Evaluating the VPIN as a trigger for single-stock circuit breakers*

David Abad
University of Alicante, Spain
goliat@ua.es

Magdalena Massot
University of the Balearic Islands, Spain
m.massot@uib.es

Roberto Pascual**
University of the Balearic Islands, Spain
rpascual@uib.es

This version:
July 2015

* We thank Terrence Hendershott, José Penalva, Andriy Shkilko, and José Yagüe for insightful comments on our results. We thank audiences at the 14th EEFS Annual Conference (Brussels, 2015), and seminars at University Carlos III (Madrid, Spain) and Universitat Rovira i Virgili (Tarragona, Spain). We acknowledge the financial support of the Spanish DGICYT projects ECO2010-18567, ECO2013-4409-P, and ECO2011-29751. Roberto Pascual also acknowledges the financial support of *Fundación* BBVA. Any errors are entirely our own.
** Contact author.

Evaluating the VPIN as a trigger for single-stock circuit breakers

Abstract

We evaluate the order flow toxicity metric called VPIN (Easley, López de Prado, and O'Hara, 2012, *Review of Financial Studies* 25, 1457-1493) as a potential trigger for single-stock circuit breakers. When signaling toxicity, we show that the VPIN is not robust to changes in the key parameters of the metric. VPIN-limit hits are preceded by abrupt rises in trading volume that often anticipate peaks in short-term volatility. However, post-event illiquidity is rarely out of the ordinary. Using actual single-stock trading halts, we find that price-limit hits seldom fall within toxic periods according to the VPIN. Therefore, VPIN limits cannot replace price limits. For some VPIN versions, bid-ask spreads and volatility are higher around toxic halts that around non-toxic halts, suggesting that a properly-calibrated VPIN might occasionally succeed in predicting true toxic events.

Keywords: VPIN, BVC, circuit breakers, trading halts, price limits, order flow toxicity, adverse selection costs.

1. Introduction

In a series of related papers, Easley, López de Prado, and O'Hara (hereafter ELO) introduce a new metric of order flow toxicity called Volume-Synchronized Probability of Informed Trading (VPIN).[1] According to ELO, the order flow is regarded as toxic when it adversely selects liquidity providers. Order flow toxicity is therefore a fancy way to talk about adverse selection costs.[2]

ELO (2012a) motivate the VPIN as an updated version of Easley et al.'s (1996) PIN metric, designed to deal with modern high-frequency environments. Like the PIN, the VPIN is a function of the order imbalance, which, according to the adverse selection costs literature, signals the presence of informed traders. The VPIN is expected to increase when information events induce unbalanced and accelerated trades over relatively short intervals. Also like the PIN, high readings of the VPIN are expected to concur or precede illiquidity shortfalls and, then, short-term volatility peaks. Yet, the VPIN overcome some of the computational difficulties of the PIN approach as it does not entail the maximum-likelihood estimation of non-observable parameters. Besides, the VPIN is updated in volume increments of a constant size, so that toxicity is measured in real time.[3]

Using data on the E-mini S&P500 futures contract, ELO (2011a) find that the VPIN achieved abnormally high values in the hours preceding the Flash Crash. The extreme and persistent toxicity signaled by the VPIN, they argue, played a part in the withdrawal of many liquidity providers from the market. ELO (2012a) find that within periods of high and persistent toxicity, as measured by the VPIN, short-term volatility is often

---

[1] See ELO (2011a, 2011b, 2012a, 2012b, 2013, 2014).
[2] Seminal papers on the adverse selection costs literature include Bagehot (1971), Glosten and Milgrom (1985), Kyle (1985), and Easley and O'Hara (1987, 1992). See also O'Hara (1995).
[3] Abad and Yagüe (2012) study the connection between PIN and VPIN.

substantial. Encouraged by these promising findings, ELO (2012a, p. 1459) recommend market authorities to use the VPIN to monitor the market in real time, and eventually trigger a trading halt, or any other regulatory action, whenever the VPIN signals that liquidity provision is at risk. Our study aims to evaluate ELO's proposal, revealing some pros and cons of using the VPIN as a trigger for circuit breakers.

Extensively used in financial markets, circuit breakers are control mechanisms aimed to protect liquidity providers in times of market stress by restricting trading and/or price movements.[4] When non-discretionary, circuit breakers often use price limits as triggers. When price limits are hit, they activate an automatic switch to a temporary trading halt or a call auction, after which the continuous session resumes. As a response to the Flash Crash, US authorities approved the "limit-up, limit-down" (LULD) mechanism, a system of short-lived trading halts triggered by price limits.[5] Yet, similar circuit breakers have been in place in European markets way before the Flash Crash.[6]

ELO (2012a, p. 1459) argue that in modern high-frequency markets, "effective regulation needs to be done on an ex ante basis, anticipating problems before, and not after, then lead to market breakdowns". On a recent report on the future of computer trading in financial markets, the UK Government Office of Science (2012) also advocates for forward-looking types of circuit breakers. In this regard, ELO (2012a)

[4] Circuit breakers are the subject of an enduring controversy. Proponents argue they provide time for price reassessment, offset overreactions, and encourage liquidity provision. For opponents, they are an unnecessary impediment for trading, delay information revelation, and can even exacerbate volatility. The Flash Crash of May 6, 2010, and the proliferation of mini flash crashes (e.g., Barr, 2010; Bowley, 2010; Johnson et al., 2012; Golub, Keane and Poon, 2012; Farrel, 2013) have reignited the interest on circuit breakers. For a recent survey on this literature, see Abad and Pascual (2013).

[5] The LULD mechanism prevents trades from occurring outside an allowable band of 5% for large caps and active ETFs and 10% for other NMS stocks. If trading does not resume within the band at most 15 seconds after a price band violation, the trading system switches to a 5-minute trading halt. The LULD was approved on May 31, 2012, and went into effect on a pilot basis on April 8, 2013. See SEC release 34-67091, May 31, 2012.

[6] For example, in the Spanish Stock Exchange (SSE) they were introduced on May 14, 2001 (e.g., Abad and Pascual, 2010), and in Xetra on October 12, 1998 (e.g., Zimmermann, 2013).

argue that circuit breaker programs based on the VPIN would have this preventive capacity, as the VPIN should be able to forestall toxicity-driven liquidity crashes. In contrast, circuit breakers triggered by price limits cannot be forward-looking. By construction, price limits are hit by unusually large price changes. Thus, they may help to cool off the market once a crash has already begun by allowing a trading halt, but they can barely prevent the crash from happening. In this paper, we evaluate if circuit breakers triggered by the VPIN can replace circuit breakers triggered by price limits.

Our exercise requires of a large sample of actual single-stock short-lived trading halts triggered by stock-specific price limits. We use 12 years (from 2002 to 2013) of trade and quote data on 45 stocks handled by the electronic order-driven platform of the Spanish Stock Exchange (SSE), covering 6,740 short-lived stock-specific trading halts triggered by price limits. As ELO (2012a), we characterize toxic events as periods of high and persistent toxicity. We presume that any rule-based circuit breaker relying on the VPIN would halt continuous trading as soon as the VPIN would reach a preset threshold we call the “VPIN limit”.

Our database includes an initiator flag, so that we know the true direction of each trade. Thus, we can estimate the VPIN using true order imbalances (hereafter, “VPIN-true”). ELO (2012a, 2013) propose a probabilistic method to assign direction to pre-aggregated volume, the Bulk Volume Classification (BVC). ELO (2013) posit that in modern high-frequency markets, where informed traders use both limit and market orders and ratios of message traffic to trades have increased exponentially, traditional order imbalance estimates based on the trade initiator no longer capture the information embedded on the order flow. Instead, the BVC is supposed to render estimates of order flow imbalance. We also estimate the VPIN under different parameterizations of the BVC algorithm (hereafter, “VPIN-BVC”).

To evaluate the VPIN as a potential trigger for circuit breakers, we answer three main questions. Firstly, does the VPIN identify the same toxic events no matter how do we implement the metric? The answer has direct implications on the implementability of the VPIN as a trigger. Our focus is on the sensitivity of the VPIN to changes in the key parameters of the metric and the BVC algorithm. If the VPIN turns out to be robust to key parameter changes, a careful calibration of the VPIN would be unnecessary.

We show that the VPIN is not robust when signaling toxic periods. In particular, changes in the size of the VPIN’s rolling window, the trade classification algorithm (true vs. BVC), and the bar type (i.e., time-, trade- or volume-based) or the bar size in applying the BVC algorithm, all result in VPIN signaling different toxic events. For example, 88% of the toxic events identified by the VPIN-true are not signaled as toxic by the VPIN-BVC, and only 39% of the toxic events identified by the VPIN-BVC using time bars are signaled as toxic by the VPIN-BVC with either trade or volume bars. Therefore, the performance of the VPIN as a trigger would be contingent on the particular parameterization chosen by market authorities. These findings do not necessarily invalidate the VPIN as a trigger for circuit breakers, but definitely complicate its application.[7] Regulators and market authorities would have to find valid criteria to optimally (re)calibrate the VPIN, not only at the market level but, most likely, at the stock level.

Secondly, is the VPIN a reliable proxy for order flow toxicity? Our focus is on whether VPIN-limit hits precede liquidity shortfalls, as should be expected under the

[7] After all, the performance of price limits is also contingent to the market authorities’ choice on the allowed range of price variability.

adverse selection costs paradigm. Failing to document a strong link between the VPIN and illiquidity would call into question the need to halt a continuous session after each and every VPIN-limit hit. We should also expect VPIN-limit hits to anticipate peaks in illiquidity-driven short-term volatility. ELO (2012a) provide evidence of unusual volatility within toxic periods, but they do not study liquidity.

