satisfying Eq. (5) for some points t_{db}, is still a semimartingale (e.g., Jacod and Protter, 2012).[3] This is necessary – but not sufficient – to exclude arbitrage trading from the market (e.g., Back, 1991; 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 measure (e.g., Theorem 4.1 in Karatzas and Shreve, 1998):

$$\int_{t-\Delta}^{t} \left(\frac{\mu_s}{\sigma_s}\right)^2 \mathrm{d}s < \infty, \tag{7}$$

which is known as a "structural condition." This cannot hold if the drift explodes in the interval $[t-\Delta, t]$, but the volatility remains bounded. It would allow for what is called a "free lunch with vanishing risk," see, e.g., Definition 10.6 in Bjork (2003). Explosive volatility is thus a necessary condition to allow for a drift burst in a market free of arbitrage.

[3] While explosive drift does not impede the semimartingale structure of X, it can on the other hand negatively affect non-parametric estimation of volatility from high-frequency data, as unveiled by Example 3.4.2 in Jacod and Protter (2012). 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 and the hoax tweet of April 23, 2014.

We say there is a volatility burst, if

$$\int_{t-\Delta}^{t} \sigma_s \mathrm{d}s = O_p\big(\Delta^{\gamma_\sigma}\big), \tag{8}$$

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

$$\sigma_t = \frac{1}{|t - t_{\text{db}}|^\beta}, \tag{9}$$

with $0 < \beta < 1/2$. We restrict β to ensure that $\int_{t-\Delta}^{t} \sigma_s^2 \mathrm{d}s < \infty$, so that a stochastic integral can be defined. Here, $\gamma_\sigma = 1-\beta$ and using the Burkholder-Davies-Gundy inequality

$$\int_{t-\Delta}^{t} \sigma_s \mathrm{d}W_s = O_p\big(\Delta^{1/2-\beta}\big). \tag{10}$$

In this model, $\mu_t/\sigma_t \to \infty$ as $t \to t_{\text{tb}}$, for $\alpha > \beta$. Moreover, if $\alpha - \beta < 1/2$, absence of arbitrage follows from the structural condition, as $\int_{t-\Delta}^{t} \left(\frac{\mu_s}{\sigma_s}\right)^2 \mathrm{d}s = O_p\big(\Delta^{1-2(\alpha-\beta)}\big)$. The drift coefficient prevails over the volatility coefficient without arbitrage, for example, if $\alpha = 0.6$ and $\beta = 0.2$.

The drift burst hypothesis can thus be formulated in such a way that it is compatible with absence of arbitrage, which requires the volatility process to explode in conjunction with the drift. Empirically, it is a widely recognised fact that volatility can spike during extreme market conditions. In fact, Kirilenko, Kyle, Samadi, and Tuzun (2016) and Andersen, Bondarenko, Kyle, and Obizhaeva (2015) report highly elevated levels of volatility during the equity flash crash (see also Bates, 2016). But volatility by itself is unlikely to generate rapid and sustained price trends because of the random walk nature of the Brownian motion which tends, probabilistically, to be symmetric. With an exploding drift term, we can generate such sample paths and – depending on the specification of μ_t – the drift burst may be followed by a price that subsequently hovers around its new level (as in a gradual jump) or partially or fully revert towards its original level (as in a flash crash). After establishing the mathematical definition of drift bursts and assessed that they are consistent with no arbitrage, we now turn to a fundamental question: which mechanism can generate a drift burst in a financial market?

3 A simple model featuring drift bursts

To provide some economic intuition and justification for the existence of a locally explosive drift coefficient, we formulate a simple model where drift bursts arise endogenously. It re-enforces the point that drift bursts are a natural and expected outcome of the interaction amongst financial intermediaries. An interesting feature of the

model is that both the drift and the volatility are inflated during a burst, as consistent with Section 2. We build on the setting of Cont and Wagalath (2013, 2014) with distressed selling, but in contrast to their work we focus on the implications for the drift term. In the proposed setup, the transaction price is affected by a random shock, mean reversion and a feedback effect. The model is initially cast in discrete-time, and – as we move to the continuous-time limit – we show that local drift explosions can erupt under suitable conditions on the shape of the price impact function of feedback trading.

We assume that trading takes place at discrete time points $t_i = i/n$ for $i = 1, \ldots, n$ and define $\Delta = 1/n$ as the distance between two consecutive transactions. The efficient log-price $\tilde{X}_t$ starts at $\tilde{X}_0 = 0$ and evolves as a martingale:

$$\tilde{X}_{t_{i+1}} = \tilde{X}_{t_i} + \sigma_{\tilde{X}} \sqrt{\Delta} \tilde{\epsilon}_{t_{i+1}}, \tag{11}$$

where $\tilde{\epsilon}_{t_i} \sim$ i.i.d.$(0,1)$ represent exogenous shocks that change the value of the asset, and $\sigma_{\tilde{X}} > 0$ controls the degree of underlying risk in the economy. The transacted log-price X_t is assumed to follow:

$$X_{t_{i+1}} = X_{t_i} + \underbrace{\sigma\sqrt{\Delta}\epsilon_{t_{i+1}}}_{\text{noise}} + \underbrace{m\big(\tilde{X}_{t_i} - X_{t_i}\big)\Delta}_{\text{mean reversion}} + \underbrace{f\Big(X_{t_i} + \sigma\sqrt{\Delta}\epsilon_{t_{i+1}} + m\big(\tilde{X}_{t_i} - X_{t_i}\big)\Delta\Big) - f(X_{t_i})}_{\text{feedback}}, \tag{12}$$

where $\epsilon_{t_i} \sim$ i.i.d.$(0,1)$, m is an increasing function with $m(0) = 0$, and f is an increasing and concave function also with $f(0) = 0$. We interpret Eq. (12) as follows:

- The first term $\sigma\sqrt{\Delta}\epsilon_{t_i}$ is the contribution of noise trading. It originates from participants that trade for exogenously motivated reasons and do not posses any signal regarding the value of the efficient price $\tilde{X}$. They are viewed as trading randomly thereby moving the transaction price with "force" σ in ways that is unrelated to the evolution of $\tilde{\epsilon}_{t_i}$.[4]

- The second term $m(\tilde{X}_{t_i} - X_{t_i})\Delta$ represents the contribution of informed traders who have knowledge (possibly via a noisy signal) of the efficient price. They buy or sell the security depending on whether the transacted price X_t is below or above the efficient price $\tilde{X}_t$. Their price impact is modelled by an increasing function m, such that $m(0) = 0$. This implies that there is a natural tendency for X_t to revert towards $\tilde{X}_t$, which ensures market efficiency is preserved in the long run.

- The last term captures the effect of feedback trading where directional price moves are re-enforced and mag-

[4] It is possible to allow for ϵ_{t_i} to be correlated with $\tilde{\epsilon}_{t_i}$. This does not change the line of thought behind the main result of this section, as the only difference is that in Eq. (13) B and W are correlated.

nified. This is achieved by letting the function f be increasing (i.e. $f' > 0$) with $f(0) = 0$.[5]

There are various mechanisms in financial markets that can justify the presence of a feedback term in Eq. (12). The first is the liquidity provision theory of Grossman and Miller (1988). As highlighted by Eq. (2), with material one-sided liquidity demand (in our model this corresponds to a large value of the shock $\sigma\sqrt{\Delta}\epsilon_{t_{i+1}}$) market makers that provide immediacy will increase their required compensation by moving X_t in the same direction as the demand. So for instance, with significant selling pressure, market makers will revise their quotes down and as the pressure persists they will do so more aggressively as their risk exposure builds up or the number of market makers diminishes. This can make the transaction price drop below the efficient price and at some stage, informed traders will enter the market and drive the price back towards its fundamental level. As already noted in the introduction, this prediction is consistent with the CFTC and SEC (2010, 2011) report, where it was noticed that during the equity market flash crash, several participants reduced liquidity provision or withdrew entirely from the market in the midst of the turmoil. Subsequently, the market recovered back to its previous levels.

The literature on price formation is also abundant with models which imply feedback. Cont and Wagalath (2013, 2014) note that fire sales – i.e., the sudden deleveraging of large financial portfolios – can be self-reinforcing, leading to a downward spiral in asset prices. Gennotte and Leland (1990) develop a rational expectations model, inspired by the portfolio insurance strategies that were accused of exacerbating the 1987 stock market crash, where hedgers sell in a falling market to prevent further losses, thus creating a snowball effect that increases the initial price drop. Danielsson, Shin, and Zigrand (2012) note that feedback is present, when derivatives on the asset are traded, and option traders with short positions (i.e., negative "gamma") follow delta-hedging strategies to rehedge their risk. Barlevy and Veronesi (2003) show that market crashes can occur irrespective of strong fundamentals, if uninformed traders sell rationally at low prices in an attempt to extract information about the true value of the asset. In the model of forced liquidation and predatory trading by Brunnermeier and Pedersen (2005), strategic traders know that other market participants are in distress (e.g., due to margin calls, stop-loss orders, etc.). The informed trader is then inclined to sell the asset upfront, because he anticipates an opportunity to buy it back at better levels later on. Morris and Shin (2004) study a game-theoretic setup (inspired by bank run models), in which mutually reinforcing selling of short-term traders against a downward sloping demand curve from long-term traders can also result in market crashes.

The following result can be used to show that the inclusion of a feedback term as in by Eq. (12) is capable of generating drift bursts in the continuous-time limit of X_t, provided that f is sufficiently concave.

[5] While we assume that both m and f are time-homogenous, in practice they can depend on different state variables of the market (e.g., volatility or liquidity).

Theorem 1 *Assume that* $E\left(\epsilon_{t_i}^4\right) < \infty$ *and* f *is three times differentiable, bounded and with bounded derivatives. Then, as* $\Delta \to 0$*, the dynamics in Eqs.* (11) – (12) *converge weakly to the processes:*

$$\begin{aligned} \mathrm{d}\tilde{X}_t &= \sigma_{\tilde{X}} \mathrm{d}B_t, \\ \mathrm{d}X_t &= \left(\left[1 + f'(X_t)\right] m\left(\tilde{X}_t - X_t\right) + \frac{\sigma^2}{2} f''(X_t) \right) \mathrm{d}t + \sigma \left[1 + f'(X_t)\right] \mathrm{d}W_t, \end{aligned} \tag{13}$$

where B *and* W *are independent standard Brownian motions.*

Proof See Appendix A. ■

The theorem shows that, while $\tilde{X}_t$ has no drift and constant volatility, in X_t both these coefficients are random. As $f' > 0$, feedback trading is to first-order enhancing the speed of mean-reversion and helping X_t converge to $\tilde{X}_t$. This is because informed trading push in the direction of $\tilde{X}_t$ via $m\left(\tilde{X}_t - X_t\right)$, and this partly determines the feedback contribution as well. In addition, because f too is a function of noise trading, $\sigma\sqrt{\Delta}\epsilon_{t_i}$, the Brownian shocks in X_t are magnified, inducing higher volatility.

The main aspect of Theorem 1 is that an additional second-order effect – proportional to f'' – appears in the drift term. This implies that the drift-to-volatility ratio of X_t is

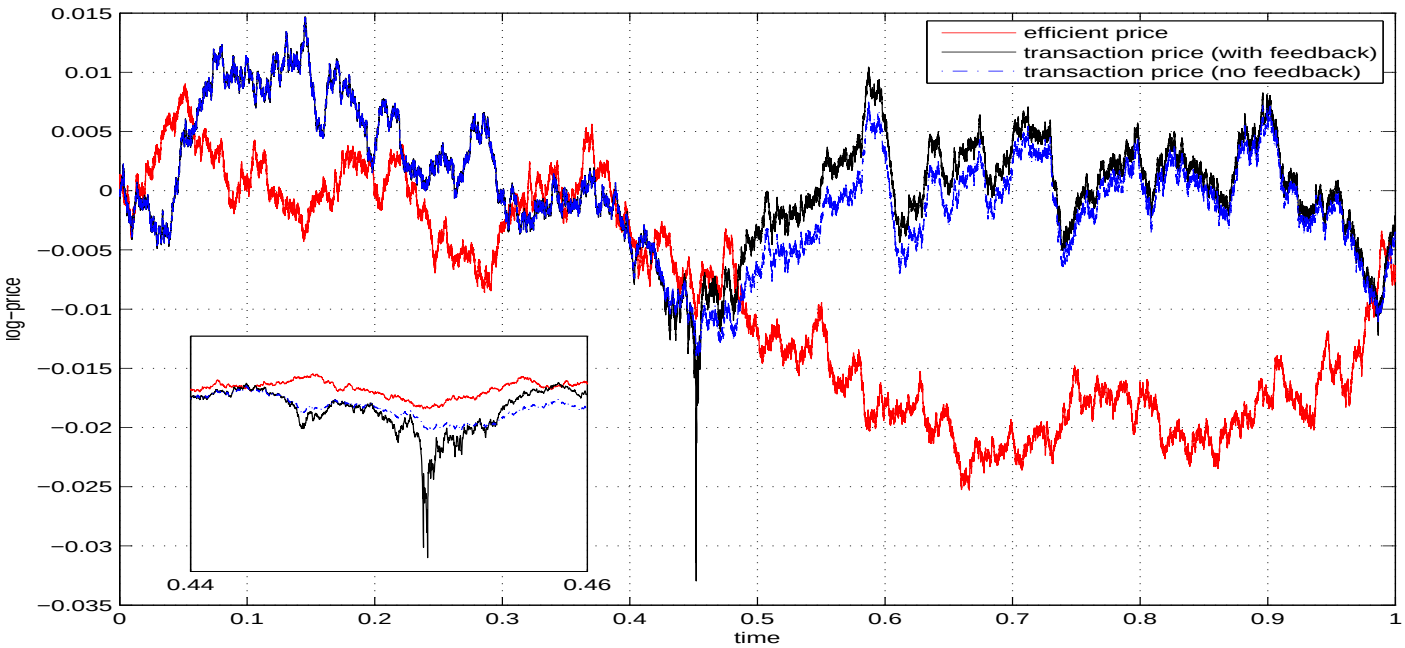
$$\frac{\mu_t}{\sigma_t} = \frac{m\left(\tilde{X}_t - X_t\right)}{\sigma} + \frac{\sigma}{2} \frac{f''(X_t)}{1 + f'(X_t)}. \tag{14}$$