We evaluate activity, liquidity, and realized volatility around VPIN-limit violations, using close-in-time non-toxic days as benchmark. We find that VPIN limits are triggered by sudden rises in trading volume, mostly within the preceding 5-minute interval. VPIN-BVC-limit hits are often followed by extraordinary increases in realized volatility, consistent with ELO (2012a). In contrast, VPIN-true-limit hits are rarely followed by unusual realized volatility, consistent with Andersen and Bondarenko (2015). Regarding liquidity, for some parameterizations of the VPIN-BVC, but never for the VPIN-true, relative spreads after VPIN-limit hits are seldom out of the ordinary, and no remarkable effects are found for the limit order book depth. Overall, our findings cast serious doubts on the reliability of the VPIN as a proxy for order-flow toxicity.

Finally, we wonder if VPIN limits can substitute price limits. To shed some light on this issue, we test if VPIN-limit hits anticipate price-limit hits. Namely, we analyze how often actual halts, or price-limit hits, fall within the boundaries of toxic events according to the VPIN (hereafter, toxic halts). With the highest VPIN limit considered, 4.5% (7-10%) of the price-limit hits are toxic according to the VPIN-true (-BVC). We find that for some VPIN-BVC implementations, illiquidity and volatility around toxic halts are significantly higher than around the average non-toxic halt, suggesting that a properly calibrated VPIN-BVC may occasionally succeed in anticipating truly toxic price-limit hits. However, since it is only occasionally successful, VPIN cannot be used on its own to prevent episodes of significant price fluctuations. Therefore, VPIN limits might complement, but never substitute, price limits.

Ours is not the first independent study in evaluating the VPIN. Wu et al. (2013) examine how well the VPIN does in predicting volatility events in 94 futures contracts. They conclude that, well calibrated, the VPIN is the best warning signal available to regulators. In contrast, Andersen and Bondarenko (2014a,b, 2015) study the incremental information carried by the VPIN beyond volatility and volume measures. They conclude that the VPIN is a poor predictor of short-run volatility once they control for trading intensity and volatility itself. ELO (2014) refute the later studies by arguing that the VPIN must do well in predicting toxicity-induced volatility, not overall volatility. We contribute to this debate by testing ELO's proposal of using the VPIN to trigger circuit breakers. On the shadow side, we point out potential operational difficulties due to the lack of robustness of the metric to key parameter changes; call in question its trustworthiness as a proxy for order flow toxicity by revealing a weak connection between high VPIN values and posterior illiquidity, and establish the non-substitutability between price limits and VPIN limits. On the bright side, we find that VPIN-BVC seldom works in anticipating true toxic halts.

The rest of the paper is structured as follows. In Section 2, we describe the VPIN approach. In section 3, we present the database and sample and provide basic market background. In Section 4, we describe the circuit breaker mechanism implemented in the SSE. In Section 5, we provide descriptive statistics on order flow toxicity according to VPIN. In Section 6, we evaluate the robustness of VPIN when signaling toxic events. In Section 7, we study liquidity and volatility around VPIN-limit violations. In Section

8, we compare simulated VPIN-limit violations with actual price-limit violations. In Section 9, we summarize some robustness tests. Finally, we conclude in Section 10.

2. The VPIN approach

In an attempt to capture pieces of new information of comparable relevance arriving to the marketplace, the VPIN is defined in volume time rather than clock time. Thus, the VPIN is a moving average of the absolute order imbalance over the most recent *n* volume increments or buckets,

$$VPIN_{i\tau(n)} = \frac{\sum_{\tau=1}^{n} |OI_{i\tau}|}{nV_i}, \qquad [1]$$

where $\tau(n)$ is the last of the *n* buckets; V_i is the size (in shares) of each bucket, and $OI_{i\tau}$ is the order imbalance in the τ-th bucket, that is, the difference between the buyer-initiated volume (V_i^B) and the seller-initiated volume (V_i^S).

Since the level of the VPIN depends on parameter choices, ELO suggest evaluating the VPIN in relative terms to its own history. They define extreme toxicity as unusually high VPIN values relative to its empirical CDF for the asset in question. As ELO, we use the empirical CDF of VPIN to convert each VPIN reading into a cumulated probability (hereafter, “relative VPIN”). ELO also stress that “it takes persistently high levels of VPIN to reliably generate large absolute returns” (p. 1486). Persistence happens when the relative VPIN reaches and stays above a critical level.

ELO (2011a, 2012a) fix the number of volume buckets needed to update the VPIN (n) at 50; we also consider n = 25 and 75. We follow Andersen and Bondarenko (2015) in computing V_i as a percentage (δ) of average daily volume over the previous month. In this way, we account for the pronounced rise in daily volume within our sample period. We fix δ at 1/50. To generate a bucket, we aggregate consecutive trades until the accumulated volume equals V_i. If a trade is for a size greater than needed to fill the bucket, the excess size is assigned to the next bucket. For each combination of n and δ, we get a different series for the VPIN. According to ELO (2012a), the VPIN should be robust to a wide range of choices of n and δ.

Unless an initiator flag is available, we need of a trade classification algorithm (TCA) to estimate $OI_{i\tau}$ in eq. [1]. ELO posit that implementing traditional tick-based TCAs, like Lee and Ready (1991), in modern high frequency markets is challenging. Instead, they introduce the BVC.[8] We use both the true direction of trades and the BVC algorithm.[9]

To apply the BVC, we must pre-aggregate volume into bars of a given size. Then, we estimate the buyer-initiated volume ($\hat{V}_b^B$) within a bar b as

$$\hat{V}_b^B = V_b \times \Phi\left(\frac{\Delta p_b}{\sigma_{\Delta p}}\right), \quad [2]$$

where V_b is the aggregated volume; $\Delta p_b = p_b - p_{b-1}$ is the price change between two consecutive bars, computed as the price of the last trade in bar b and the price of the last trade in bar b-1; $\sigma_{\Delta p}$ is the volume-weighted standard deviation of Δp_b, and $\Phi(.)$ is the CDF of the probabilistic distribution assumed for $\Delta p_b / \sigma_{\Delta p}$. The relative weights of

[8] The relative accuracy of BVC versus tick based TCAs is analyzed by ELO (2013) concluding that BVC is as accurate as the tick rule. However, the independent studies of Chakrabarty, Pascual, and Shkilko (2015), Andersen and Bondarenko (2014b), and Moos, Pöppe, and Schiereck (2014) conclude just the opposite. ELO (2013) downplay the accuracy concern by asserting that initiator-based order imbalances no longer signal the information embedded in the order flow.

[9] Andersen and Bondarenko (2014b) show that the VPIN-true is negatively related to volatility whilst the VPIN-BVC is mechanically and positively related to volatility.

buyer- and seller-initiated volume within a bar depend on how large Δp_b is relative to the assumed distribution of price changes.

It turns out that, BVC-based $OI_{i\tau}$ estimates and, consequently, VPIN estimates, depend on the choice of the type-size of bars, and the assumed CDF of $\Delta p_b / \sigma_{\Delta p}$. In our implementation of BVC, we consider time, trade, and volume bars. Time bars' length is either 60, 300, 600 or 1800 seconds. Volume and trade bars' sizes are stock-specific to account for differences in the average daily trading activity across assets and overtime. Volume bar sizes are computed as υV_i, where υ is either 2, 5, 10 or 20%, implying 50, 20, 10 or 5 volume bars per bucket, respectively. Trade bar sizes are defined analogously, but on the average daily number of trades over the preceding month.[10]

Regarding the distribution of $\Delta p_b / \sigma_{\Delta p}$, ELO (2011a) and Andersen and Bondarenko (2014c) assume normality. More recently, however, ELO (2013) and Wu et al. (2013) suggest a t-student distribution with 0.25 and 0.1 degrees of freedom, respectively. Although we initially considered all three possibilities, our main results are independent on this particular choice. Without loss of generality, we provide results with the t-student with 0.25 degrees of freedom only.

In Table I, we summarize all our parameter choices, resulting in 189 different VPIN series estimates per stock.

[Table I]

Finally, we need to define extreme toxicity. In previous studies, a toxic period starts when the relative VPIN crosses bottom-up a given threshold or VPIN limit (*p*). We

[10] We have also considered trade bars of 50, 100, 150 and 200 trades, and volume bars of 500, 1000, 3000, and 5000 shares. Results are available from the authors upon request.

choose either the 95th or 99th percentile of the empirical CDF of the VPIN as limits (e.g., Wu et al., 2013; Andersen and Bondarenko; 2015; Chakrabarty et al., 2015). We assume that a toxic period ends when the relative VPIN crosses up-bottom a second threshold $q<p$. For each p, we choose q low enough to prevent detecting nested short-lived toxic events that essentially constitute the same event. Thus, we consider three combinations of p and q: [0.99, 0.85], [0.99, 0.9], and [0.95, 0.85].

3. The market, the database, and the sample

Our database covers 12 years (from 2002 to 2013) of intraday trade and quote high-frequency data from the electronic trading platform of the Spanish Stock Exchange (SSE), called *Sistema de Interconexión Bursatil Español* (SIBE).[11] The SIBE handles the trading activity of the most active and liquid Spanish stocks. Trading is continuous from 9:00 am to 5:30 pm GMT+1, with regular call auctions at the opening (8:30-9:00 am) and closing (5:30-5:35 pm).

Liquidity supply depends entirely on the open limit order book (LOB). Our database includes snapshots of the five best ask and bid quotes. For each quote, we know the displayed depth. The LOB file adds a new register after each order submission, execution, and cancellation. Iceberg orders are allowed, but not hidden orders, meaning that all quotes are displayed but only part of the available depth is observed (see Pardo and Pascual, 2012). We only use the LOB data from the continuous trading phase.

For each trade, we know the marginal price, the direction (i.e., buy or sell), the size, the time stamp in hundredths of a second (milliseconds after June 2013), and the best

[11] According to the World Federation of Exchanges, in 2013 the SSE was the 5th stock exchange by domestic market capitalization, and the 3th by total value of share trading in the Europe-Africa-Middle East Region.

quotes prevailing before each trade. We only keep ordinary trades.[12] There are no round-lots (the minimum trade size is 1 share). An aggressive order may be filled against orders stored in the LOB with different limit prices. The resulting trade is reported as a single record in the database, with a price equal to the marginal price, that is, the price of the last share transacted. The SSE is the least fragmented of the major EU stock exchanges, meaning that our database contains almost all the trades for the Spanish stocks.[13] Trade files and quote files are matched by means of an internal sequence code.

We restrict our sample to SIBE-listed stocks that traded continuously during at least three years within our sample, and were constituents of the official market index (the IBEX-35) at least half of the time. We drop stocks with prices below 1€ at some point, or that were temporarily delisted. Our final sample consists of 45 stocks.

In Table II, Panel A, we provide cross-sectional average daily statistics on liquidity (relative spread and depth), trading activity (trades and volume) and volatility (high-low and realized volatility). We form two subsamples with the 10 largest stocks (LC) and the 10 smallest stocks (SC) by average market capitalization. We revise the composition of the LC and SC subsamples at the end of each month.

[Table II]

LC stocks are more than 10 times larger than SC stocks. They are also more liquid, with half the relative spread and 1.7 times the displayed quoted depth of SC stocks. LC

[12] We drop the first and last trade of each day, which provide the allocation price and volume of the corresponding auction. For the same reason, we drop the first trade after each intraday trading halt. Finally, we exclude after-hours and prearranged trades

[13] According to Liquid Metrix, the SIBE market share of trading for the Spanish blue chips in April 2012 was 92.55%, 6.03% for BATS Chi-X, and 1.42% for Turquoise. The SIBE is far more liquid than the competing MTFs both in terms of lower immediacy costs and higher depth. See http://www.liquidmetrix.com/LiquidMetrix/Articles/LM010 (last access December 2013).

stocks show 2.6 times more trades and 8 times more euro volume than SC stocks. Finally, LC stocks are less volatile, with realized volatility being 0.4 times that of the SC stocks. These differences are statistically significant at ordinary levels.

In Table II, Panel B, we report cross-sectional average statistics on volume bucketing for the whole sample. The average stock has 124,740 volume buckets (VPIN updates), a bucket size of 79,194 shares, and about 54 VPIN updates per day.

In Figure 1, we plot cross-sectional average monthly statistics on trading activity and liquidity through the entire sample period. We smooth each time series by computing simple moving averages over the preceding 6 months. Figure 1.a reveals a steady growth in the number of trades over the sample, only temporarily interrupted by the short selling ban in place from July 23rd, 2012, to January 31st, 2013. Figure 1.b shows that the average trade size has halved over the sample, from 3,000 to about 1,350 shares. In Figure 1.c, we plot a proxy for HFT, message traffic per trade (e.g., Hendershott, Jones, and Menkveld, 2011). Whilst it only doubled during the first half of the sample, message traffic increased from 4 messages per trade in January 2007 to 28 messages per trade right before the 2012 ban.

[Figure 1]

Figure 1.d shows that relative spreads have fallen from 0.24% to 0.1% over the sample, but with two remarkable peaks: the 2008 financial crisis and the 2012 short selling ban. Figure 1.e exposes a sharp decrease in the accumulated book depth from 344,037 shares in June 2009 to 86,156 in December 2013. A program initiated in May 2009 to reduce the tick size from 0.01 to 0.001 or 0.005, depending on the stock, might explain this structural change (e.g., Goldstein and Kavajecz, 2000). We control for tick

regime changes in posterior analyses. Figure 1.f shows that the smaller tick sizes cheapened competition for price priority, tightening the book near the best quotes.

4. The SSE's circuit breaker mechanism

Since May 2001, the SSE implements a single-stock circuit-breaking mechanism to handle episodes of extraordinary volatility. It consists of static and dynamic price limits that trigger short-lived call (a.k.a. volatility) auctions. Static price limits set the maximum permitted variation around the static price, which is the allocation price of the last (opening, closing or volatility) auction. After each volatility auction, the static price and the static price limits are revised. Dynamic price limits set the maximum allowed price around the last trade price. Static and dynamic ranges are revised every month, based on the stock's volatility over the last 6 months. Non-regular revisions also occur.

In Figure 2, we plot the cross-sectional average static and dynamic ranges for our 45 stocks. Static ranges can vary from 4% to 8%. Nonetheless, ranges up to 20% have been implemented during recent periods of extreme market turmoil, like October 2008. Dynamic ranges fluctuate between 1% and 8%. The dynamic range is always smaller than or equal to the corresponding static range.

[Figure 2]

A volatility auction is triggered when an incoming order is to be executed at a price at or above (below) the upper (lower) price limit. The call auction lasts 5 minutes plus a random end of at most 30 seconds to avoid price manipulation.[14] During the auction, the static price is set equal to the price limit that has been hit. If there is no allocation price after the auction, the static price remains the same.

[14] Volatility auctions are never mechanically extended. Discretionary extensions, however, may happen. We find 50 cases, lasting about 27 minutes on average. Our results are not driven by these auctions.

Whilst a violation of the static price limit implies a remarkable intraday variation in the stock price, a violation of the dynamic price limit may happen because of a single incoming aggressive order encountering an unusually thin book. Within our sample, there were 6,740 price limit violations; 3,794 (56.3%) dynamic, and 2,946 (43.7%) static. For the LC (SC) subsample, there were 1,553 (2,194) violations, 51.3% (53.9%) dynamic, and 48.7% (46.1%) static.

In Figure 3, we plot the yearly distribution of the number of volatility auctions per stock. Dynamic (static) limit hits are relatively more common in the first (second) half of the sample, echoing the decrease in immediacy costs observed in Figure 1. In 2009, there were only 1.24 (1.09) static (dynamic) limit hits per stock, in line with the marked increase in both dynamic and static ranges shown in Figure 2. This action was the market authority's response to the peak in the number of halts observed during 2008.

[Figure 3]

5. General statistics on toxic events according to the VPIN

In Table III, Panel A, we provide cross-sectional average statistics on the number and persistence of toxic events using n = 50, p = 0.95, and q = 0.85. In Panel B, for selected VPIN implementations, we provide changes with respect to Panel A's statistics as we vary n (25, 75) or p (0.99).

[Table III]

The VPIN-true signals less but more persistent toxic events than the VPIN-BVC. The median number of events per stock using the VPIN-true is 32, whilst using the VPIN-BVC varies from 154 to 132. The median duration of a VPIN-true toxic event is 122 buckets (1,126 minutes), whilst a VPIN-BVC toxic event lasts 47-52 buckets (287-

404 minutes), depending on the bar type and size. These differences are statistically significant at the 1% level. By definition, 5% of the relative VPIN readings must exceed the 0.95 threshold. We find that with the VPIN-true these extreme readings cluster in fewer toxic events than with the VPIN-BVC.[15]

In Panel B of Table III, we show that the number (persistence) of toxic events decreases (increases) with *n*, the size of the rolling window in eq. [1]. This is a mechanical fact. Further smoothing the VPIN time series (by increasing *n*) causes larger clusters of extreme VPIN readings. Yet, the choice of *n* may be of no substance after all if shorter small-*n* toxic events compose longer large-*n* toxic events. We provide evidence on this issue in Section 6.

Finally, increasing the VPIN limit to $p = 0.99$ causes the number of toxic events to decrease. This is again by construction, since most violations of the 0.95 limit will never reach the 0.99 limit. For example, the VPIN-BVC with v = 5% trade bars hits the 0.95 VPIN limit 6,516 times for our 45 stocks, but only in 27% of these occasions the VPIN reaches the 0.99 limit. When they do, however, toxicity persists longer.

In Figure 4, we provide the cross-sectional average number of toxic events per year for $n = 50$, $p = 0.95$, $q = 0.85$, for different VPIN parameterizations. Like trading halts in Figure 3, the number of toxic events peaks in 2002 and 2008 but, unlike trading halts, it remains flat during 2011-2012. Because toxicity is measured in relative terms, in the presence of major events within the sample, like the 2008 crisis, the VPIN might fail to signal toxic events that are local but not global highs. This shortcoming of the VPIN

[15] According to the Ansari-Bradley (1960) test, the time series of the VPIN-true changes is statistically less volatile than that of the VPIN-BVC changes for every stock within our sample and for every version of the VPIN-BVC considered, which may explain the pattern reported.

approach might be alleviated by restarting the metric periodically. We address this possibility later, in Section 9.