While the contribution of informed trading is typically bounded, the second term can diverge. As consistent with Section 2, the model can therefore exhibit explosions in both μ_t and σ_t in the domain of X_t via f', while $\mu_t/\sigma_t \to \infty$ is a possibility via f''. In Figure 2, we illustrate a sample path of an implementation of the model, which as predicted exhibits a drift burst in the form of a flash crash.[6]

Although stylized, the model captures many features of the price dynamics of interest in this paper. It supports the notion that drift can prevail locally over volatility. As we show, this can be achieved as the limit of a discrete-time process in a setting with asymmetric information, where a subset of market participants feedback on prices. As documented above, there is both theoretical and empirical support for feedback effects in financial markets. If the

[6] Alternatively, we can impose a drift-to-volatility ratio by solving the ODE in Eq. (14) by setting $0.5\sigma f''(x)/[1 + f'(x)] = g(x)$, where $g(x)$ models the drift-to-volatility ratio. The boundary conditions are $f(0) = f'(0) = 0$, and if we further require $g(0) = 0$, then $f''(0) = 0$ and the feedback is negligible in a neighborhood of zero (up to third-order terms in the deviation). The solution to this ODE is analytic for a broad variety of functions. For instance, setting $\sigma = 1$, in the polynomial case with $g(x) = c\,x^\alpha$, $c > 0, \alpha > 0$, we have that $f'(x) = e^{2c\,x^{1+\alpha}} - 1$ and $f''(x) = 2c\,x^\alpha e^{2c\,x^{1+\alpha}}$. This also illustrates that we do not need $g(x)$ to diverge to generate arbitrarily large drift-to-volatility ratio and still get an analytical solution.

Figure 2: A simulated sample path of the feedback model in Eq. (13).



Note. We simulate a model with feedback trading and mean reversion. We set $\sigma_{\tilde{X}} = \sigma = 0.4$ per annum and $f(x) = 5x^5$, $x \leq 0$. Also, $m(x) = x^7$, so mean reversion is small (relative to the drift induced by feedback), if the deviation from the reference point $X_t = 0$ is small, but it prevails as the deviation gets larger. This ensures the stationarity of the process and the efficiency of the market, in the sense that the unconditional expected value of X_t is equal to that of $\tilde{X}_t$.

feedback is strong enough, it leads to a drift burst. It also shows that the mathematical continuous-time definition of the drift burst hypothesis provided in Section 2 is natural. Moreover, as consistent with our empirical results in Section 6, because informed trading eventually dominates the feedback for large deviations between $\tilde{X}_t$ and X_t, the model generates reversals, as also exhibited in Figure 1 above and illustrated here in Figure 2.

4 Identification

We here develop a non-parametric in-fill asymptotic theory that can be applied to detect drift bursts in practice. It exploits the message of Eq. (5), namely if there is a drift burst in X at time t_{db}, log-returns in the vicinity of t_{db} are mostly caused by drift and not volatility. We therefore propose a t-statistic, which compares the ratio of estimates of μ_t and σ_t over time. Later in this section, we prove that our "signal-to-noise" measure uncovers drift bursts in X, if they are present.

We assume X is observed at (potentially irregular) times $0 = t_0 < t_1 < \ldots < t_n = T$, with T fixed. The associated

discretely sampled log-return on $[t_{i-1}, t_i]$ is:

$$\Delta X_{i,n} = X_{t_i} - X_{t_{i-1}}, \qquad \text{for } i = 1, \ldots, n.$$

We borrow from – and extend – existing work within non-parametric kernel-based estimation of the coefficients of diffusion processes to estimate μ_t and σ_t (e.g., Bandi, 2002; Kristensen, 2010). As usual in that branch of the literature, we require a kernel K and a bandwidth h_n. K is assumed to fulfill some weak regularity conditions, which are succinctly listed in Assumption 3 in Appendix A. While most of these are standard and not restrictive, we should point out that we limit attention to the class of so-called left-sided kernels that are backward-looking in time. This avoids inducing a look-ahead bias and allows our t-statistic to be computed in real-time. The bandwidth h_n is a sequence of positive real numbers, such that $h_n \to 0$ and $n h_n \to \infty$, as $n \to \infty$.

We define:

$$\hat{\mu}_t^n = \frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n} \right) \Delta X_{i,n}, \quad \text{for } t \in (0, T]. \tag{15}$$

In absence of a drift burst, it follows that:

$$\sqrt{h_n}\left(\hat{\mu}_t^n - \mu_t\right) \xrightarrow{d} N\left(0, K_2 \sigma_t^2\right), \tag{16}$$

where $K_2 = \int_{-\infty}^{0} K^2(x) \mathrm{d}x$.

As shown by Eq. (16), $\hat{\mu}_t^n$ is unbiased but inconsistent, because its variance explodes as $h_n \to 0$. This appears to rule out detection of drift bursts via $\hat{\mu}_t^n$. On the other hand, if we rescale the left-hand side of Eq. (16) with a consistent estimator of σ_t – and a constant related to K – it appears the right-hand side has a standard normal distribution. It is this key idea that facilitates the construction of a t-statistic that can identify drift bursts, as we prove in Theorem 2 – 3.

To estimate σ_t, we set:

$$\hat{\sigma}_t^n = \left(\frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n} \right) \left(\Delta X_{i,n}\right)^2 \right)^{1/2}, \quad \text{for } t \in (0, T]. \tag{17}$$

Theorem 2 (The null) *Assume that X is a continuous semimartingale as defined by Eq. (1), and that Assumption 1 – 3 in Appendix A hold true for the stochastic coefficients μ_t and σ_t. For every fixed $t \in (0, T]$, we define:*

$$T_t^n = \sqrt{\frac{h_n}{K_2}} \frac{\hat{\mu}_t^n}{\hat{\sigma}_t^n}. \tag{18}$$

As $n \to \infty$, such that $h_n \to 0$ and $nh_n \to \infty$, it holds that

$$T_t^n \xrightarrow{d} N(0,1). \tag{19}$$

Proof See Appendix A. ■

This shows that in absence of a drift burst, the t-statistic has a limiting standard normal distribution.[7] Thus, although it is not possible to consistently estimate μ_t, we can exploit its asymptotic distribution to form an unbiased test of the drift burst hypothesis.

T_t^n has an intuitive interpretation in terms of the log-return over a time-interval h_n. It is approximately the ratio of the part of that log-return due to drift over that due to volatility (this is exact, if K is the indicator function). A large value of the t-statistic thus signals that the realized log-return is mostly induced by drift. As explained above, while a prevailing drift over a time-interval Δ violates absence of arbitrage, it is still perfectly possible (and confirmed by our empirical application) that the drift contribution can prevail over a longer time-interval h_n (recall that $h_n/\Delta \to \infty$). The bandwidth h_n therefore represents, if the t-statistic is large, a time-interval in which the log-return is mostly composed by drift.

To prove that the t-statistic in Eq. (18) is also consistent under the alternative, we assume that both μ_t and σ_t explode at a given rate as in Eq. (6) and (9), but with the drift diverging faster. In this setting, as the next theorem shows, T_t^n goes to infinity.

Theorem 3 (Drift burst alternative) *Assume that X is a continuous semimartingale of the form:*

$$\mathrm{d}X_t = \left(\mu_t + \frac{1}{|t - t_{\mathrm{db}}|^{\alpha}}\right)\mathrm{d}t + \left(\sigma_t + \frac{1}{|t - t_{\mathrm{db}}|^{\beta}}\right)\mathrm{d}W_t, \tag{20}$$

with $0 < \beta < 1/2$ and $\beta < \alpha < 1$, and that Assumption 1 – 3 in Appendix A hold true for the stochastic coefficients μ_t and σ_t. Then, as $n \to \infty$, such that $h_n \to 0$ and $nh_n \to \infty$, it holds that

$$T_{t_{\mathrm{db}}}^n \xrightarrow{p} \infty. \tag{21}$$

Proof See Appendix A. ■

[7] While this statement appears to follow trivially from Eq. (16) – i.e., via application of Slutsky's Theorem – this is not true. In general, we can only use Eq. (16) to deduce Eq. (19), 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. (16) 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. In both these directions, Theorem 2 extends Kristensen (2010).

As noted in Section 2, inference is not compromised by the presence of a jump process in X. Theorem 4 verifies this formally. In it, we study a departure from the null that – in addition to the drift and diffusive volatility components in Eq. (1) – has a jump term. It implies that jumps are immaterial for drift burst detection.

Theorem 4 (Jump alternative) *Assume that X is a jump-diffusion semimartingale of the form:*

$$\mathrm{d}X_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_t + \mathrm{d}J_t, \tag{22}$$

where $\mathrm{d}J_{t_J} = J$ is a random variable expressing a jump size at time $t = t_J$, while $\mathrm{d}J_t = 0$ otherwise. We also require that Assumption 1 – 3 in Appendix A hold true for the stochastic coefficients μ_t and σ_t. Then, as $n \to \infty$, such that $h_n \to 0$ and $nh_n \to \infty$, it holds that

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

Proof See Appendix A. ■

It is not difficult to select a kernel that can discriminate the occurrence of a jump from that of a drift explosion. In particular, the left-sided exponential kernel adopted in this paper has $\sqrt{\frac{K(0)}{K_2}} = 1$, so that $|T^n_{t_J}| \xrightarrow{p} 1$. Thus, our proposed t-statistic is—asymptotically—small under the null (standard normally distributed) and under the jump alternative (equal to one in absolute value), while it is arbitrarily large under the drift burst alternative.

4.1 Robustness to microstructure noise

In practice, we cannot measure the true, efficient log-price from transaction or quotation data, as it is disrupted by multiple layers of "noise" or "friction" (e.g., Black, 1986; Stoll, 2000). The notion of a minimum tick size, for example, implies that price increments are discrete, which contradicts the Gaussian semimartingale description of price formation presented in Eq. (1). A limitation of our analysis above is that it lacks robustness against such market imperfections that operate at the tick level. In this section, we therefore show how to modify our test for drift bursts, so it is resistant to such features of the market microstructure.

To incorporate noise, we assume that:

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

where $(\epsilon_{t_i})_{i=0}^n$ is an error term with $E(\epsilon_{t_i}) = 0$. The difficulty brought by noise is that in order to do inference about drift bursts in X, we are forced to work with the contaminated high-frequency record of Y.

The observed, noisy log-return is:

$$\Delta Y_{i,n} = \Delta X_{i,n} + \Delta \epsilon_{i,n}, \tag{25}$$

where we further require $E\left[(\Delta \epsilon_{t_i})^4\right] < \infty$ and, with a slight abuse of notation, we redefine

$$\hat{\mu}_t^n = \frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) \Delta Y_{i,n}. \tag{26}$$

The additive noise model is standard in the literature. As ϵ has mean zero, the estimator in Eq. (26) is in expectation equal to the one in Eq. (15). The core of the problem is that microstructure noise creates serial correlation in $\Delta Y_{i,n}$ and therefore affects the variance of $\hat{\mu}_t^n$. As microstructure effects are typically perceived to be transitory, it is commonly assumed the noise is covariance stationary, so that it eventually dies out. Still, a large strand of literature suggests the noise process is complicated in practice. In empirical work, it has been noticed that serial dependence in returns often extends beyond the first lag (e.g., Aït-Sahalia, Mykland, and Zhang, 2011), implying the noise is autocorrelated.[8] Also, the variance of the noise is probably time-varying, as it tends to follow a pronounced U-shaped intraday profile (e.g., Bandi and Russell, 2006; Oomen, 2006; Kalnina and Linton, 2008). Morever, a large part of market microstructure theory – based on asymmetric information and strategic trading – suggests noise contributes to price discovery, as it reveals private information about fundamentals, see, for instance, Glosten and Milgrom (1985); Kyle (1985); Stoll (1989); Easley and O'Hara (1992), or Diebold and Strasser (2013). This means noise is, potentially, endogenous (i.e., ϵ and X are dependent). Last, in practice the properties of the noise change a lot, depending on how you sample the data and whether the analysis is done with trade or quote data (e.g., Hansen and Lunde, 2006).