[Figure 4]

In Figure 5, we plot the intraday distribution of VPIN-limit hits (with p = 95%) and price-limit hits (both static and dynamic). VPIN-limit hits' intraday distribution is U-shaped, recalling the regular patterns in volatility reported in the literature for decades now.[16] Andersen and Bondarenko (2015) find that the VPIN-BVC (true) is positively (negatively) correlated with concurrent realized volatility. The regular patterns in Figure 5 only partially agree, since we find U-shaped patterns for both the VPIN-BVC and the VPIN-true. Dynamic halts display a regular intraday L-shaped pattern, mimicking the regular pattern in relative spreads within the sample, reinforcing the notion that these halts are driven by illiquidity shocks. Static halts follow an inverse L-shaped pattern, as large price variations with respect to the static price are more likely as the trading session advances. We study the overlap between toxic events and trading halts in Section 8.

[Figure 5]

6. The robustness of the VPIN to parameter changes

In this section, we address our first main research question: Is the VPIN robust to key parameter changes when signaling toxic events? We have just shown that the number of toxic events identified by the VPIN is contingent on, at least, the size of rolling window (n) and the TCA used to estimate OI in eq. [1]. Now, we evaluate if

[16] See Wood, McInish and Ord (1985), Admati and Pfeiderer (1988), McInish and Wood (1990), Andersen and Bollerslev (1997, 1998), Glezakos, Vafiadis and Mylonakis (2011), and Scholtus, van Dijk and Frinjs (2014).

alternative VPIN parameterizations may result in different VPIN-limit violations and, eventually, different trading halts.

Firstly, we focus on n. For alternative versions of the VPIN, we compute the percentage of events located with $n = 50$ that overlap with toxic events identified with $n = 25$ or 75. We assume that two or more overlapped toxic periods correspond to the same toxic event. Results are provided in Table IV. We also report the magnitude of the overlap, computed as a percentage of the focal toxic event duration, both in seconds and volume buckets. In Panel A, we report results with $p = 0.99$ and $q = 0.85$. In Panel B, we report average statistics across VPIN specifications and (p, q) choices to show that our findings are robust to changes in p and q.

[Table IV]

We find that 85.8% to 96.2% (67.2% to 72.3%) of the toxic events obtained with n = 50 overlap with one or more toxic events found with $n = 25$ (75). Conditional on a match, the overlap is close to 1 with $n = 75$ and 52 to 69% with $n = 25$, differences being statistically significant at the 1% level. From Table III, we learned that as n decreases (increases), the number of toxic events increases (decreases) and their duration decreases (increases). We now learn that shorter small-n toxic events are not just the result of cracking longer large-n toxic events into pieces. Indeed, Panel B of Table IV reveals that 57% (70%) of the $n = 25$ (50) toxic events pinpointed across all the VPIN specifications overlap with $n = 50$ (75) toxic events. Hence, a large fraction of small-n toxic events are n-specific. On the contrary, large-n toxic events tend to overlap with small-n events. For example, 93.5% (88.16%) of the $n = 75$ events overlap with one or more $n = 50$ (25) events. We can therefore preclude that any VPIN-based circuit breaker program would trigger more halts the smaller the n chosen by the market authorities.

Secondly, we study the robustness of the VPIN-BVC to changes in the bar size. In Table V, we provide results for $n = 50$, $p = 0.99$ and $q = 0.85$.[17] No matter the type of bar (time, trade or volume), we find that the performance of the VPIN-BVC as a trigger would be contingent on the bar size too. The percentage of toxic events that overlap can be as low as 48.9% for time bars of 60 and 1800 seconds or 53% for trade bars with $v = 0.02$ and 0.2. Conditional on a match, though, the median magnitude of the overlap is generally high, frequently above 90%.

[Table V]

Finally, we analyze the robustness of the VPIN to changes in the TCA. In Table VI, we summarize our findings for the VPIN-true, and the VPIN-BVC with 60-second time bars, $v = 0.02$ trade bars, and $v = 0.02$ volume bars.[18] For each VPIN implementation, we show the percentage of toxic events that overlap with those found using the other VPIN versions. We also provide average statistics across all the VPIN parameterizations considered in this study. We find that most of the toxic events pinpointed by the VPIN-true are not considered toxic according to the VPIN-BVC. In average across all the VPIN-BVC implementations, about 88% of the VPIN-true toxic events do not match a toxic event identified by the VPIN-BVC. The average magnitude of a match, when it happens, is 22% of the duration of the VPIN-true toxic event.

[Table VI]

[17] Results for other (n, p, q) parameter choices are consistent, and available upon request from the authors.
[18] Results of other VPIN-BVC choices are similar and available upon request from the authors.

In line with Chakrabarty et al. (2015), who find that the VPIN-true and the VPIN-BVC show low correlation, and Andersen and Bondarenko (2015), who reveal that the VPIN-true is negatively correlated with concurrent volatility whilst the VPIN-BVC is positively correlated, our results show that any circuit breaker program based on the VPIN-true will halt the continuous session at different moments than the VPIN-BVC. ELO (2013) maintain that, in modern high frequency markets, *OI* metrics based on the trade initiator cannot capture the information embedded in the order flow. Accordingly, we would expect the VPIN-true to be a poorer signal of order flow toxicity within the second half of our sample period. However, restricting our analysis to 2000-2006 renders similar results.[19] Besides, the lack of robustness of the VPIN approach to changes in the TCA extends to the VPIN-BVC. For example, only 39% (20%) of the toxic events identified by the VPIN-BVC using time (volume) bars, are matched with toxic events identified by the VPIN-BVC using trade or volume (time or trade) bars.

Overall, we find that the VPIN performance as a trigger for circuit breakers will critically depend on parameter and TCA choices. We notice that the performance of price limits is also contingent on choices like the reference price or the maximum allowed variation. Therefore, our findings so far do not necessarily undermine the VPIN as a potential trigger for circuit breakers. However, they do stress the need to find valid criteria to optimally calibrate the VPIN. As the trading environment changes, periodic recalibrations of the VPIN would certainly be necessary too. By the time this paper is written, the proponents of the VPIN have not dealt with this important implementation difficulty.

7. The reliability of the VPIN as a proxy for order flow toxicity

[19] These results are available upon request.

VPIN-limit violations are supposed to signal the relatively highest levels of order flow toxic within the sample period. Accordingly, the adverse selection costs literature would predict that VPIN-limit hits should precede extraordinary liquidity withdrawals and, as a result, peaks in short-term volatility. In this section, we evaluate the consistency of the VPIN as a proxy for order flow toxicity by testing the expected connection between VPIN-limit hits and illiquidity.

For each VPIN-limit hit, we compute trading activity, liquidity, and volatility statistics for 24 5-minute intervals centered on the time of the VPIN-limit violation.[20] As liquidity metrics, we use the relative spread, that is, the quoted bid-ask spread divided by the quote midpoint (*RS*); the average displayed depth in euros at the best ask and bid quotes (*D*), and the average accumulated depth in euros at the 5 best ask and bid quotes of the limit order book (*Db*). All these measures are averaged weighting by time. We measure realized volatility as the standard deviation of the 1-minute returns within each interval (*RV*). We also collect the trading volume in shares (*V*) and the number of trades (*T*) within each interval.

To properly test for abnormal market conditions around VPIN-limit violations, we need a benchmark to compare with. It is also necessary to control for any regular intraday pattern. To this end, for each toxic event we pick the 250 days without abnormal order flow toxicity and with the same tick regime closest to the event day.[21] We assume that there is no remarkable toxicity if the relative VPIN remains below 0.9. We standardize the value of each metric per interval by subtracting the metric's mean and dividing by the metric's standard deviation over the benchmark days and during the exact same interval.

---

[20] We exclude those intervals that start after the end of the toxic event.
[21] Similar results are obtained with a 50-day benchmark.

We firstly look for average abnormal illiquidity and volatility around VPIN-limit violations across all events. In Figure 6, we plot the estimated average abnormal levels of each metric across all toxic events, centered on the time of the VPIN-limit hit (t = 0). We use p = 0.99 and q = 0.85. We only report abnormal levels that are statistically different from zero at the 1% level. Figure 6.a shows the results with the VPIN-true, and Figure 6.b shows the results with the VPIN-BVC with v = 0.02 volume bars.[22]

[Figure 6]

We find that VPIN-limit violations happen because of abrupt rises in volume (V) within the last 5 minutes (t = -1). According to the VPIN-true (BVC), V deviates from its benchmark mean more than 10 (17) times its benchmark standard deviation. Abnormal levels persist at least 30 minutes after the VPIN-limit violation, but they are never as high as in t = -1. The number of trades (T) is also noticeably elevated in the 10-minute interval centered on the VPIN-limit hit (from t = -1 to t = 1). The relatively smaller increase in T (1.5 to 2.1) with respect to V suggests that VPIN limits are hit by unusually large trades.

For VPIN-true violations (Figure 6.a), realized volatility (RV) is only abnormally high (1.07) at t = 1. For the VPIN-BVC (Figure 6.b), RV progressively increases up to the time of the hit (2.32 at t = -1), peaks immediately after the hit (3.24 at t = 1), and then falls below pre-hit levels. Thus, in average terms across all VPIN-limit violations, we show that VPIN-limit hits precede short-lived abnormal increases in RV.

What about liquidity? Relying on the VPIN-true to measure order flow toxicity, we find relative spreads (RS) to be only slightly wider than usual (0.034) at t = 1. With the VPIN-BVC, RS steadily increases from t = -12 (0.434) to t = -1 (0.739), peaks at t = 1

[22] Our findings with alternative VPIN parameterizations are consistent and available upon request.

(0.915), and quickly drops to pre-hit levels. We do not find the expected negative connection between the VPIN and quoted depth. For both the VPIN-true and the VPIN-BVC, *D* and *Db* are never statistically below the benchmark average levels. Overall, the average liquidity after a VPIN-limit hit seems less dramatic than expected. We must not forget that we are supposedly dealing with the top information asymmetry risk episodes for our sample stocks over 12 years.

The average patterns we reveal in Figure 6 suggest that the VPIN may often fail as an advanced indicator of order flow toxicity. So as to provide further evidence on this issue, we take a closer look to illiquidity and volatility after each post VPIN-limit hit. Namely, we classify post event liquidity (*RS*, *Db*) and volatility (*RV*) as either ordinary or extraordinary by means of the empirical distribution of each metric during non-toxic "ordinary" periods. As ordinary days, we pick the same 250 benchmark days per stock-event as before, split their trading sessions into regular 5-minute intervals, and compute the standardized liquidity and volatility proxies. Then, we obtain the 1st, 5th, 10th, 90th, 95th and 99th percentiles of the resulting benchmark empirical distribution per stock-event. Finally, post-event standardized *RS* and *RV* (*Db*) are compared with the RHS (LHS) percentiles of the benchmark distribution. Intuitively, under ordinary trading conditions, we would expect *RV* and *RS* (*Db*) to exceed (be lower than) the corresponding 95th (5th) percentile during ordinary days in about 5% of the events.

Our results are summarized in Table VII. In Panel A, we use the VPIN-true to signal order flow toxicity; in Panels B, C, and D we use the VPIN-BVC with 60-second bars, v = 0.02 trade bars, and v = 0.02 volume bars, respectively.[23] For selected 5-minute intervals, we provide the proportion of VPIN-limit hits followed by non-ordinary

[23] Our findings with other parameter choices are alike and available upon request.

illiquidity and volatility levels. We also provide average deviations with respect to the benchmark percentiles across all toxic events.

[Table VII]

According to the VPIN-true, 12.5%, 6.1%, and 2.2% of the toxic events show *RS* levels above the 90th, 95th, and 99th benchmark percentiles, respectively, within the first 5 minutes after the VPIN-limit violation (t = 1). The occurrence of such extreme readings declines during intervals t = 3 and 5. In average terms, the post-event *RS* is significantly below the 90th percentile at the 1% level. For the *Db*, 9.7%, 5.5%, and 1.1% of the toxic events reach levels below the 10th, 5th, and 1st benchmark percentile at t = 1. In average terms, the post-event *Db* is significantly above the 10th percentile at the 1% level. Our findings suggest that liquidity levels following VPIN-true-limit violations are nothing but ordinary.

For the VPIN-BVC results are consistent across alternative implementations. About 31.5-35.9%, 22-25.4%, and 9.2-10.7% of the VPIN-limit violations show a *RS* at t = 1 above the 90th, 95th, and 99th benchmark percentiles. Regarding the *Db*, the VPIN-BVC with time bars renders similar results to those obtained with the VPIN-true. With trade and volume bars, about 16-17%, 10%, and 3.4-2.3% of the limit hits head severe illiquidity as compared with the 10th, 5th, and 1st benchmark percentiles. Cross-sectional average post-event levels for the *RS* (*Db*) are statistically below (above) the 90th (10th) benchmark percentile. Our findings suggest that, at least occasionally, VPIN-BVC-limit hits precede extraordinary upturns in illiquidity, mostly in terms of immediacy costs.

VPIN-limit violations often precede extremely high *RV* levels, most notably when toxicity is gauged by the VPIN-BVC. Depending on the type of bar, 48.2%-60.6% (22.6%-33.2%) of the toxic events show realized volatility levels at t = 1 above the 90th

(99^{th}) benchmark percentile. Outstandingly, the average *RV* at $t = 1$ across all events is statistically above (below) the non-toxic 95^{th} (99^{th}) benchmark percentile. For the VPIN-true, the *RV* at $t = 1$is extraordinary for 27.4% (8.3%) of the limit hits, and the cross-sectional average *RV* at $t = 1$ is statistically below the 90^{th} (99^{th}) benchmark percentile.

According to ELO (2011a, 2012a), high relative VPIN records indicate that toxicity is at its topmost within the sample period. Under the presumption that the VPIN works as an early warning signal that liquidity provision is at risk, we should observe liquidity providers to repeatedly withdraw from the market soon after a VPIN limit is hit. Overall, we find that VPIN-BVC-limit hits often (up to 60% of hits) precede severe increases in short-term volatility, but only seldom (up to 35% of hits) anticipate exceptionally high illiquidity. Therefore, our findings question the reliability of the VPIN as a proxy for order-flow toxicity, and cast doubts on the appropriateness of halting the continuous session after every VPIN-limit violation. After all, liquidity providers do not carry severe exposure risk in most of the occasions.

8. VPIN-limit vs. price-limit violations

Consistently with ELO (2012a), we have found that within periods of high and persistent toxicity volatility is often substantial. Since price limits are hit by unusually large absolute price moves, we could expect periods of persistent high toxicity to frequently lead and comprise actual trading halts. In that case, market authorities could consider replacing traditional price limits by VPIN limits as a trigger for circuit breakers. Conversely, ELO (2012a, 2014) argue that extreme volatility should not always be preceded by high VPIN levels, since not all volatility is due to toxicity. Therefore, we presume that trading halts that do not fall within toxic periods (hereafter "non-toxic" halts) must be of a different nature than toxic halts. We formally test this hypothesis next.

In Figure 7, we display cross-sectional average abnormal levels of the relative spread (*RS*), volume in shares (*V*), and realized volatility (*RV*) around static halts (Figure 7.a) and dynamic halts (Figure 7.b). For each trading halt, we consider 24 5-minute intervals around the recession. In this case, the benchmark comprises the 250 days closest in time to the event day, with the same tick regime and no trading halts.[24] We control for intraday regular patterns by standardizing each observation using the 250 benchmark values for the same metric and the same 5-minute interval. We only plot observations for which the null that the cross-sectional average equals zero is rejected at the 1% level.

[Figure 7]

Figure 7a reveals that static halts are preceded by progressive increases in *V* and *RV*, but peaks are reached in the 5-minute interval following the halt. Therefore, SSE halts do not succeed in cooling off the market, at least not instantly. The *RS* also peaks right after the resumption of the continuous session. *RS*, *V* and *RV* all remain abnormally high at least 1 hour after the halt.[25] In contrast, Figure 1b shows that dynamic halts are likely driven by a liquidity shortfall that is more remarkable within the last 15 minutes before the halt. Liquidity crises enhance *RV*, which attains its topmost right after the halt. *V* shows less striking rises than around static halts. Overall, Figure 7 exposes the different nature of both types of SSE volatility auctions.

[24] We do not control for toxicity because, as previously shown, a day may be toxic or not depending on the particular VPIN parameterization chosen. We will control for toxicity in posterior tests, as we look at particular VPIN implementations and we distinguish between toxic and non-toxic halts.
[25] These patterns are consistent with those provided by Abad and Pascual (2010) using a more limited sample of SSE static halts.

In Table VIII, Panel A, we provide the proportion of toxic halts (limit hits) within the sample.[26] A halt is toxic if it totally or partially falls within a toxic period according to the VPIN. We find that the majority of the SSE halts happen in periods of no remarkable toxicity. With the VPIN-true, with $p = 0.99$, only 301 out of 6,734 halts (4.5%) are toxic. Depending on the bar type and size, the VPIN-BVC classifies between 494 (7.3%) and 680 (10.1%) halts as toxic. With $p = 0.95$, the number of toxic halts increases to 727 with the VPIN-true (10.8%), and 1,173 (17.4%) to 1,546 (23%) with the VPIN-BVC. Our findings do not change if we control for the type of halt. The VPIN-true, classifies 158 out of 2,943 (5.37%) static halts and 143 out of 3,791 (3.8%) dynamic halts as toxic. Except for volume bars, the VPIN-BVC detects more toxic static halts than toxic dynamic halts; but still, the proportion of toxic static halts (5.4-12.5% with $p = 0.99$) is notably low.

[Table VIII]

Toxic periods may comprise two or more halts. Using the VPIN-true (BVC with 1,800-second time bars) and $p = 0.99$, 109 (304) out of 578 (1,307) toxic events comprise the 301 (574) toxic halts. For the VPIN-true, the first halt within a toxic period is triggered on average about 15 hours after the VPIN-limit hit. With the VPIN-BVC, the toxic halt begins 1.3 to 3.3 hours after the VPIN-limit hit.[27] We could interpret these findings as evidence that, at least occasionally, the VPIN succeeds as an early warning signal for "truly" toxic events (halts). Quite the opposite, it could be the case that the toxic periods embrace actual halts just by chance. To provide evidence on which of

[26] We drop 6 halts that occur before the VPIN series of the corresponding stock is initiated, that is, before we collect the first 50 volume buckets. We are left with 6,734 halts, 2,943 static and 3,791 dynamic.

[27] For VPIN-true with p=0.99 (0.95), 74% (77%) of all the toxic halts happen during periods of highly persistent toxicity, that is, with duration above the 75% percentile across all toxic periods. For VPIN-BVC, this percentage varies between 65-75% (69-75%).

these alternative interpretations is correct, we next evaluate if toxic halts are distinguishable from non-toxic halts.