In the noisy setting, we prove formally in Theorem 5 in Appendix A that the leading term in the (diverging) variance of the drift estimator is sourced from the noise. This implies that the t-statistic suggested in Eq. (18) continues to work, provided we update the estimator for σ_t to handle these additional complications. In small samples, however, the actual variance to be estimated is a mixture of the continuous variation part and that due to market microstructure noise (with the latter dominating asymptotically), so we advice using a robust estimator, which handles both. In doing so, we avoid imposing overly strict and unrealistic assumptions on ϵ by using an estimator that adapts naturally to the market environment, thereby keeping with the non-parametric approach of our investigation. That is, following the huge literature on heteroscedasticity and autocorrelation consistent (HAC)

[8] For example, assume that ϵ is independent of X (i.e., exogenous) and i.i.d. with $E(\epsilon_{t_i}^2) = \omega^2$. Then, conditional on σ, $\text{var}(\Delta Y_{i,n}) = \int_{t_{i-1}}^{t_i} \sigma_s^2 \mathrm{d}s + 2\omega^2$, $\text{cov}(\Delta Y_{i,n}, \Delta Y_{i-1,n}) = -\omega^2$, and higher-order autocovariances are zero. While this suffices to capture some of the leading effects of bid-ask bounce on transaction data, such as spurious return variation and negative first-order autocorrelation (e.g., Niederhoffer and Osborne, 1966; Roll, 1984; French and Roll, 1986), the independent i.i.d. noise model is not well-suited for our analysis at the tick frequency, as we argue below.

covariance matrix estimation (e.g., Newey and West, 1987; Andrews, 1991), we propose to set:

$$\hat{\Sigma}_t^n = \sum_{l=-L}^{L} w_n(l)\hat{\gamma}(l), \tag{27}$$

where

$$\hat{\gamma}(l) = \sum_{i=|l|+1}^{n} K\left(\frac{t_{i-1}-t}{h_n}\right)\Delta Y_{i,n} \cdot K\left(\frac{t_{i-|l|-1}-t}{h_n}\right)\Delta Y_{i-|l|,n}, \tag{28}$$

and $w_n(l) = w(l/n)$ is a kernel.[9]

In the above, $\hat{\gamma}(l)$ is the realized lth order autocovariance of $\left(K\left(\frac{t_{i-1}-t}{h_n}\right)\Delta Y_{i,n}\right)_{i=1}^{n}$, while the lag length L determines the number of these to include in the computation of $\hat{\Sigma}_t^n$. L grows with n, so that $L \to \infty$, $n/L \to \infty$ and $h_n/L \to 0$, as $n \to \infty$.

Throughout the paper, we set:

$$w(x) = \begin{cases} 1-6x^2+6|x|^3, & \text{for } 0 \le |x| \le 1/2, \\ 2(1-|x|)^3, & \text{for } 1/2 < |x| \le 1, \\ 0, & \text{otherwise.} \end{cases} \tag{29}$$

The function in Eq. (29) is the Parzen kernel. It has some profound advantages in our framework. Firstly, with the Parzen kernel $\hat{\Sigma}_t^n$ is ensured to be positive (with probability one), so we can always compute the t-statistic. This 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 $\hat{\Sigma}_t^n$ can be computed on the back of the first L lags of the autocovariance function, while more efficient weight functions typically require all n lags. In the high-frequency framework, where n is often large, the latter can be prohibitively slow to compute. In contrast, L is typically small compared to n in practice, rendering our choice of w much less time-consuming.

The robust t-statistic is then computed as:

$$T_t^n = \sqrt{\frac{h_n}{K_2}}\frac{\hat{\mu}_t^n}{\sqrt{\hat{\Sigma}_t^n}}. \tag{30}$$

In Theorem 6 presented in Appendix A, we prove the consistency of $\hat{\Sigma}_t^n$ under the null of bounded coefficients and

[9] In high-frequency estimation of volatility, the estimator in Eq. (27) based on $(\Delta Y_{i,n})_{i=1}^n$ is called a realized kernel. We refer to Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008, 2009) for more details, while noting that we follow their recommendations to compute $\hat{\Sigma}_t^n$ (e.g., they also advocate working with the Parzen kernel).

standard mixing conditions on ϵ, as in Newey and West (1987). This implies the convergence of the modified T_t^n statistic in Eq. (30) to a standard normal distribution also in the presence of microstructure noise.

5 Simulation study

In this section, we adopt a Monte Carlo approach to further explore the t-statistic proposed in Eq. (18) 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 coefficient.

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

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

where W and B are correlated 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 $(\kappa, \theta, \xi, \rho) = (5, 0.0225, 0.4, -0.5)$. 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 (see Figure 5). 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 6 (see Table 2). It corresponds to second-by-second sampling in a 6.5 hours trading session.

The noisy log-price is:

$$Y_{i/n} = X_{i/n} + \epsilon_{i/n}, \quad i = 0, 1, \ldots, n, \tag{32}$$

where $\epsilon_{i/n} \sim N\big(0, \omega_{i/n}^2\big)$ is the noise component. We set $\omega_{i/n} = \gamma \dfrac{\sigma_{i/n}}{\sqrt{n}}$, 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 a scalar that controls the strength of the noise-to-volatility ratio. In this paper, we assume that $\gamma = 0.5$, which amounts to medium contamination, e.g., Christensen, Oomen, and Podolskij (2014).

$\hat{\mu}_t^n$ and $\hat{\Sigma}_t^n$ are constructed from $(\Delta Y_{i,n})_{i=1}^n$ based on Eq. (26) and (27) with a left-sided exponential kernel $K(x) = \exp(-|x|)$, for $x \leq 0$. An automatic lag length algorithm selects a data-driven choice of autocovariances n_{ac} in the computation of $\hat{\Sigma}_t^n$ (see, e.g., Newey and West, 1994; Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2009).[10] The bandwidth for $\hat{\mu}_t^n$ is varied in $h_n = (120, 300, 600)$ seconds. We use a slightly larger bandwidth of $5h_n$ for $\hat{\Sigma}_t^n$ to capture the persistence in volatility and more accurately estimate the microstructure-induced return variation. We then compute T_t^n at $t_i \in (0,1]$, for $i = 1, \ldots, m$, where m is the total number of tests. In our simulations, we record a new value of T_t^n at every 60th transaction update, so that more than 340 tests are run in each sample.[11] As expected, this leads to a multiple comparison problem, which needs to be accommodated. Noting our test is two-sided, we propose to compute in each Monte Carlo replication:

$$T_m^* = \max_{i=1,\ldots,m} |T_{t_i}^n|. \tag{33}$$

As $T_{t_i}^n \xrightarrow{d} N(0,1)$, it follows that a normalized version of T_m^* has a limiting Gumbel distribution, as $m \to \infty$ (e.g., David, 1970; Lee and Mykland, 2008). However, in our setting the choice of K and h_n, coupled with the frequency of the testing times, implies that T_t^n is highly autocorrelated, so the asymptotic extreme value theory is overly conservative in finite samples. As explained in Appendix B, we therefore adopt a simulation-based approach, which accounts for this dependency, to evaluate the critical values for T_m^*.

We create drift and volatility bursts with the parametric model:

$$\mu_t^{\text{db}} = c_1 \frac{\text{sign}(t - t_{\text{db}})}{|t - t_{\text{db}}|^\alpha}, \qquad \sigma_t^{\text{db}} = c_2 \frac{\theta}{|t - t_{\text{db}}|^\beta}, \quad \text{for } t \in [0.475, 0.525], \tag{34}$$

with $t_{\text{db}} = 0.5$. Here, the asset experiences a short-lived flash crash at t_{db}, as consistent with our empirical finding that most of the identified drift bursts are followed by a partial or full recovery.[12] The window $[0.475, 0.525]$ can be interpreted as making the duration of the drift burst last about 20 minutes. The parameters α and β are varied across a broad interval in order to gauge their impact on the size and power of our t-statistic. We examine all combinations of $\alpha = (0.55, 0.65, 0.75)$ and $\beta = (0.1, 0.2, 0.3, 0.4)$.[13] These values are selected to yield bursts of comparable size to what we observe in the real data. In particular, fixing the tuning parameters at $c_1 = 3$ and $c_2 = 0.15$, our choices of α

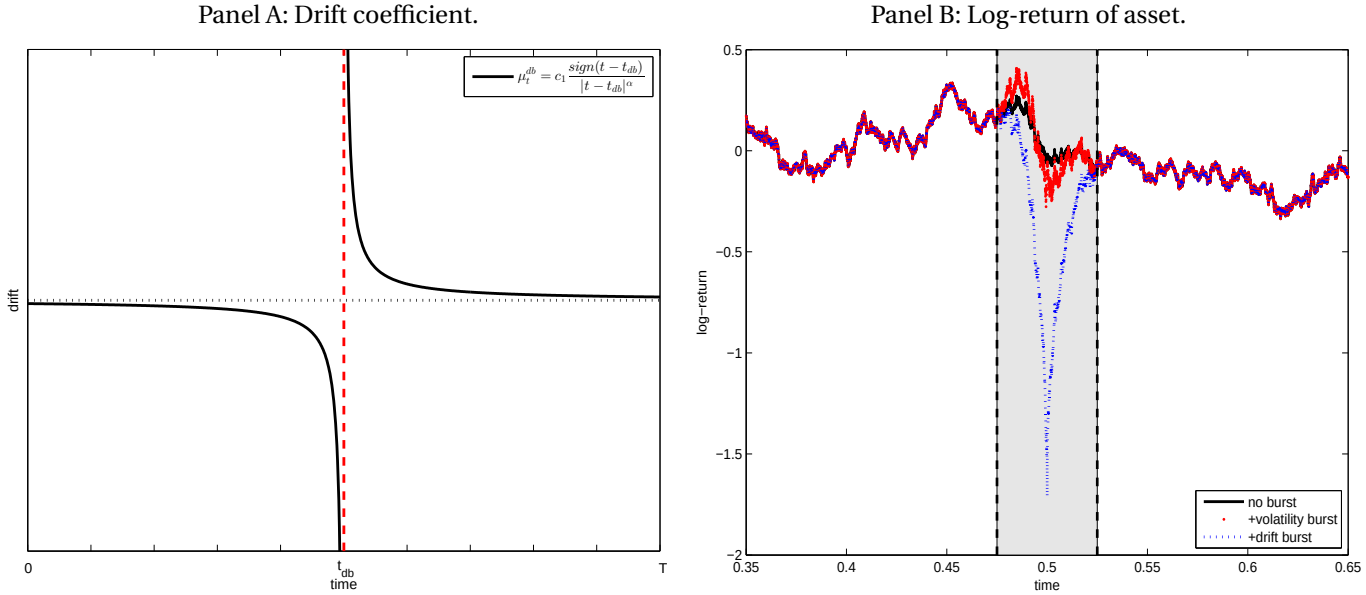
[10] In our numerical experiments, the average value of n_{ac} is 9.3, while its interquartile range is 7 – 12.

[11] The process is started after a full volatility bandwidth of trading time has elapsed to allow for a sufficient number of observations in the construction of T_t^n.

[12] In order to ensure that the log-price also reverts during a pure volatility burst, we recenter the log-return series associated with σ_t^{db} in each simulation, so that $\int_0^T \sigma_t^{\text{db}} \mathrm{d}W_t = 0$. This has almost no impact on the outcome of the t-statistic, but it makes the price processes comparable across settings.

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

Figure 3: Illustration of simulation with a drift burst.



Note. This figure plots a drift burst in our simulated price process. In Panel A, the drift coefficient is shown against time, while Panel B is the evolution of the log-price with a burst in: (i) nothing, (ii) volatility, and (iii) drift and volatility. The latter are based on Eq. (34) with $c_1 = 3$, $c_2 = 0.15$, $\alpha = 0.75$ and $\beta = 0.4$.

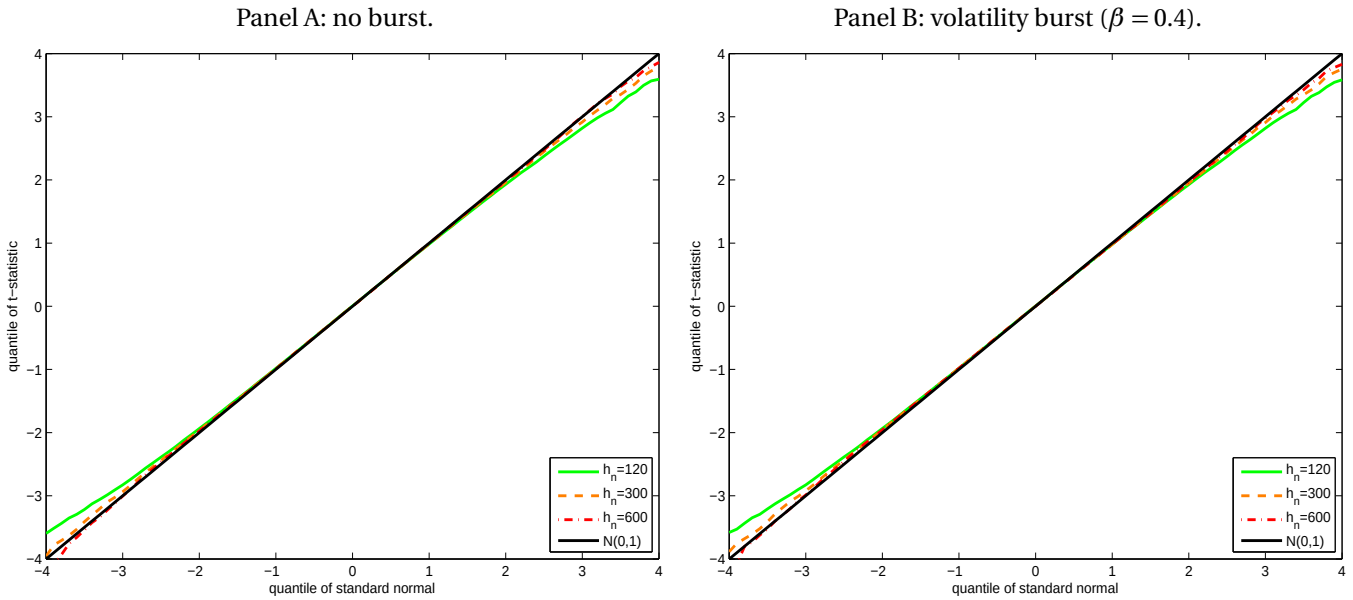
induce a cumulative return $\int_0^{t_{\text{db}}} \mu_t^{\text{db}} \mathrm{d}t$ of about -0.5% (with opposite sign after the crash) for $\alpha = 0.55$ to slightly less than -1.5% for $\alpha = 0.75$. β produces a 25% ($\beta = 0.1$) to more than 100% ($\beta = 0.4$) increase in the average standard deviation of log-returns in the drift burst window relative to its unconditional level across simulations. A drift burst is thus accompanied by often extremely elevated volatility, making it challenging to detect the signal. Figure 3 plots the drift burst function and an illustrative simulation based on the setting with $\alpha = 0.75$ and $\beta = 0.4$.