In Table IX, we compare standardized liquidity (*RS*, *Db*), and realized volatility (*RV*) around non-toxic halts vs. toxic halts according to different parameterizations of the VPIN. In this case, we standardize the variables using the 250 days closest in time to the event (i.e., halt) day, with the same tick regime, no trading halts, and no toxicity (i.e., relative VPIN<0.9). Static halts (dynamic halts) are evaluated in Panels A to D (E to H). We focus on the 15-minute window before the price limit and the 15 minute window after the resumption of the continuous session.

[Table IX]

With the VPIN-true, liquidity (*RS* and *Db*) and *RV* around toxic static halts (Panel A) are never worse than around non-toxic halts. For dynamic halts (Panel E), differences are negligible at the 1% level. Our findings suggest that the intersections between toxic periods based on VPIN-true and actual trading halts are purely random. In contrast, with the VPIN-BVC, we find *RS* and *RV* to be significantly higher around toxic halts than around non-toxic halts for all bar types. In other words, toxic halts based on the VPIN-BVC do remarkably differ from the average non-toxic halt. Therefore, our analysis suggests that a properly calibrated VPIN-BVC may seldom succeed as an early warning signal. However, as it is only seldom successful, the VPIN should not be used on its own to prevent episodes of significant price fluctuations. Accordingly, VPIN limits might complement, but never substitute, price limits.

9. Robustness section

9.1. Robustness of the VPIN in signaling toxic halts

In Section 6 we showed that the VPIN is not robust when signaling extreme order flow toxicity. Is the VPIN at least robust in signaling toxic halts? In Table X, we show that the same trading halt may be classified as toxic or not depending on the VPIN version chosen. About 89%-95% of the toxic halts identified by the VPIN-true are not toxic according to the VPIN-BVC versions considered. Besides, 48% (49%) of the toxic halts identified by the VPIN-BVC with 60-second (v = 0.02 trade) bars are missed by the VPIN-BVC with 1800-second (v = 0.2 trade) bars. Results are slightly better for volume bars: 73% (58%) of the toxic halts identified with the VPIN-BVC with v = 0.02 (v = 0.2) are also pointed out by the VPIN-BVC with v = 0.2 (v = 0.02). Between bar types, only 17% (20%) of the toxic halts located with 1800-second (v=0.02 volume) bars are also classified as toxic when we use v = 0.02 volume (1800-second) bars.

[Table X]

9.2. Illiquidity and volatility around SSE trading halts: Out of the ordinary?

In Section 7, we showed that VPIN-limit violations often precede extraordinary realized volatility (*RV*), but only seldom anticipate extraordinary relative spreads (*RS*). For comparative purposes, we perform a similar analysis around the SSE volatility auctions. We split the one-hour window around each halt into 5-minute intervals. As in previous sections, we measure liquidity by *RS* and *Db*, and volatility by *RV*. We use the 250 days closest to the event day, with no halts, and the same tick regime, to standardize each event-day observation. Finally, as in Section 7, we compare the resulting statistics with the tails of the empirical distribution of each standardized metric over the benchmark days. We summarize our findings in Table XI. We distinguish between static (Panel A) and dynamic (Panel B) halts.

[Table XI]

We find that both illiquidity and, foremost, realized volatility are frequently severe around the halt. For 63% (75%) of the static halts, realized volatility right before (after) the halt is above the 90% of its distribution during the benchmark days. For dynamic halts, we report extreme volatility before (after) the halt in 45% (70%) of the events. Illiquidity is markedly acute around dynamic halts; for 44% (57%) of them, the relative spread right before (after) the halt falls in the RHS tail of its benchmark distribution. For static halts, extreme pre (post) illiquidity is found in 24% (34%) of the halts, which is in line with different nature of static and dynamics halts exposed in Figure 7. The illiquidity statistics reported for the VPIN-BVC in Table VII resemble those found around static halts.

By comparing Tables XI and VII, one might conclude that peaks in volatility and drops in liquidity are more common around price-limit hits than around VPIN-limit hits. Yet, opponents to circuit breakers claim that trading halts may exacerbate rather than curtail volatility both before and after the halt.[28] Accordingly, the differences observed in *RV* could be driven, at least partially, by the circuit breaker itself.

9.3. Restarting the VPIN series

In Section 5, we discussed that the VPIN measures toxicity in relative terms to the whole history of the metric. As a result, the VPIN may fail to identify toxicity periods that are local but not global highs. Restarting the VPIN series periodically could

[28] The magnet effect hypothesis asserts that circuit breakers may cause traders concerned with a likely impediment to trade to advance their trading plans, aggravating order imbalances, and eventually pushing prices towards their limits. Abad and Pascual (2007) show that there is no magnet effect in the SSE. The volatility spillover hypothesis claims that circuit breakers spread volatility over long periods of time by impeding immediate price adjustments and order imbalance corrections. The delaying information hypothesis claims that circuit breakers prevent prices from incorporating new information, and may induce informed traders to postpone trading until the resumption of the continuous session (e.g., Kyle, 1988; Lehmann, 1989; Abad and Pascual, 2013).

alleviate this potential limitation.[29] We provide some preliminary evidence by shifting the starting point of the VPIN series to December 2008 (VPIN09). In Table XII, we evaluate the robustness of the VPIN to changes in the starting point.

[Table XII]

Firstly, we find that the number of toxic events from 2009 to 2013 according to VPIN09 is larger than according to the VPIN with starting point January 2002 (VPIN02). For example, the VPIN02-true identifies 157 toxic events, whilst VPIN09-true identifies 217. With the VPIN-BVC differences are even larger, especially with volume bars. In terms of the average number of toxic events per stock, differences are statistically significant for all VPIN-BVC implementations. Secondly, with the notable exception of the VPIN-true, the number and proportion of toxic halts tends to be larger with VPIN09 than with VPIN02. Finally, the VPIN09 corroborates most of the toxic halts identified by the VPIN02. Thus, 100% (97%) of the VPIN02-BVC toxic halts with v = 0.02 trade (1800 sec. time) bars are also signaled as toxic by the VPIN09-BVC. The exception is, again, the VPIN-true. In contrast, the VPIN02 misses many toxic halts located by the VPIN09. Our findings confirm that the VPIN may miss local highs in toxicity unless it is restated regularly. Therefore, the performance of the VPIN as a trigger for circuit breakers will also depend on the starting point of the metric.

ELO design the VPIN to be a warning signal of order flow toxicity in high frequency environments. Arguably, high-frequency trading should be in an embryonic stage during most of the first-half of our sample period.[30] However, with the VPIN09

[29] Determining the optimal frequency of these updates, however, may represent another downside of the VPIN approach.
[30] The Spanish *Comisión Nacional del Mercado de Valores* (CNMV) reports that HFTs account for 25-30% of the SSE volume traded in 2010 (see CNMV, 2011). Using 2013 data on 12 large SSE-listed stocks, the European Securities and Markets Authority attributes 32% of the euro volume traded, 29% of the trades, and 46% of the orders to HFT (see ESMA, 2015).

we corroborate the conclusions obtained with the VPIN02, meaning that our results are not driven by the low-frequency part of our sample.

9.4. Static and dynamic price variations within toxic periods

In Section 8, we showed that most of the SSE trading halts are not toxic, that is, they do not happen within toxic periods defined by the VPIN. However, SSE static and dynamic price limits are regularly revised by the market authorities according to the most recent historical volatility of the asset. Therefore, it could be the case that toxic periods still contain extraordinary price changes even though the price limits in place are never hit. We investigate this possibility by analyzing the maximum price variation with respect to actual static and dynamic prices within each toxic period. We look at toxic periods with no price limit violations. Table XIII summarizes our findings.

[Table XIII]

We provide the average of the maximum dynamic (Panel A) and static (Panel B) price variations across all the toxic periods considered for different VPIN implementations. For VPIN-true, for example, the average maximum price variation with respect to the dynamic (static) price is 0.82% (2.79%). For comparative purposes, we extract the 99th percentile and the maximum of these same price variations over a benchmark period per toxic event. The benchmark period is given by the 250 days closest to the toxic event with no remarkable toxicity, no trading halts, and the same tick regime. In Table XIII, we provide the average of these benchmark statistics across toxic events. We find that the maximum toxic price variations are statistically larger than the corresponding benchmark 99th percentile, but remarkably lower than the benchmark maximum price variations.

We also provide summary statistics of the distribution of both the maximum dynamic and static price changes within toxic periods with respect to some of the SSE standardized dynamic and static price limits. We find that for 71.7% (72.9%-80%) of the VPIN-true (-BVC) toxic events, the minimum SSE dynamic price limit category (1%) is never reached. Similarly, 82% (78.2%-91.7%) of the VPIN-true (-BVC) toxic events do not reach the minimum SSE static price limit category (4%). Therefore, static and dynamic price variations within the majority of the toxic periods identified by the VPIN would be qualified as acceptable by the SSE authorities.

Conclusions

Episodes of severe price swings that occur in extremely short periods, of which the Flash Crash of May 6, 2010 is the most representative and dramatic case, are becoming more and more frequent in modern financial markets. Traditionally, market regulators have relied on circuit breakers triggered by price limits, like the US LULD mechanism, to constrict volatility peaks. This kind of circuit breaker, however, cannot prevent crashes from happening since price limits are hit, precisely, by unusually large price changes.

In this paper, we evaluate ELO's (2011, 2012) proposal of using the VPIN, a new order flow toxicity measure, to build a forward looking type of circuit breaker. Namely, ELO recommend using the VPIN to monitor the conditions under which liquidity is provided and, eventually, trigger a trading halt when the metric signals that liquidity provision is at risk.

We use high frequency data on 45 SSE-listed stocks over 12 years (2002-2013). Our database includes 6,740 single-stock rule-based short-lived trading halts triggered by price-limit hits. Using 189 different parameterizations of the VPIN per stock, we identify toxic events, which begin when the VPIN crosses a threshold we call the VPIN limit. We assume that any VPIN-based circuit breaker would trigger a halt immediately after a VPIN-limit hit.

Firstly, we expose some practical difficulties in using the VPIN as a trigger. The VPIN is not robust to key parameter changes when signaling highly toxic periods. Changing the size of the VPIN's order imbalance rolling window, the TCA used to estimate order imbalances, the starting point of the sample period, or the bar type or size in applying the BVC algorithm, results in different toxic periods. Therefore, regulators willing to use the VPIN to design a circuit breaker mechanism will need to fix valid criteria to optimally (re)calibrate the metrics' parameters per stock. Such criteria are not currently available.

Secondly, we question the reliability of the VPIN as a proxy for order flow toxicity. VPIN-limit violations follow sudden increases in trading volume over relatively short intervals. If these volume peaks were driven by toxic order flow, we should observe remarkable liquidity shortfalls to follow and, then, short-term volatility peaks. We find that VPIN-limit violations are often followed by sudden short-term increases in realized volatility that only seldom concur with severe liquidity drops. This finding is at odds with the presumed nature of the VPIN, and questions the pertinence of halting a continuous session immediately after each VPIN-limit violation. Apparently, liquidity providers do not bear an unusual exposure risk in most of these cases.

Thirdly, we show that VPIN limits cannot be implemented in isolation. According to the VPIN approach, most of the SSE trading halts are not driven by order flow toxicity since they do not happen within toxic periods. Therefore, VPIN limits cannot substitute traditional price limits. Moreover, the same trading halt can be classified as toxic or non-toxic depending on the VPIN parameterization chosen. For some VPIN implementations, however, we find that illiquidity and volatility around toxic halts are significantly higher than around non-toxic halt. Therefore, properly calibrated, the VPIN might help to forestall episodes of true toxicity.

ELO (2011, 2012) argue that market regulators need of new instruments to deal with the new threats inherent to high frequency markets. Our analysis suggests that the VPIN is far from being the desired instrument, but it may definitely help to find the proper one.

References

Abad, D., and R. Pascual, 2007. On the magnet effect of price limits. *European Financial Management Journal* 13, 833-852.

Abad, D., and R. Pascual, 2010. Switching to a temporary call auction in times of high uncertainty. *The Journal of Financial Research* 33, 45-75.

Abad, D., and R. Pascual, 2013. Holding back volatility: circuit breakers, price limits and trading halts" in Baker, K., and H. Kiymaz (Eds.) *The Kolb Series in Finance: Market Microstructure in Emerging and Developed Markets*, John Wiley and Sons, Inc.

Abad, D., and J. Yagüe, 2012. From PIN to VPIN: An introduction to order flow toxicity. *The Spanish Review of Financial Economics* 10, 74-83.

Admati, A.R., and P. Pfeiderer, 1988. A theory of intraday patters: volume and price variability. *The Review of Financial Studies* 1, 3–40.

Anand, A., and K. Venkataraman, 2014. Should exchanges impose market maker obligations? *SSRN working paper* (http://dx.doi.org/10.2139/ssrn.2179259)

Andersen, T.G., and T. Bollerslev, 1997. Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance* 2–3, 115–158.

Andersen, T.G., and T. Bollerslev, 1998. DM–dollar volatility: intraday activity patters, microeconomic announcements, and longer run dependencies. *Journal of Finance* 53, 219–265.

Andersen, T.G., and O. Bondarenko, 2014a. VPIN and the flash crash. *Journal of Financial Markets* 17, 1-46.

Andersen, T.G., and O. Bondarenko, 2014b. Reflecting on the VPIN dispute. *Journal of Financial Markets* 17, 53-64.

Andersen, T.G., and O. Bondarenko, 2015. Assessing measures of order flow toxicity and early warning signals for market turbulence, *Review of Finance* 19, 1-54.

Bagehot, W., 1971. The only game in town. *Financial Analysts Journal* 27, 12-14.

Barr, C., 2010. Progress Energy joins flash crash crowd. *Fortune*, September 27.

Boehmer, E., K. Y. L. Fong, and J. Wu, 2014. International evidence on algorithmic trading. *SSRN working paper* (http://dx.doi.org/10.2139/ssrn.2022034)

Bowley, G., 2010. Stock swing still baffles, with an ominous tone. *New York Times*, August 22.

Brogaard, J., T. Hendershott, and R. Riordan, 2014. High frequency trading and price discovery. *Review of Financial Studies* 27, 2267-2306.

Brogaard, J., R. Riordan, A. Shkilko, and K. Sokolov, 2014. High frequency trading and extreme price movements. *SSRN Working Paper* (http://dx.doi.org/10.2139/ssrn.2531122).

Chakrabarty, B., R. Pascual, and A. Shkilko, 2015. Evaluating trade classification algorithms: bulk volume classification versus the tick rule and the Lee-Ready algorithm. *Journal of Financial Markets*, forthcoming.

CNMV, 2011. Desarrollos recientes en la microestructura de los mercados secundarios de acciones. Working Paper, #50.

Easley, D., N. Kiefer, M. O'Hara, and J.B Paperman, 1996. Liquidity, information, and infrequently traded stocks. *Journal of Finance* 51, 1405–1436.

Easley, D., M. López de Prado, and M. O'Hara, 2011a. The microstructure of the "Flash Crash": flow toxicity, liquidity crashes, and the probability of informed trading. *Journal of Portfolio Management* 37, 118–128.

Easley, D., M. López de Prado, and M. O'Hara, 2011b. The exchange of flow toxicity. *Journal of Trading* 6, 8–13.

Easley, D., M. López de Prado, and M. O'Hara, 2012a. Flow toxicity and liquidity in a high-frequency world. *Review of Financial Studies* 25, 1457-1493.

Easley, D., M. López de Prado, and M. O'Hara, 2012b. The volume clock: insights into the high frequency paradigm. *Journal of Portfolio Management* 37, 118-128.

Easley, D., M. López de Prado, and M. O'Hara, 2013. Bulk classification of trading activity. *SSRN Working Paper* (http://dx.doi.org/10.2139/ssrn.1989555).

Easley, D., M. López de Prado, and M. O'Hara, 2014. VPIN and the flash crash: a rejoinder. *Journal of Financial Markets* 17, 47-52.

Easley, D., and M. O'Hara, 1987. Price, trade size and information in security markets. *Journal of Financial Economics* 19, 69–90.

ESMA, 2014. High frequency trading activity in EU equity markets. Economic Report #1.

Egginton, J., B. Van Ness, and R. Van Ness, 2012. Exchange entrances, mergers and the evolution of order flow on NASDAQ 1993-2010. *Journal of Accounting and Finance* 12, 66-85.

Farrell, M., 2013. Mini flash crashes: a dozen a day. *CNN Money*, March 20.

Glezakos, M., K. Vafiadis, and J. Mylonakis, 2011. Analysis of Intra-Day Volatility under Economic Crisis Conditions. *International Journal of Economics and Finance* 3, 61-69.

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, 71-100.

Goldstein, M.A., and K.A. Kavjecz, 2000. Eighths, sixteenths, and market depth: changes in tick size and liquidity provision on the NYSE. *Journal of Financial Economics* 56, 125-149.

Golub, A., J. Keane, and S-H. Poon, 2012. High frequency trading and mini flash crashes. *Working Paper* (http://arxiv.org/abs/1211.6667)

Hasbrouck, J., 2013. High frequency quoting: short-term volatility in bids and offers. *SSRN Working Paper* (http://dx.doi.org/10.2139/ssrn.2237499).

Hasbrouck, J., and G. Saar, 2009. Technology and liquidity provision: the blurring of traditional definitions. *Journal of Financial Markets* 12, 143-172.

Hasbrouck, J., and G. Saar, 2013. Low-latency trading. *Journal of Financial Markets* 16, 646-679.

Hendershoot, T., M.C. Jones, and A. Menkveld, 2011. Does Algorithmic Trading Improve Liquidity? *Journal of Finance* 66, 1-33.

Johnson, N., G. Zhao, E. Hunsader, J. Meng, A. Ravindar, S. Carran, and B. Tivnan, 2012. Financial black swans driven by ultrafast machine ecology. *Working Paper* (http://arxiv.org/abs/1202.1448).

Jovanovic, B., and A. J. Menkveld, 2015. Middlemen in limit order markets. *SSRN Working Paper* (http://dx.doi.org/10.2139/ssrn.1624329).

Kirilenko, A., A.S. Kyle, M. Samadi, and T. Tuzun, 2014. Flash crash: the impact of high frequency trading on an electronic market. *SSRN Working Paper* (http://dx.doi.org/10.2139/ssrn.1686004).

Korajczyk, R.A., and D. Murphy, 2015. High frequency market making to large intstitutional trades. *SSRN working paper* (http://dx.doi.org/10.2139/ssrn.2567016)

Kyle, A., 1985. Continuous auctions and insider trading. *Econometrica* 53, 1315-1336.

Kyle, A.S., 1998. Trading halts and price limits. *Review of Futures Markets* 7, 426-434.

Lee, C. M. C., and M. Ready, 1991. Inferring trade direction from intraday data. *Journal of Finance* 46, 733-747.