Figure 4 reports Q-Q plots of the distribution of 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 T_t^n in this setup, although the t-statistic starts to be slightly thin-tailed a few standard deviations out with $h_n = 120$. In Panel B, the outcome of the process featuring no drift burst but a large volatility burst ($\beta = 0.4$) is plotted.[14] While the volatility burst does put some mass further into the tails of the distribution of T_t^n this is hardly noticeable, and the normal continues to be a good approximation also in this setting.

This is further corroborated by Table 1, where we compute the propensity of T_m^* to reject the null hypothesis for the 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). We note that without a volatility burst ($\beta = 0.0$),

[14] As expected, the results for other values of β fall in-between those of Panel A and B and are therefore not reported.

Figure 4: Q–Q plot of T_t^n without drift burst.



Note. We present a Q-Q plot of 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. (34) with $c_2 = 0.15$ and $\beta = 0.4$.

the test is conservative compared to the nominal level if h_n is small, as also reflected in Figure 4. As β increases, T_m^* is mildly inflated yielding a tiny size distortion, but this effect is benign and only present for the largest β and h_n. Otherwise, the test is roughly unbiased. This is critical, as it 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. It is worth noting that, 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. In the empirical analysis, we settle for a 5-minute bandwidth.

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*	0.5		49.5	89.8	100.0	0.3		36.2	79.2	98.7	0.1		30.0	73.7	96.2
	0.1	0.5		30.4	80.4	99.4	0.3		17.1	65.1	96.1	0.1		12.7	57.6	92.7
	0.2	0.6		20.3	71.8	98.8	0.3		9.1	53.9	93.2	0.1		6.4	44.5	87.3
	0.3	0.6		10.9	52.0	94.8	0.3		2.9	30.5	83.4	0.1		1.8	22.3	72.2
	0.4	0.9		5.2	26.6	80.2	0.3		1.2	10.6	52.9	0.1		0.8	6.6	40.6
Panel B: $h_n = 300$																
$\beta =$	0.0*	3.4		58.1	92.4	99.9	0.7		46.8	87.2	99.6	0.4		42.3	84.2	99.3
	0.1	2.8		46.0	87.9	99.7	0.6		33.6	80.0	99.2	0.3		29.4	75.0	98.7
	0.2	2.9		40.3	84.5	99.6	0.6		25.8	73.5	98.8	0.4		21.0	69.2	97.9
	0.3	3.4		29.3	74.4	99.2	0.6		15.4	60.3	97.2	0.5		11.8	54.2	94.8
	0.4	4.1		20.2	53.6	94.8	1.2		9.1	37.1	88.1	0.7		6.6	30.1	82.5
Panel C: $h_n = 600$																
$\beta =$	0.0*	4.7		54.0	89.8	99.8	0.9		41.2	82.0	99.4	0.5		37.4	79.4	98.7
	0.1	4.3		45.4	85.9	99.7	0.7		35.1	76.5	98.7	0.4		30.7	73.8	97.9
	0.2	4.4		42.0	83.0	99.6	0.7		30.0	73.4	98.4	0.3		24.9	69.9	97.1
	0.3	4.9		34.7	75.5	99.1	0.9		21.9	66.1	96.6	0.5		17.2	60.1	95.2
	0.4	6.6		25.2	60.0	96.4	1.9		14.6	46.3	92.3	1.2		11.9	39.9	89.2

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. (33), 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 $\hat{\Sigma}_t^n$ is $5h_n$.

6 Drift bursts in financial markets

We now apply the drift burst test developed above to a large set of intra-day tick data, covering a broad range of financial assets. The aim here is to establish whether drift bursts are an empirical phenomena and – if so – to illustrate some of their basic properties. The analysis also provides an opportunity to examine whether some of the predictions made in theoretical work about liquidity provision, particularly by Huang and Wang (2009), are supported by the data.

6.1 Data

We have available a comprehensive set of tick data – trades and quotes with milli-second timestamps – for futures contracts traded on the Chicago Mercantile Exchange (CME). For each of the main assets classes, we select the most actively traded futures contract, namely the E-mini S&P500 futures (ES) for equities, the 10-Year Treasury Note

Table 2: CME futures data summary statistics

			volume		# quote	inside	sub-sample retained	
code	name	# days	# contracts	notional	updates	spread	by volume	by quotes
ES	E-mini S&P500	1650	1,865,404	$143.0bn	28,950	1.66bps	97.5%	92.4%
ZN	10-Year T-Note	1139	1,118,867	$111.9bn	6,359	1.22bps	95.1%	90.1%
6E	Euro FX	1139	221,465	$34.5bn	49,485	0.77bps	93.4%	88.4%
GC	Gold	1143	148,685	$20.9bn	55,049	1.03bps	89.4%	85.4%
CL	Crude oil	1142	241,131	$17,3bn	74,973	1.58bps	96.0%	92.3%
ZC	Corn	1130	112,281	$2.8bn	5,409	5.52bps	86.5%	69.8%

Note. This table reports for each futures contract, the number of days in the sample, the average daily volume by number of contracts and by 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 Jan 2012 – Jun 2016 for all contracts, except for ES where the sample starts in Jan 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.

futures (ZN) for rates, the Euro FX futures (6E) for currencies, the Gold futures (GC) for precious metals, the Crude oil futures (CL) for energy, and the Corn (ZC) futures for agricultural commodities. The sample period is January 2012 – June 2016 for all futures, except for the S&P500 futures where we backdate the sample to January 2010 in order to capture the widely documented May 2010 flash crash. The contracts selected are amongst the most liquid financial instruments in the world. To illustrate, the average daily notional volume traded in just a single E-mini S&P500 futures contract on the CME is comparable to the trading volume of the entire US cash equity market covering over 5000 stocks traded across more than ten different exchanges.[15] While the CME is open nearly all day, we restrict attention to the European and US trading sessions: from 01:00 – 15:15 Chicago time or 07:00 – 21:15 London time. The only exception here 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 some summary statistics of the data.

6.2 Drift burst identification

The implementation of the drift burst test is done as follows. Using the quote data, we construct for each futures contract a simple mid-price as the average of the best bid and offer price available at any point in time, and then only select those observations where the mid-price changes. Because we operate on the finest granularity tick data, it is inevitable that market microstructure effects are present and if left unaccounted for these can produce substantial biases and a reduction of power in our drift burst test statistic. As discussed above, the local drift estimation proceeds as usual but the estimation of the local price volatility requires a HAC-type adjustment. We use the Parzen kernel with a bandwidth of $L = 10$. Figure 5 provides an illustration of the pronounced microstructure-induced serial

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

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
ES	E-mini S&P500	1.01	1.01	3.2	376	140	48	21	9
ZN	10-Year T-Note	1.11	1.06	2.7	82	34	18	6	2
6E	Euro FX	0.96	0.97	3.8	788	356	162	75	34
GC	Gold	0.92	0.95	4.0	834	332	139	58	20
CL	Crude oil	0.96	0.99	3.9	874	356	152	59	29
ZC	Corn	1.08	1.08	3.3	106	41	14	5	1

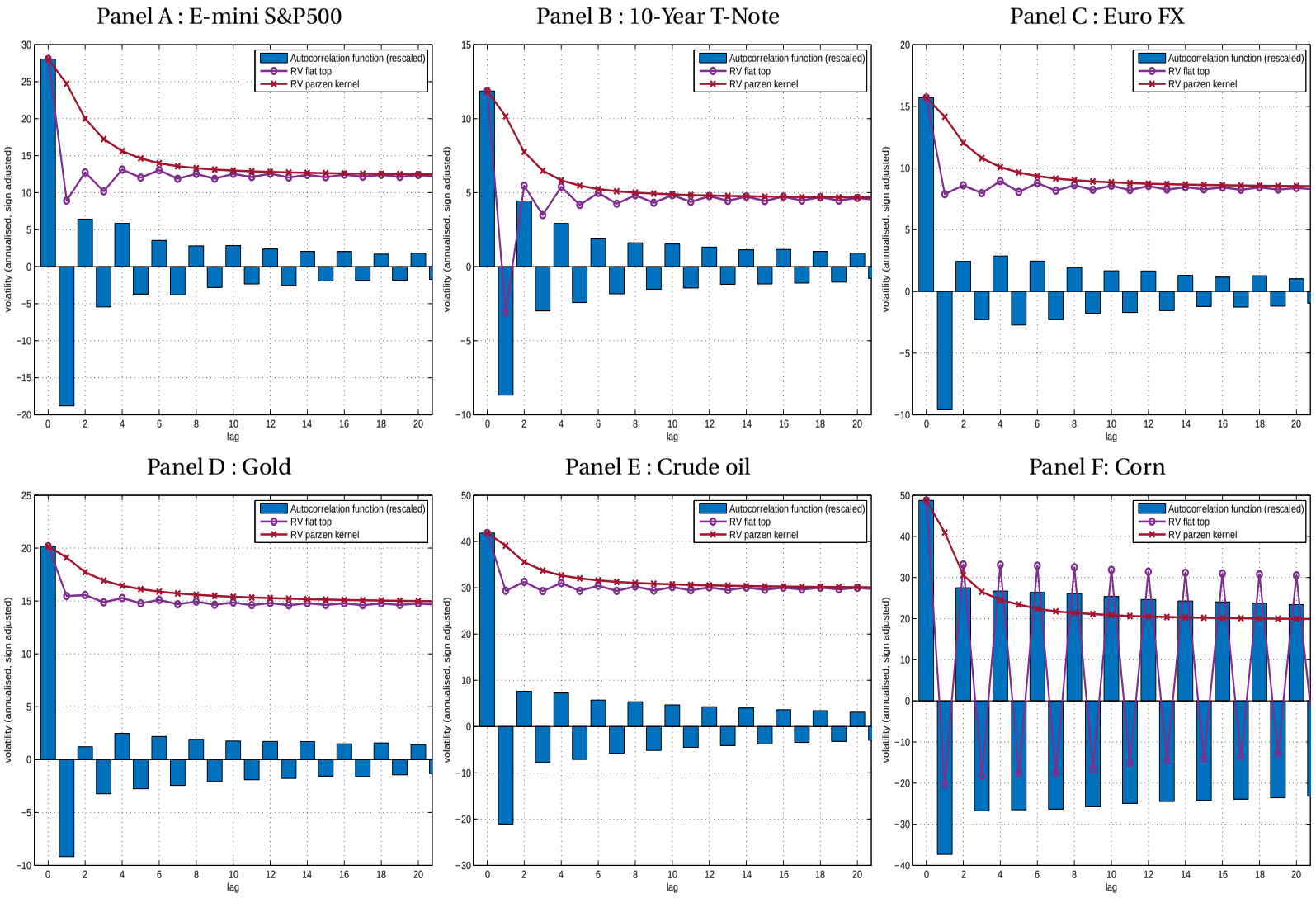
Note. This table reports for each futures contract, the standard deviation and kurtosis of the empirical drift burst test values. The test is calculated at five second intervals across the full sample period, provided that there was a mid-quote update over that interval. The standard deviation is calculated in the usual way ("σ") and 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 also reported. The number of false positives that we expect, which can be computed using the techniques described in Appendix B, is virtually zero.

correlation present in the mid-quote returns. The alternating pattern is due to the way we sample the returns here. Overlaid in this chart are two measures of realized volatility as in Eq. (27) with one that sets $w_n(\cdot) = 1$ and the other uses the Parzen kernel. From here it is clear that with a bandwidth of $L = 10$ the realized variance estimator is largely free of any microstructure biases and that the Parzen kernel has the added advantage that it is guaranteed to produce positive variance estimates. This feature is particularly important when the test is applied to relatively small samples over illiquid parts of the trading day, or to futures contracts where the minimum price increment – and hence the microstructure noise – is relatively large (e.g. ZN).

Because our primary interest lies in identifying any short-lived intra-day drift bursts, we set the bandwidth of the drift estimator to five minutes. To estimate the local spot volatility process, however, we use a longer bandwidth of 25 minutes. This is in recognition that to obtain reliable volatility estimates a reasonable number of data points is required (in contrast to the drift estimator which only obtains marginal efficiency gains from using more data points). Also, because the volatility estimator appears in the denominator of the test statistic, the impact of any measurement error can severely distort the properties of the test. And finally, while volatility is known to vary over time, an extensive literature suggests the process is relatively persistent which enables one to use a local window around the point of interest to make inference on the spot volatility. The simulation experiments above support this choice. To keep computations manageable, we calculate the drift burst test on a regular grid at five-second intervals and then only sample those points that are preceded by a mid-quote change.

Table 3 provides some summary statistics for the calculated drift burst test values over the full sample. Judging by the estimates of the standard deviation and kurtosis, the test appears to be well behaved and in line with the theory which suggests that, asymptotically, the test will have a standard normal distribution under the null hypothesis of

Figure 5: Autocorrelation function and realized variance estimation



Note. This figure draws for each security the autocorrelation function of the sampled tick returns together with the realized kernel-based volatility estimates using a flat (unit weight) kernel and the Parzen kernel with varying lag length. To facilitate comparison, the autocorrelations are multiplied by the square root of the realized variance so that all coincide at lag length 0.

no drift bursts. This is quite remarkable considering that the test here is calculated over short intra-day intervals and across a wide range of asset classes, liquidity conditions, diurnal patterns, and microstructure effects. Concentrating on the tails of the distribution, we identify a large number of significant drift bursts in the data. To account for the rolling calculation of the test statistic and to avoid double counting of significant events, we allow for at most one drift burst to be established at the point where the test statistic attains its local extremum and exceeds a set critical value over any five-minute window. For instance, at a critical value of 4.5, we identify 140 drift bursts in the E-mini S&P500 futures, with the number of expected false positives (computed using extreme value theory as described in Appendix B) being practically zero. That is about one every two weeks. Drift bursts are even more frequent in the Euro FX, Gold, and Oil contracts but much less frequent in the Treasury and Corn futures. Figure 6 provides some examples of significant drift bursts identified by the test. Observing the price evolution over these episodes, it is clear that neither price jumps or bursts in volatility are driving the dynamics. The drift burst hypothesis provides a plausible alternative to model the data.

6.3 Some stylised facts of drift bursts

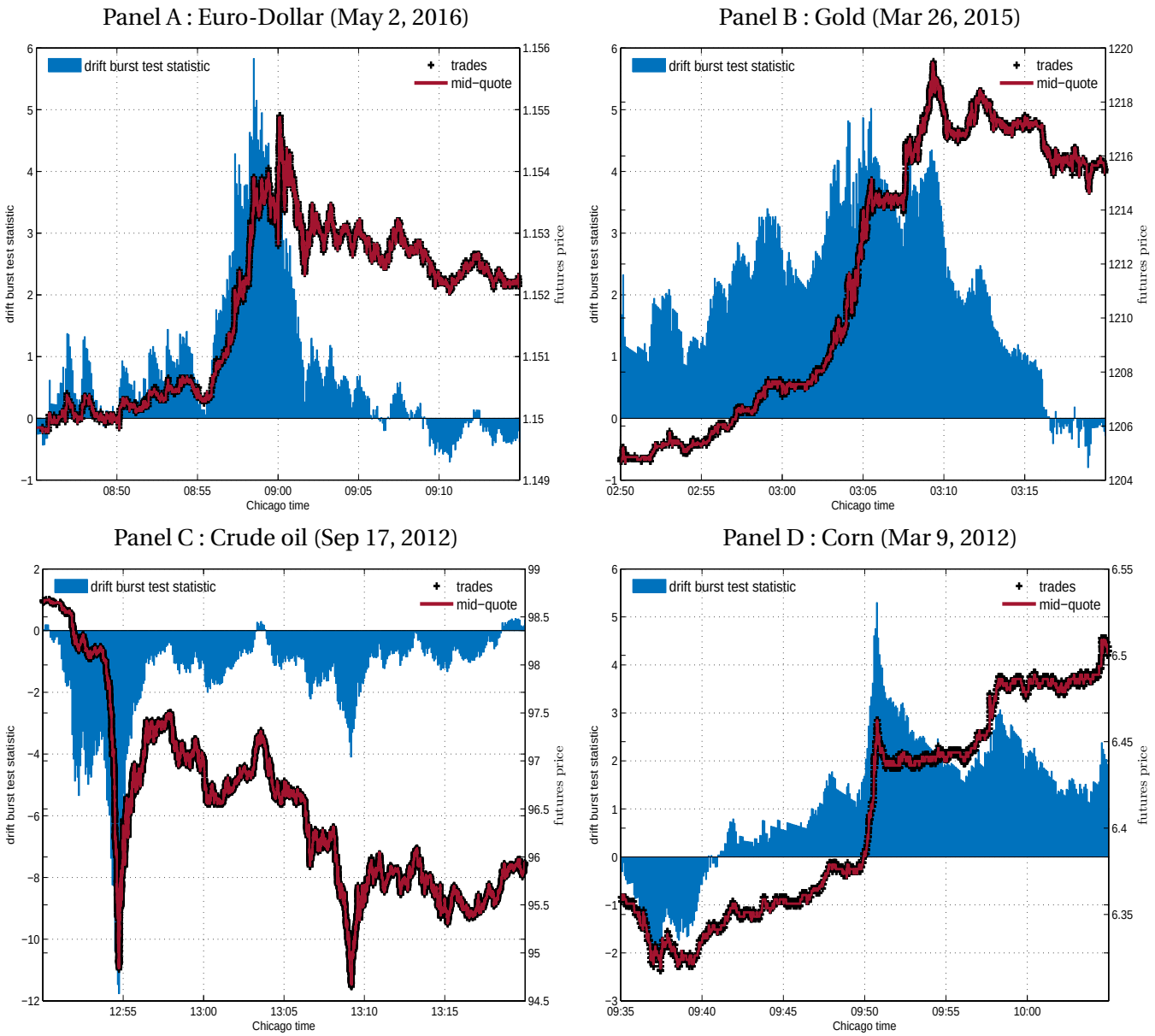
We now investigate in some more detail the price dynamics over a drift burst episode. In particular, we study whether the very strong mean reversion observed over the high profile May 2010 S&P500 flash crash and the October 2014 Treasuries flash crash is more broadly associated with a drift burst. For this purpose, we define the start of a drift burst at the latest point in time (prior to the peak) where the test statistic is less than one in absolute value. The time-interval from start to peak then forms a natural frequency at which to sample the price process post peak. Let $\{t_j\}_j$ denote the set of time points where a drift burst peak is identified. Let τ_j denote the time-interval from start to peak, which is our definition of the duration of the drift burst event. Then define:

$$D_{t_j} = X_{t_j} - X_{t_j-\tau_j} \qquad \text{and} \qquad R_{t_j,k} = X_{t_j+k\tau_j} - X_{t_j}. \tag{35}$$

D_{t_j} is the logarithmic return during the $j-$th drift burst event. $R_{t_j,k}$ is the logarithmic post-drift burst return immediately after the $j-$th event, over a time interval $k\tau_j$, so that we standardize time measurement to the duration of the drift burst event. Figure 7 provides a graphical illustration of the behavior of D_{t_j} and $R_{t_j,k}$ with $k=3$, pooled over all the considered markets. The figure shows that drift bursts can be associated with both positive and negative returns. Most importantly, it clearly indicates that most of drift bursts (75%, in this specific case) are reversals, so that the percentage of "gradual jumps" is small.

To evaluate the magnitude of the reversals more formally, and to evaluate whether drift bursts can be associated

Figure 6: Drift Burst Examples



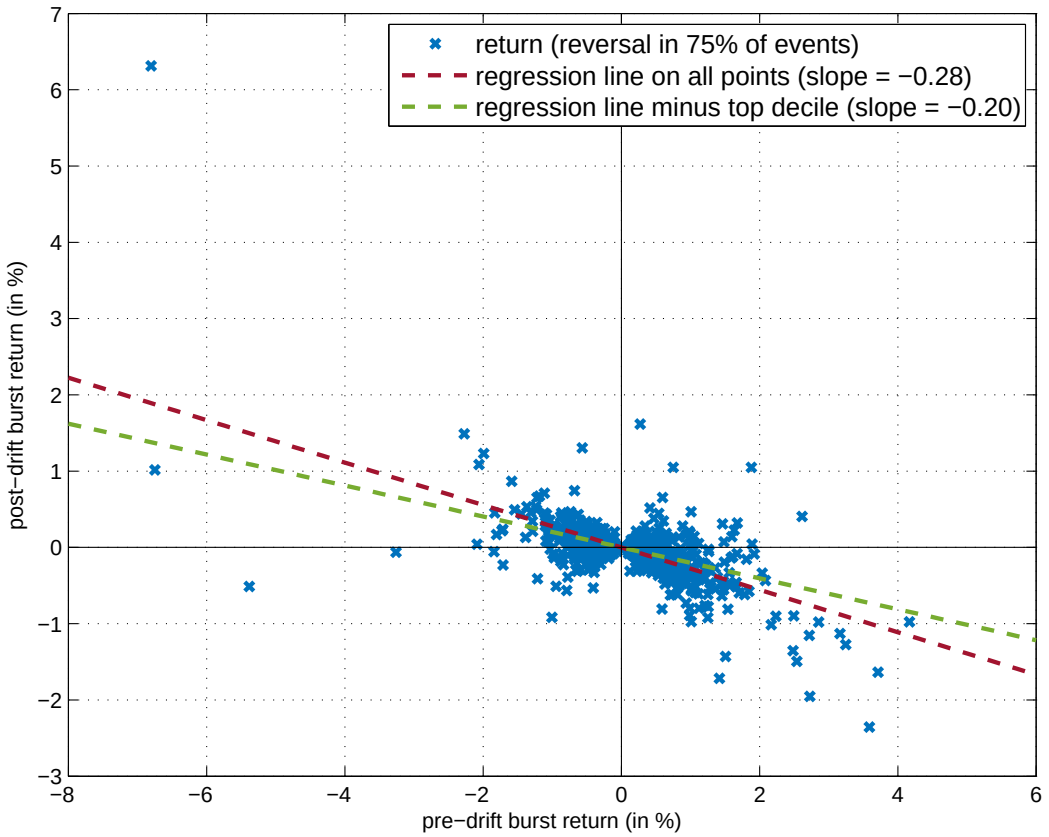
Note. This Figure draws, for some identified drift bursts, the sample path of the mid-quote and traded price (right axis) together with the test statistic (left axis) over a 30-minute window that includes the peak of the drift burst.

with short term, high-frequency, return predictability, we run the following backward-forward regressions:

$$R_{t_j,k} = \alpha + \beta D_{t_j} + \varepsilon_j \tag{36}$$

When $\beta < 0$, then a drift burst tends to be followed by a (partial) retracement of the price. In general, a value of β

Figure 7: Drift burst reversals



Note. This Figure draws on the horizontal axis the percentage price move from the start to the peak of a drift burst (as defined D_{t_j} in Eq. 35) against the subsequent price move (as defined by $R_{t_j,k}$ in Eq. 35) on the vertical axis. The horizon is set to $k = 3$.

significantly different from zero would indicate short term predictability conditional on the drift burst event.

Panel A of Table 4 reports the regression results for each futures contract separately and when all are pooled together for different values of the forward looking horizon post the drift burst peak. Strong mean reversion is evident and this is consistent across all asset classes and horizons. Estimates of the β coefficients are negative and highly significant (even for assets with relatively small number of observations) whereas the intercept is insignificantly different from zero (not reported). The regression R^2 indicates there is a substantial amount of predictive power here. We also report the fraction of reversals (measured by counting the relative occurrence when D_{t_j} and $R_{t_j,k}$ have opposite sign) which rarely drops below 60%.

To check the robustness of these results, we remove the most significant decile of drift bursts episodes and rerun the regression with results reported in Panel B. As expected, the regression R^2 drops, but the same qualitative finding remains overly evident in the data with significant levels of mean reversion and frequent reversals. The critical value

Table 4: Reversion in the drift burst

		horizon $k=1$			horizon $k=3$			horizon $k=5$			horizon $k=10$		
sym	#	β	R^2	$\%R$	β	R^2	$\%R$	β	R^2	$\%R$	β	R^2	$\%R$
Panel A: all drift bursts (critical value = 4.5)													
ES	140	−0.55 (−17.42)	68.6%	75.0%	−0.60 (−16.51)	66.2%	74.3%	−0.59 (−13.49)	56.7%	62.9%	−0.59 (−12.66)	53.6%	62.9%
ZN	34	−0.37 (−5.92)	51.5%	79.4%	−0.40 (−5.33)	46.3%	70.6%	−0.42 (−3.68)	29.1%	67.6%	−0.40 (−2.62)	17.2%	58.8%
6E	356	−0.21 (−16.73)	44.1%	77.0%	−0.24 (−16.39)	43.1%	71.1%	−0.26 (−11.65)	27.7%	66.0%	−0.22 (−8.61)	17.3%	61.0%
GC	332	−0.22 (−13.84)	36.6%	82.2%	−0.23 (−11.95)	30.1%	78.3%	−0.19 (−6.94)	12.7%	71.1%	−0.18 (−5.34)	7.9%	64.5%
CL	356	−0.25 (−21.12)	55.7%	83.4%	−0.26 (−16.73)	44.1%	75.6%	−0.21 (−9.42)	20.0%	66.0%	−0.22 (−8.38)	16.5%	60.4%
ZC	41	−0.17 (−6.26)	49.5%	85.4%	−0.14 (−4.35)	32.1%	75.6%	−0.08 (−1.93)	8.5%	75.6%	−0.20 (−3.29)	21.3%	61.0%
ALL	1259	−0.27 (−34.56)	48.7%	80.3%	−0.28 (−29.32)	40.6%	74.7%	−0.23 (−18.70)	21.7%	67.4%	−0.26 (−18.07)	20.6%	61.9%
Panel B: all drift bursts after removing largest decile													
ES	126	−0.18 (−5.67)	20.5%	73.8%	−0.20 (−5.11)	17.3%	73.0%	−0.10 (−1.95)	2.9%	59.5%	−0.13 (−2.09)	3.4%	61.1%
ZN	31	−0.29 (−7.11)	62.7%	80.6%	−0.28 (−4.91)	44.5%	71.0%	−0.20 (−2.35)	15.6%	67.7%	−0.16 (−1.05)	3.6%	61.3%
6E	320	−0.20 (−15.29)	42.3%	75.3%	−0.19 (−10.62)	26.1%	69.1%	−0.15 (−6.03)	10.2%	64.4%	−0.13 (−4.16)	5.1%	59.1%
GC	299	−0.19 (−11.11)	29.3%	81.6%	−0.20 (−8.19)	18.4%	77.6%	−0.17 (−5.42)	9.0%	70.9%	−0.13 (−3.24)	3.4%	63.9%
CL	320	−0.22 (−17.71)	49.6%	82.8%	−0.21 (−13.28)	35.6%	75.0%	−0.19 (−7.74)	15.8%	65.6%	−0.18 (−5.51)	8.7%	59.4%
ZC	37	−0.22 (−6.67)	55.3%	86.5%	−0.18 (−4.30)	33.9%	75.7%	−0.14 (−2.36)	13.4%	75.7%	−0.21 (−2.97)	19.7%	62.2%
ALL	1133	−0.21 (−28.75)	42.2%	79.4%	−0.20 (−21.10)	28.2%	73.7%	−0.17 (−12.39)	11.9%	66.4%	−0.17 (−9.90)	8.0%	60.8%

Note. This table reports for each security, the number of identified drift bursts (#) using a critical value of 4.5, the estimated slope coefficient β 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 test statistic peaks is opposite in sign compared to that in the run-up. The forward looking horizon $k\tau_j$ is varied. To confirm robustness of the results, Panel B removes the decile of strongest drift bursts as measured by the absolute value of the test statistic.

is set to 4.5 here but qualitatively similar results are obtained for larger more conservative values.

These results are in keeping with the interpretation of reversals as returns earned by market makers from liquidity provisions. Also Nagel (2012) notices that returns on short-term reversals can be interpreted as liquidity signatures, identifying high levels of VIX as a symptom of reduced liquidity supply. In particular, the model of Huang and Wang (2009), which can be seen as a development of the model of Grossman and Miller (1988), formulates a series of theoretical predictions about the strength of the reversal and volume/volatility during the episode, predicting that the reversal should be stronger with high volume and high volatility. We are able to test this prediction conditionally on the detection of drift burst events. Table 5 shows the estimates of the regression in Eq. (36) (with $k=3$) after conditioning on i) the traded volume during the drift burst event (above median/below median);[16] ii) the volatility at the peak, measured with $\hat{\Sigma}_t$ as defined in Eq. (27) (above median/below median). Results are broadly consistent with the theory in Huang and Wang (2009): predictability of the reversals (as measured by a negative value of the estimate of β) is stronger with higher volume and higher volatility. For instance, the strength of reversion is nearly twice as strong for high volume/volatility episodes compared to lower volume/volatility ones. Also the predictability is much stronger in those scenarios with R^2 at around 50% in the "high" regimes while it doesn't exceed 35% across any of the lower regimes. Results are robust to the removal of the largest most extreme drift bursts. For ZN and ZC the results are weaker, which is unsurprising given the relative small number of observations available here. Stronger predictability associated with high volatility is also consistent with Nagel (2012). As such, the findings presented here provide empirical support for the validity and applicability of the above mentioned theories of liquidity provision.

7 Conclusion

We develop mathematical and statistical methods for the modelling and detection of short-term explosive trends – or drift bursts – in financial time series. We show that drift bursts can be modelled using the standard technology of continuos-time finance without invalidating the basic no-arbitrage property. Applying the proposed methodology to a comprehensive set of tick data, we provide unprecedented insights into these potentially disruptive but poorly understood events. We show that the majority of drift bursts are followed by strong price reversals, and can hence

[16] We use a normalised gross trading volume measure defined as follows. We take the gross trading volume in notional value over the drift burst period, and after dividing it by the number of 5-seconds interval during the period, we normalize by the average trading rate at that time of the day. This procedure allows to depurate for time-of-the-day effects. Results with simple gross trading volume are qualitatively very similar.

Table 5: Conditional reversal regressions

	high volume				low volume				high volatility				low volatility			
sym	#	β	R^2	%R	#	β	R^2	%R	#	β	R^2	%R	#	β	R^2	%R
Panel A: all drift bursts (critical value = 4.5)																
ES	70	−0.64 (−17.46)	81.5%	78.6%	70	−0.15 (−2.98)	11.4%	71.4%	70	−0.59 (−14.75)	75.9%	75.7%	70	−0.10 (−1.59)	3.5%	74.3%
ZN	17	−0.38 (−3.73)	46.5%	64.7%	17	−0.35 (−9.17)	84.0%	94.1%	17	−0.38 (−3.94)	49.2%	76.5%	17	−0.31 (−5.54)	65.7%	82.4%
6E	178	−0.24 (−14.37)	53.9%	82.0%	178	−0.15 (−7.59)	24.5%	71.9%	178	−0.22 (−12.04)	45.0%	80.3%	178	−0.17 (−9.28)	32.7%	73.6%
GC	166	−0.25 (−10.18)	38.6%	81.9%	166	−0.18 (−8.94)	32.6%	82.5%	166	−0.22 (−9.08)	33.3%	82.5%	166	−0.21 (−10.99)	42.3%	81.9%
CL	178	−0.29 (−17.56)	63.5%	89.3%	178	−0.18 (−10.66)	39.1%	77.5%	178	−0.26 (−15.37)	57.2%	87.6%	178	−0.14 (−8.66)	29.8%	79.2%
ZC	20	−0.21 (−5.82)	64.0%	90.0%	21	−0.08 (−2.55)	24.5%	81.0%	20	−0.15 (−4.15)	47.5%	85.0%	21	−0.27 (−7.05)	71.3%	85.7%
ALL	629	−0.32 (−29.21)	57.6%	83.1%	630	−0.18 (−18.27)	34.7%	77.5%	629	−0.27 (−24.98)	49.8%	83.3%	630	−0.17 (−14.86)	26.0%	77.3%
Panel B: all drift bursts after removing largest decile																
ES	63	−0.26 (−9.81)	60.8%	77.8%	63	−0.11 (−2.01)	6.1%	71.4%	63	−0.24 (−8.50)	53.8%	76.2%	63	−0.08 (−1.02)	1.6%	73.0%
ZN	15	−0.01 (−0.10)	0.1%	53.3%	16	−0.36 (−9.38)	85.4%	100.0%	15	−0.15 (−2.12)	24.3%	73.3%	16	−0.32 (−5.81)	69.3%	81.3%
6E	160	−0.24 (−13.60)	53.8%	80.0%	160	−0.11 (−5.68)	16.9%	71.3%	160	−0.22 (−11.09)	43.6%	78.1%	160	−0.16 (−8.01)	28.8%	73.1%
GC	149	−0.22 (−8.37)	32.1%	81.9%	150	−0.18 (−7.87)	29.4%	80.7%	149	−0.20 (−7.64)	28.3%	80.5%	150	−0.20 (−10.00)	40.1%	82.0%
CL	160	−0.27 (−15.53)	60.3%	88.8%	160	−0.17 (−9.68)	37.1%	76.9%	160	−0.24 (−13.45)	53.2%	86.9%	160	−0.14 (−8.65)	32.0%	78.8%
ZC	18	−0.21 (−5.49)	64.0%	83.3%	19	−0.10 (−2.90)	31.8%	89.5%	18	−0.16 (−4.11)	49.8%	88.9%	19	−0.28 (−7.08)	73.6%	84.2%
ALL	566	−0.25 (−25.64)	53.8%	81.8%	567	−0.17 (−16.77)	33.2%	77.4%	566	−0.21 (−22.18)	46.5%	82.9%	567	−0.17 (−12.78)	22.4%	76.4%

Note. This table reports for each security and after conditioning on high/low volume and high/low volatility, the number of identified drift bursts (#) using a critical value of 4.5, the estimated slope coefficient β 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 test statistic peaks is opposite in sign compared to that in the run-up. In these regressions, $k = 3$. To confirm robustness of the results, Panel B removes the decile of strongest drift bursts as measured by the absolute value of the test statistic.

be viewed as flash crashes. Rather than being rare events, we find drift bursts and flash crashes are quite common across all major asset classes. As such they form a stylised feature of financial market dynamics. Our results provide support for a number of theoretical prediction made in the market microstructure literature on price formation in markets with trading frictions. Taken together, our methodology and empirical results contribute towards a better understanding of the microstructure dynamics of financial markets and, as such, may help to inform the regulatory policy agenda.

A Mathematical Appendix

Proof of Theorem 1. Under the Assumptions, using Taylor expansion we can write

$$\mathbf{E}\Big[X_{t_{i+1}} - X_{t_i} \mid X_{t_i}\Big] = \Big[m\big(\tilde{X}_{t_i} - X_{t_i}\big) + f'(X_{t_i})m\big(\tilde{X}_{t_i} - X_{t_i}\big) + \frac{\sigma^2}{2} f''(X_{t_i})\Big]\Delta + o_p(1),$$

and

$$\mathbf{E}\Big[\big(X_{t_{i+1}} - X_{t_i}\big)^2 \mid X_{t_i}\Big] = f'(X_{t_i})m\big(\tilde{X}_{t_i} - X_{t_i}\big)\sigma^2\Delta + o_p(\Delta).$$

Then, the result follows along the lines of the proof of Theorem 4.1 in Cont and Wagalath (2013).

Assumption 1 (The model under the null) *The underlying (logarithmic) price is assumed to follow the dynamics in Eq. (1), where the coefficients μ_t, σ_t are bounded adapted processes with càdlàg paths, and σ_t is a.s. strictly positive. Further, given a fixed point $t \in [0,T]$, let $B_\epsilon(t) = [t-\epsilon, t+\epsilon]$, with fixed $\epsilon > 0$, and assume that there exist $\Gamma > 0$, a sequence of stopping times $\tau_m \uparrow \infty$ and constants $C_t^{(m)}$ such that for all m, for $(\omega, s) \in \Omega \times B_\epsilon(t) \cap [0, \tau_m(\omega)[$, and $u \in B_\epsilon(t)$,*

$$E_{u\wedge s}[|\mu_u - \mu_s|^2 + |\sigma_u - \sigma_s|^2] \le C_t^{(m)} |u-s|^\Gamma, \tag{37}$$

where $\mathrm{E}_t[\cdot]$ denotes $\mathrm{E}[\cdot|\mathscr{F}_t]$. The initial value X_0 is measurable with respect to $\mathscr{F}_0$; γ is a constant such that $1/2 < \gamma < 1$.

Assumption 2 (The trading times) *The process X_t is observed $n+1$ times at deterministic instants $0 = t_0 < t_1 < \ldots < t_n = T$, not necessarily equally spaced and with T fixed. We set $\Delta_{i,n} = t_i - t_{i-1}$ and $\overline{\Delta}_n = \frac{T}{n}$ and assume $\max_{i=1,\ldots,n}\{\Delta_{i,n}\} = O(\overline{\Delta}_n)$. The quadratic variation of time up to a given $t \le T$ is defined as $H(t) = \lim_{n\to\infty} H_n(t)$, where*

$$H_n(t) = \frac{1}{\overline{\Delta}_n} \sum_{t_i \le t} \big(\Delta_{i,n}\big)^2. \tag{38}$$

Assuming that the above limit exists, we require that H is Lebesgue-almost surely differentiable in $[0,T]$, with H' such that for some $K \ge 0$ (not depending on i)

$$\left|H'(t_i) - \frac{\Delta_i}{\overline{\Delta}_n}\right| \le K\Delta_i. \tag{39}$$

Assumption 3 (The kernel) *The kernel $K(\cdot)$ is a positive real function defined on the negative real values (left-sided kernel) satisfying the following assumptions:*

(K0) $K(x) = 0$ *for* $x > 0$.

(K1) the kernel is bounded, differentiable with bounded first derivative;

(K2) $\int_{-\infty}^0 K(x)dx = 1$; $\mathbf{K}_2 = \int_{-\infty}^0 K^2(x)dx < \infty$;

(K3) for every positive sequence $g_n \to \infty$, $\int_{-\infty}^{-g_n} K(x)dx \le C g_n^{-\beta}$ for some $\beta > 0$ and $C > 0$ (fast vanishing tails);

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

Lemma 1 (discretization error) *Under Assumptions 1 and 3, as* $n \to \infty$ *and* $h_n \to 0$, *we have for every* $t \in]0, T[$,

$$A_n = \frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) \int_{t_{i-1}}^{t_i} \mu_s ds - \int_0^T \frac{1}{h_n} K\left(\frac{s-t}{h_n}\right) \mu_s ds = O_{a.s.}\left(\frac{1}{nh_n}\right). \tag{40}$$

$$B_n = \frac{1}{\Delta_n h_n} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) \left(\int_{t_{i-1}}^{t_i} \mu_s ds\right)^2 - \int_0^T \frac{1}{h_n} K\left(\frac{s-t}{h_n}\right) \mu_s^2 ds = O_{a.s.}\left(\frac{1}{nh_n}\right), \tag{41}$$

and the same applies replacing μ_t *with* σ_t.

Proof Write:

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

Using the mean value theorem and the boundedness of $K'(\cdot)$ and the drift process, for each interval there exists ξ_{i-1} such that

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

The proof for the term B_n is similar.

Lemma 2 (localization error) *Under Assumptions 1 and 3, as* $n \to \infty$ *and* $h_n \to 0$, *we have for every* $t \in]0, T[$,

$$B_n = \int_0^T \frac{1}{h_n} K\left(\frac{s-t}{h_n}\right) \mu_s ds - \mu_{t^-} = O_p\left(h_n^{\Gamma/2} + h_n^{\beta}\right), \tag{42}$$

and the same applies replacing μ_t *with* g_t *or* σ_t.

Proof Notice that, by the properties of the kernel,

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

so that we can write

$$\begin{aligned} |B_n| &= \left| \int_0^T \frac{1}{h_n} K\left(\frac{s-t}{h_n}\right) (\mu_s - \mu_{t^-}) ds + \mu_{t^-} \int_{-\infty}^{-t/h_n} K(x) dx \right| \\ &\leq \int_0^T \frac{1}{h_n} K\left(\frac{s-t}{h_n}\right) |\mu_s - \mu_{t^-}| ds + C h_n^{\beta} \end{aligned}$$

where we used the property $(K3)$. Now, by Jensen inequality and property (37):

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

so that for the first term, using property (K4), we have:

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

which concludes the proof.

Proof of Theorem 2. Write:

$$\begin{aligned}\widehat{T}_t^n&=\sqrt{\frac{h_n}{K_2}}\frac{\left(\frac{1}{h_n}\sum_{i=1}^nK\left(\frac{t_{i-1}-t}{h_n}\right)\Delta X_i-\mu_{t^-}\right)}{\sqrt{\frac{1}{h_n}\sum_{i=1}^nK\left(\frac{t_{i-1}-t}{h_n}\right)(\Delta X_i)^2}}+\sqrt{\frac{h_n}{K_2}}\frac{\mu_{t^-}}{\sqrt{\frac{1}{h_n}\sum_{i=1}^nK\left(\frac{t_{i-1}-t}{h_n}\right)(\Delta X_i)^2}}\\&=\underbrace{\sqrt{\frac{h_n}{K_2}}\frac{\left(\widehat{\mu}_n(t)-\mu_{t^-}\right)}{\sqrt{\widehat{\sigma}_n^2(t)}}}_{T_1}+\underbrace{\sqrt{\frac{h_n}{K_2}}\frac{\mu_{t^-}}{\sqrt{\widehat{\sigma}_n^2(t)}}}_{T_2}\end{aligned}$$

Using the results in Mancini, Mattiussi, and Renò (2015) we have $\widehat{\sigma}_n^2(t)\xrightarrow{p}\sigma_{t^-}^2$, which implies, given the boundedness of μ_t and σ_t,

$$|T_2|\leq C\sqrt{h_n}\frac{1}{\sqrt{\widehat{\sigma}_n^2(t)}}=O_p\left(\sqrt{h_n}\right).$$

Now, using Lemmas 1 and 2, we can write

$$\begin{aligned}\sqrt{h_n}\left(\widehat{\mu}_n(t)-\mu_{t^-}\right)&=\sqrt{h_n}\left(\frac{1}{h_n}\sum_{i=1}^nK\left(\frac{t_{i-1}-t}{h_n}\right)\Delta X_i-\frac{1}{h_n}\sum_{i=1}^nK\left(\frac{t_{i-1}-t}{h_n}\right)\int_{t_{i-1}}^{t_i}\mu_s\,ds\right)\\&\quad+O_p\left(\frac{\sqrt{h_n}}{nh_n}+h_n^{\Gamma/2+1/2}+h_n^{\beta+1/2}\right)\\&=\underbrace{\frac{1}{\sqrt{h_n}}\sum_{i=1}^nK\left(\frac{t_{i-1}-t}{h_n}\right)\int_{t_{i-1}}^{t_i}\sigma_s\,dW_s}_{G_n}+O_p\left(\frac{\sqrt{h_n}}{nh_n}+h_n^{\Gamma/2+1/2}+h_n^{\beta+1/2}\right)\end{aligned}$$

Write $\mathscr{G}_n=\sum_{i=1}^nu_{i,n}$. To prove stable convergence of $\mathscr{G}_n$ to $\mathscr{MN}(0,V_t)$, we use Theorem IX.7.28 in Jacod and Shiryaev (2003),

wich states the following sufficient conditions:

$$\sum_{i=1}^{n} E_{i-1}\left[u_{i,n}\right] \xrightarrow{p} 0 \tag{43}$$

$$\sum_{i=1}^{n} E_{i-1}\left[u_{i,n}^2\right] \xrightarrow{p} V_t \tag{44}$$

$$\sum_{i=1}^{n} E_{i-1}\left[u_{i,n}^4\right] \xrightarrow{p} 0 \tag{45}$$

$$\sum_{i=1}^{n} E_{i-1}\left[u_{i,n}\Delta Z_i\right] \xrightarrow{p} 0 \tag{46}$$

where $Z_t = W_t$ or $Z_t = W_t'$ with W_t' orthogonal to W_t. Condition (43) is immediate. From Itô's lemma:

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

so that

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

Now the first term converges to $\mathbf{K}_2\sigma_{t^-}^2$, as in Mancini, Mattiussi, and Renò (2015). The second term is negligible since, given condition (37),

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

To deal with condition (45), we notice that by Burkholder-Davis-Gundy inequality and the boundedness of σ_s that

$$E_{i-1}\left[\left(\int_{t_{i-1}}^{t_i} \sigma_s dW_s\right)^4\right] \leq C\left(\Delta_n\right)^2$$

which implies

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

$$=O\left(\frac{\Delta_n}{h_n}\right).$$

Finally, for condition (46), when $Z = W$ we have, using Cauchy-Schwartz inequality,

$$\begin{aligned} E_{i-1}\left[\Delta_i W \int_{t_{i-1}}^{t_i} \sigma_s dW_s\right] &\leq \sqrt{E_{i-1}\left[(\Delta_i W)^2\right]}\sqrt{E_{i-1}\left[\left(\int_{t_{i-1}}^{t_i} \sigma_s dW_s\right)^2\right]} \\ &\leq \sqrt{\Delta_n}\sqrt{E_{i-1}\left[\int_{t_{i-1}}^{t_i} \sigma_s^2 ds\right]} \\ &\leq C\sqrt{\Delta_n} \end{aligned}$$

which implies

$$\sum_{i=1}^{n} E_{i-1}\left[u_{i,n}\Delta W_i\right] \leq C\sqrt{h_n}\frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-t}{h_n}\right)\Delta_n \to 0.$$

When instead $Z = W'$, orthogonality implies that $W'\int \sigma_s dW_s$ is also a martingale, so that

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

We thus proved that $\mathscr{G}_n$ converges, stably in law, to $\mathscr{MN}(0, \mathbf{K}_2\sigma_{t^-}^2)$, which implies $\widehat{T}_n(t) \Rightarrow \mathscr{N}(0,1)$.

Proof of Theorem 3 Without loss of generality, we can set $T = t_{db} = 1$. Then we have:

$$\hat{T}_t^n = \sqrt{\frac{h_n}{K_2}}\frac{\frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-t}{h_n}\right)\Delta X_{i,n}}{\left(\frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-t}{h_n}\right)\left(\Delta X_{i,n}\right)^2\right)^{1/2}}.$$

where

$$\Delta X_{i,n} = \underbrace{\int_{i-1}^{i} \mu_s ds + \int_{i-1}^{i} \sigma_s dW_s}_{\widetilde{\Delta X_{i,n}}} + \underbrace{\int_{i-1}^{i} \frac{1}{(1-s)^\alpha} ds}_{D_{i,n}} + \underbrace{\int_{i-1}^{i} \frac{1}{(1-s)^\beta} dW_s}_{S_{i,n}}.$$

Write:

$$\hat{T}_t^n = \sqrt{\frac{h_n}{K_2}}\frac{\frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-t}{h_n}\right)\left(\widetilde{\Delta X_{i,n}} + D_{i,n} + S_{i,n}\right)}{\left(\frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-t}{h_n}\right)\left(\Delta X_{i,n}\right)^2\right)^{1/2}}.$$

From Theorem 2, we have $\frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-t}{h_n}\right)\widetilde{\Delta X_{i,n}} = O_p(1/\sqrt{h_n})$. For the second term in the denominator, using the same

derivation of Lemma 1, we have

$$\frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-t}{h_n}\right) D_{i,n} = \frac{1}{h_n}\int_0^1 K\left(\frac{s-1}{h_n}\right)\frac{1}{(1-s)^{\alpha}}\,ds + O_{a.s.}\left(\frac{1}{nh_n}\right)$$
$$= \int_{-1/h_n}^{0} K(x)|x|^{-\alpha} h_n^{-\alpha}\,dx + O_{a.s.}\left(\frac{1}{nh_n}\right) = O_{a.s.}\left(h_n^{-\alpha}\right)$$

For the third term in the numerator, we have, $\frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-1}{h_n}\right) E\left[S_{i,n}\right] = 0$, and

$$\frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-1}{h_n}\right) E\left(S_{i,n}^2\right) = \frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-1}{h_n}\right)\int_{t_{i-1}}^{t_i}\frac{1}{(1-s)^{2\beta}}\,ds = O_p(h_n^{-2\beta}),$$

and, using Burkholder-Davies-Gundy and Jensen inequalities,

$$\frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-1}{h_n}\right) E\left[S_{i,n}^4\right] \leq C\,\frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-1}{h_n}\right)\left(\int_{t_{i-1}}^{t_i}\frac{1}{(1-s)^{2\beta}}\,ds\right)^2$$
$$\leq C\,\frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-1}{h_n}\right)\Delta_n\int_{t_{i-1}}^{t_i}\frac{1}{(1-s)^{4\beta}}\,ds = O_p(\Delta_n h_n^{-4\beta}).$$

These result, invoking again Theorem IX.7.28 in Jacod and Shiryaev (2003), imply that $\frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-1}{h_n}\right) S_{i,n} = O_p(h_n^{-\beta})$, so that, since $\alpha > \beta$, the numerator is $O_p(h_n^{-\alpha})$. Similarly, the leading term in the denominator is:

$$\frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-t}{h_n}\right) D_{i,n}^2 = O_p(\Delta_n h_n^{-2\alpha}),$$

where we used a reasoning similar to that leading to Lemma 1, second part. We can thus conclude that $\hat{T}_t^n(1) = O_p(\sqrt{nh_n}) \to \infty$.

Proof of Theorem 4 Without loss of generality, we can set $T = t_J = 1$. Write

$$\widetilde{\Delta X_{i,n}} = \int_{i-1}^{i}\mu_s\,ds + \int_{i-1}^{i}\sigma_s\,dW_s.$$

Then we have, using Theorem 2:

$$\hat{T}_t^n = \sqrt{\frac{h_n}{K_2}}\,\frac{\frac{1}{h_n}K\left(\frac{\Delta_n}{h_n}\right)J + \frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-1}{h_n}\right)\widetilde{\Delta X_{i,n}}}{\left(\frac{1}{h_n}K\left(\frac{\Delta_n}{h_n}\right)J^2 + \frac{2}{h_n}K\left(\frac{\Delta_n}{h_n}\right)J\widetilde{\Delta X_{n,n}} + \frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-1}{h_n}\right)\left(\widetilde{\Delta X_{i,n}}\right)^2\right)^{1/2}}$$
$$= \frac{\sqrt{\frac{1}{h_n K_2}}K\left(\frac{\Delta_n}{h_n}\right)J + \mathcal{N}(\mu_t,\sigma_t^2) + o_p(1)}{\left(\frac{1}{h_n}K\left(\frac{\Delta_n}{h_n}\right)J^2 + \frac{2}{h_n}K\left(\frac{\Delta_n}{h_n}\right)JO_p(\sqrt{\Delta_n}) + \sigma_t^2 + o_p(1)\right)^{1/2}} \xrightarrow[n\to\infty]{p} \sqrt{\frac{K(0)}{K_2}}\cdot\text{sign}(J)$$

Theorem 5 (The drift estimator with market microstructure noise) *Assume that X is a continuous semimartingale as defined by Eq. (1), and that Assumption 1 – 3 in Appendix A hold true for the stochastic coefficients μ_t and σ_t. Define $Y_{i,n} = X_{i,n} + \epsilon_{i,n}$, and, for every fixed $t \in (0,T]$, with a slight abuse of notation, $\hat{\mu}_t^n$ as in Eq. (26). Assume $E(\Delta\epsilon_{t_i}) = 0$, $E((\Delta\epsilon_{t_i})^2) = V_{\Delta\epsilon} < \infty$ and $E((\Delta\epsilon_{t_i})^4) < \infty$. As $n \to \infty$, such that $h_n \to 0$ and $nh_n \to \infty$, it holds that*

$$\sqrt{\Delta_n h_n}\hat{\mu}_t^n \xrightarrow{d} \mathcal{N}(0, K_2 V_{\Delta\epsilon}). \tag{47}$$

Proof Using Theorem 2, write:

$$\hat{\mu}_t^n - \mu_t = \frac{1}{\sqrt{h_n}}\mathcal{N}(0, K_2\sigma_t^2) + o_p(1) + \underbrace{\frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-t}{h_n}\right)\Delta\epsilon_{i,n}}_{M_n^\epsilon}.$$

It is straightforward that $\sqrt{\Delta_n h_n}\frac{1}{h_n}\sum_{i=1}^n K\left(\frac{t_{i-1}-t}{h_n}\right)E_{i-1}[\Delta\epsilon_{i,n}] = 0$. Now write:

$$(h_n\Delta_n)\frac{1}{h_n^2}\sum_{i=1}^{n} K^2\left(\frac{t_{i-1}-t}{h_n}\right)E_{i-1}[(\Delta\epsilon_{i,n})^2] = \frac{1}{h_n}\sum_{i=1}^{n} K^2\left(\frac{t_{i-1}-t}{h_n}\right)V_{\Delta\epsilon}\Delta_n \underset{n\to\infty}{\longrightarrow} K_2 V_{\Delta\epsilon}.$$

Finally,

$$(h_n\Delta_n)^2\frac{1}{h_n^4}\sum_{i=1}^{n} K^4\left(\frac{t_{i-1}-t}{h_n}\right)E_{i-1}[(\Delta\epsilon_{i,n})^4] \underset{n\to\infty}{\longrightarrow} 0.$$

This proves that $\sqrt{\Delta_n h_n}M_n^\epsilon \xrightarrow{d} \mathcal{N}(0, 2(V_\epsilon - AC_\epsilon)K_2)$, so that M_n^ϵ is the dominating term, and this concludes the proof.

Theorem 6 (The null with market microstructure noise) *Assume that X is a continuous semimartingale as defined by Eq. (1), and that Assumption 1 – 3 in Appendix A hold true for the stochastic coefficients μ_t and σ_t. Define $Y_{i,n} = X_{i,n} + \epsilon_{i,n}$, and, for every fixed $t \in (0,T]$, with a slight abuse of notation, $\hat{\mu}_t^n$ as in Eq. (26), and $\hat{\Sigma}_t^n$ as in Eq. (27). Assume $E(\Delta\epsilon_{t_i}) = 0$, $E(\Delta\epsilon_{t_i}) = V_{\Delta\epsilon} < \infty$ and $E((\Delta\epsilon_{t_i})^4) < \infty$. Define $\gamma_j = Cov(\Delta\epsilon_{t_i}, \Delta\epsilon\, t_{i-|j|})$, $Z_{i|j|} = \Delta\epsilon_{t_i}\Delta\epsilon_{t_{i-|j|}} - \gamma_j$ and assume that*

1. $\sum_{j=-\infty}^{\infty}\gamma_j < \infty$;
2. $E(|Z_{i|j|}|^r) < C_1$ *for all* i, j *for some* $r > 2$ *and a given constant* C_1;
3. $Z_{i|j|}$ *is mixing with* $\phi(\ell)$ *of size* $-r/(r-1)$ *or* $\alpha(\ell)$ *of size* $-2r/(r-2)$.

As $n \to \infty$, such that $h_n \to 0$ and $nh_n \to \infty$, it holds that $\Delta_n\hat{\Sigma}_t^n \xrightarrow{p} K_2 V_{\Delta\epsilon}$, which implies $T_t^n = \sqrt{\frac{h_n}{K_2}}\frac{\hat{\mu}_t^n}{\sqrt{\hat{\Sigma}_t^n}} \xrightarrow{d} \mathcal{N}(0,1)$.

Proof Follows from Theorem 5 and Theorem 2 in Newey and West (1987).

B Critical value of drift burst t-statistic

In the main text, we propose to detect drift bursts by calculating our t-statistic T_t^n progressively during the time-interval $(0,T]$. This leads to a multiple testing problem, which can cause severe size distortions, if the quantile function of the standard normal distribution is used to determine the significance of the individual T_t^n, as asserted by the asymptotic theory in Theorem 2. To control the family-wise error rate, we borrow from extreme value theory by evaluating a standardized version of the daily maximum of the absolute value of our drift burst t-statistic.

In absence of a drift burst in X, recall that for $t \in (0,T]$:

$$T_t^n \xrightarrow{d} N(0,1), \tag{48}$$

as $n \to \infty$, $h_n \to 0$, such that $nh_n \to \infty$.[17]

Now, assume that we compute T_t^n at a set of time points $t_i \in (0,T]$, for $i = 1,\dots,m$. The value of the t-statistic at time t_i is $T_{t_i}^n$. Then, we set:

$$T_m^* = \max_{t_i} |T_{t_i}^n|, \quad i = 1,\dots,m. \tag{49}$$

It follows that, as $m \to \infty$,

$$(T_m^* - b_m)a_m \xrightarrow{d} \xi, \tag{50}$$

where

$$a_m = \sqrt{2\ln(m)}, \qquad b_m = a_m - \frac{1}{2}\frac{\ln(\pi \ln(m))}{a_m}, \tag{51}$$

and the CDF of ξ is the Gumbel, i.e. $P(\xi \leq x) = \exp(-\exp(-x))$.[18]

While the convergence in Eq. (50) is known to hold, when the sequence $\left(T_{t_i}^n\right)_{i=1}^m$ is i.i.d. Gaussian, it also remains true for the absolute value of covariance-stationary normal random variables (Berman, 1964; Pakshirajan and Hebbar, 1977), because weak serial dependence does not alter the distribution of the maximum term, asymptotically.[19] On the other hand, persistence induces to large discrepancies in finite samples (e.g., Arellano-Valle and Genton, 2008). The route we follow to find drift bursts with frequent sampling of our t-statistic leads $\left(T_{t_i}^n\right)_{i=1}^m$ to be constructed from overlapping data, the extent of which 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 particular, in our implementation $T_{t_i}^n$ typically exhibits a strong serial correlation, so that m tends to severely overstate the effective number of "independent" copies in a given sample. This implies our test is too conservative, when evaluated against the Gumbel distribution.

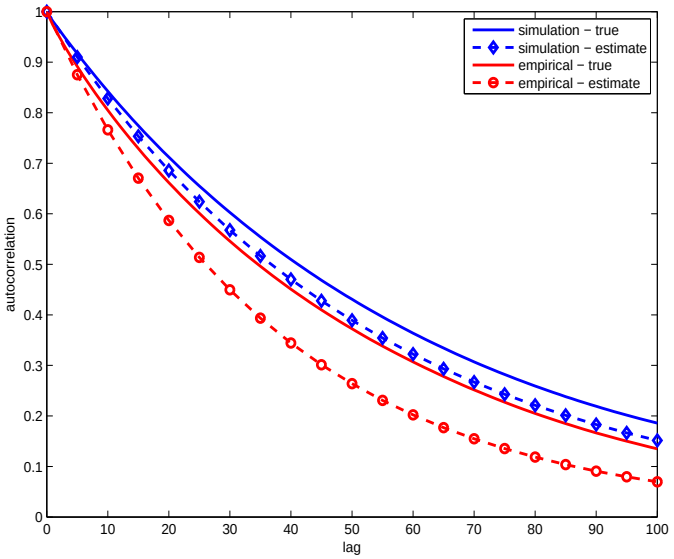
[17] Note that Theorem 2 precludes market frictions, while the corresponding result is also verified in the presence of microstructure noise in Appendix A above.

[18] David (1970) is a classic reference to the necessary extreme value theory behind this statement. Bajgrowicz, Scaillet, and Treccani (2016); Lee and Mykland (2008) 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 in this paper.

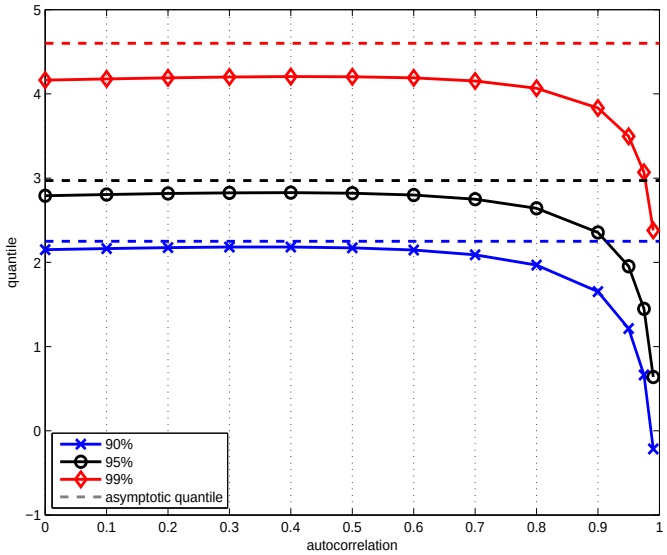
[19] In particular, sufficient conditions for this to hold is that either a) $\lim_{i\to\infty} \rho_i \log(i) = 0$ or that $\sum_{i=1}^{\infty} \rho_i^2 < \infty$, where ρ_i is the ith-order autocorrelation coefficient.

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. (52) 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. (52) 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 $1-\alpha$.

It turns out that with the left-sided exponential kernel advocated in this paper, the autocorrelation function (ACF) of $T_{t_i}^n$ decays – to a very good approximation – as that of a covariance-stationary AR(1) process with positive autoregressive coefficient (see Panel A in Figure 8):

$$Z_i = \rho Z_{i-1} + \epsilon_i, \quad i = 1,\ldots,m, \tag{52}$$

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

To account for dependence in $T_{t_i}^n$ 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. (52) from each individual series of t-statistics (i.e. OLS of Eq. (52) based on $\left(T_{t_i}^n\right)_{i=1}^m$). 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, which are discarded. In each simulation, we record the extreme value Z_m^* based on $\left(Z_i\right)_{i=1}^m$. We tabulate the quantile function of the raw and normalized Z_m^* series from Eq. (49) – (50) across the entire universe of simulations and use this table to draw inference.[20]

In Figure 8, we provide an illustration of this approach. In Panel A, we show the ACF of our t-statistic for the stochastic volatility model considered in Section 5 and the empirical high-frequency data analyzed in Section 6. We also plot the curve

[20] 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.

fitted using the above AR(1) approximation. The estimated ACF is close to the observed one, although there is a slight attenuation bias. 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).

References

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

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

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.

Arellano-Valle, R. B., and M. G. Genton, 2008, "On the exact distribution of the maximum of absolutely continuous dependent random variables," *Statistics and Probability Letters*, 78(1), 27–35.

Back, K., 1991, "Asset prices for general processes," *Journal of Mathematical Economics*, 20(4), 371–395.

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

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

Bandi, F. M., and R. Renò, 2016, "Price and volatility co-jumps," *Journal of Financial Economics*, 119(1), 107–146.

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

Barlevy, G., and P. Veronesi, 2003, "Rational panics and stock market crashes," *Journal of Economic Theory*, 110(2), 234Ű263.

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., 2016, "How crashes develop: Intradaily volatility and crash evolution," Working paper, University of Iowa.

Berman, S. M., 1964, "Limit theorems for the maximum term in stationary sequences," *Annals of Mathematical Statistics*, 35(2), 502–516.

Bjork, T., 2003, *Arbitrage theory in continuous time*. Wiley.

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.

Cont, R., and L. Wagalath, 2013, "Running for the exit: Distressed selling and endogenous correlations in financial markets," *Mathematical Finance*, 23(4), 718–741.

——— , 2014, "Fire sales forensics: Measuring endogenous risk," *Mathematical Finance*, (Forthcoming).

Danielsson, J., H. S. Shin, and J.-P. Zigrand, 2012, "Endogenous extreme events and the dual role of prices," *Annual Review of Economics*, 4(1), 111–129.

David, H. A., 1970, *Order Statistics*. John Wiley & Sons, 1 edn.

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

Diebold, F. X., and G. H. Strasser, 2013, "On the correlation structure of microstructure noise: A financial economic approach," *Review of Economic Studies*, 80(4), 1304–1337.

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.

Easley, D., and M. O'Hara, 1992, "Time and the process of security price adjustment," *Journal of Finance*, 47(2), 577–605.

French, K. R., and R. Roll, 1986, "Stock return variances: The arrival of information and the reaction of traders," *Journal of Financial Economics*, 17(1), 5–26.

Gennotte, G., and H. Leland, 1990, "Market liquidity, hedging, and crashes," *American Economic Review*, 80(5), 999–1021.

Glosten, L. R., and P. R. Milgrom, 1985, "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders," *Journal of Financial Economics*, 14(1), 71–100.

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.

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

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

Jacod, J., and A. N. Shiryaev, 2003, *Limit Theorems for Stochastic 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, 2016, "The Flash Crash: High frequency trading in an electronic market," *Journal of Finance*, (Forthcoming).

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

Kyle, A. S., 1985, "Continuous auctions and insider trading," *Econometrica*, 53(6), 1315–1335.

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*, (Forthcoming).

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.

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

Morris, S., and H. S. Shin, 2004, "Liquidity black holes," *Review of Finance*, 8(1), 1–18.

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.

Niederhoffer, V., and M. F. M. Osborne, 1966, "Market making and reversal on the stock exchange," *Journal of the American Statistical Association*, 61(316), 897–916.

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

Pakshirajan, R. P., and H. V. Hebbar, 1977, "Limit theorems for extremes of absolute values of Gaussian sequences," *Sankhya: The Indian Journal of Statistics; Series A*, 39(2), 191–195.

Roll, R., 1984, "A simple implicit measure of the effective bid-ask spread in an efficient market," *Journal of Finance*, 39(4), 1127–1139.

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., 1989, "Inferring the components of the bid-ask spread: Theory and empirical tests," *Journal of Finance*, 44(1), 115–134.

——— , 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.

Todorov, V., and G. Tauchen, 2011, "Volatility jumps," *Journal of Business and Economic Statistics*, 29(3), 2011.

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.