Lehmann, B.N., 1989. Commentary: volatility, price resolution, and the effectiveness of price limits. *Journal of Financial Services Research* 32, 205-209.

Madhavan, A., 1992. Trading mechanisms in securities markets. *Journal of Finance* 47, 607-641.

Madhavan, A., 2012. Exchange-traded funds, market structure, and the flash crash. *Financial Analysts Journal* 68, 20-35.

McInish, T.H., and R.A. Wood, 1990. A transaction data analysis of the variability of common stock returns during 1980–1984. *Journal of Banking and Finance* 14, 99-112.

Moos, S., T. Pöppe, and D. Schiereck, 2014. The sensitivity of VPIN to the choice of trade classification algorithm, *Working Paper*.

Näes, R., and J. Skjeltorp, 2006. Order book characteristics and the volume-volatility relation: Empirical evidence from a limit order market. *Journal of Financial Markets* 9, 408–432.

Pardo, A., and R. Pascual, 2012. On the hidden side of liquidity. *European Journal of Finance* 18, 949-967.

Popper, N., 2012, Stock market flaws not so rare, data shows. *New York Times*, March 28.

Subrahmanyam, A., 2013. Algorithmic trading, the Flash Crash, and coordinated circuit breakers. *Borsa Istanbul Review* 13, 4-9.

Scholtus, M., D. van Dijk, and B. Frijns, 2014. Speed, algorithmic trading, and market quality around macroeconomic news announcements. *Journal of Banking and Finance* 38, 89-105.

Wilxocon, F., 1945. Individual comparisons by ranking methods. *Biometrics Bulletin* 1, 80-83.

Wood, R.A, T.H. McInish, and J.K. Ord, 1985. An investigation of transactions data for NYSE stocks. *Journal of Finance* 40, 723-739.

Wu, K., W. Bethel, M. Gu, D. Leinweber, and O. Ruebel, 2013. A big data approach to analyzing market volatility. *Algorithmic Finance* 2:3-4, 241-267.

Zimmermann, K., 2013. Price discovery in European volatility interruptions. *kSSRN Working Paper* (http://dx.doi.org/10.2139/ssrn.2365772).

Table I
Parameterization of the VPIN and the BVC

This table summarizes our particular parameter choices for implementing the order flow toxicity measure VPIN with or without the bulk-volume classification (BVC) algorithm.

VPIN	**Description**	**Alternatives considered**
n	Number of volume buckets in each update of the VPIN	25, 50, 75
δ	Percentage of the daily average volume	1/50
V_i	Size of the volume bucket for stock *i*	δ times the average daily trading volume of the asset over the preceding month, rounded to the closest integer.
$OI_{i\tau}$	Order imbalance for stock *i* in the τ-th volume bucket	Option #1: True *OI* based on the actual direction of trades. Option #2: BVC-based order flow imbalance
[p, q]	Thresholds for the CDF of VPIN	[0.99,0.85], [0.99,0.9], [0.95,0.85]
BVC	**Description**	**Alternatives considered**
Bar types	Data is pre-aggregated into bars of equal size	Time, trade and volume bars.
Bar size	Pre-determined bar sizes	<u>Time bars (in seconds):</u> 60, 300, 600 and 1800. <u>Trade bars (in number of trades):</u> Fixed: 50, 100, 150 and 200. Stock-specific: 0.02, 0.05, 0.1 and 0.2 of the average daily number of trades over the preceding month, rounded to the closest integer. <u>Volume bars (in shares traded):</u> Fixed: 500, 1000, 3000 and 5000. Stock-specific: 0.02, 0.05, 0.1 and 0.2 of V_i, rounded to the closest integer.
$\Phi(\cdot)$	CDF of the probabilistic distribution assumed for the standardized changes in prices between two consecutive bars	t-student with 0.1 and 0.25 degrees of freedom Normal

Table II
Sample statistics

Panel A contains cross-sectional average daily statistics on trading activity, liquidity, and volatility for our sample of 45 Spanish SIBE-listed stocks. We provide statistics for the whole sample and for two subsamples with the 10 largest and 10 smallest stocks (by market capitalization at the beginning of each year) within the sample, respectively. "Size" is the market capitalization in millions of Euros. For each trade, we compute the "relative spread" as the quoted spread prevailing before each trade, divided by the quote midpoint. Similarly, the "quoted depth" is the average of the displayed depth at the best quotes prevailing before each trade. We compute the daily average liquidity statistics weighting by trade size. "Trades" is the number of trades registered each day. "€Volume" is the daily volume in Euros. "High-low" is the ratio between the maximum and the minimum trade price of the day. "Realized volatility" is the daily standard deviation of 1-minute returns. *Price* is the daily average trade price. Standard deviations are reported in parenthesis. Panel B provides cross-sectional average volume bucket statistics. A volume bucket is defined as 1/50 the average daily volume of the asset over the preceding month. Statistical test are based on the non-parametric Wilcoxon (1945) test for equality of medians.

Panel A: Sample statistics

		All (45)	10 largest	10 smallest
Size	Capitalization	12079.42	24992.82 ***	2392.48
	($x10^{-6}$)	(15069.08)	(16734.15)	(526.98)
Liquidity	Relative spread	0.1577	0.1241 ***	0.2162
	(x100)	(0.065)	(0.061)	(0.105)
	€ Depth	4028.01	4696.52 ***	2726.51
	($x10^{-2}$)	(7972.41)	(6832.76)	(5659.45)
Activity	Trades	1506.55	2286.42 ***	873.10
		(1475.42)	(2050.18)	(480.80)
	€ Volume	3888.18	7687.05 ***	961.80
	($x10^{-4}$)	(7517.35)	(10634.96)	(462.25)
Volatility	High-Low	2.4991	2.1189 **	2.8748
	(x100)	(0.508)	(0.670)	(1.122)
	Realized Vol	0.7992	0.4631 ***	1.2125
	(x100)	(0.449)	(0.512)	(0.721)
Price		18.00	19.65 *	14.34
		(14.42)	(11.20)	(11.11)

Panel B: Volume bucketing

Number of volume buckets	124740.16
(VPIN updates)	(49359.31)
Volume bucket size	79194.18
	(49460.83)
VPIN updates per day	54.41
(δ = 1/50)	(4.60)

***,**,* means statistically different from the small caps subsample

Table III
Toxic events

In Panel A, we provide cross-sectional median statistics on the number and duration of toxic events for our sample of 45 SSE-listed stocks from January 2002 to December 2013 (interquartile ranges in parenthesis). A toxic event begins when the relative VPIN crosses bottom-up a threshold (p) given by the 95th percentile of the VPIN's empirical CDF for the corresponding stock, and ends when it crosses up-bottom another threshold (q) that correspond to the 85^{th} percentile. In Panel A, n (the length of the VPIN's rolling window) is set to 50. "True" means we use the use actual order imbalances. "BVC" means we use the order imbalance estimates based on the BVC algorithm. We report results with time bars of 60, 300 and 1800 seconds, and trade bars (volume bars) of 2%, 5%, and 20% of the average daily number of trades (volume bucket size) over the preceding month. The volume bucket size is $1/50^{th}$ of the average daily volume over the preceding month. Persistence in toxicity is measured both in minutes and number volume buckets. We also report the percentage of toxic events that begin and end within the same trading session. In Panel B, we provide deviations with respect to Panel A statistics as we change n (25 and 75) and p (99^{th} percentile). Statistical significance is evaluated using on the non-parametric test of Wilcoxon (1945) for equality of medians.

Panel A: Number and duration of toxic events (n = 50, p = 0.95, q =0.85)

Trade classification	# Events	Persistence: minutes	Persistence: buckets	% intraday events
True	32	1126	122.0	11.76
	(28.5)	(1430.5)	(117.5)	(17.14)
BVC(time,60s)	150 ***	287 ***	50.0 ***	54.79 ***
	(96.8)	(82.5)	(5.8)	(13.93)
BVC(time,300s)	132 ***	310 ***	51.0 ***	52.27 ***
	(81.8)	(88.8)	(9.3)	(18.27)
BVC(time,1800s)	132 ***	300 ***	52.0 ***	52.87 ***
	(90.5)	(90.0)	(7.3)	(11.57)
BVC(trade,v =0.02)	151 ***	404 ***	48.0 ***	41.67 ***
	(84.8)	(106.7)	(7.6)	(13.71)
BVC(trade,v =0.05)	142 ***	382 ***	48.0 ***	41.94 ***
	(83.5)	(111.3)	(7.9)	(12.20)
BVC(trade,v =0.2)	154 ***	371 ***	47.5 ***	40.30 ***
	(90.3)	(110.8)	(11.8)	(10.43)
BVC(volume,v =0.02)	142 ***	385 ***	50.5 ***	34.1 ***
	(73.3)	(148.8)	(8.0)	(10.2)
BVC(volume,v =0.05)	145 ***	378 ***	51.0 ***	35.0 ***
	(76.0)	(151.4)	(7.0)	(16.8)
BVC(volume,v =0.2)	149 ***	374 ***	49.0 ***	39.1 ***
	(98.0)	(169.9)	(7.0)	(18.2)

*** means statistically different from VPIN based on true classification at the 1% level

Table III (Cont.)
Toxic events

Panel B: Sensitivity to *n* and *p*

Case 1: n = 25 (p = 0.95, q = 0.85)				
True	46.00 ***	-683.14 ***	-71.00 ***	18.54 ***
BVC(time,300s)	165.00 ***	-170.00 ***	-25.00 ***	16.84 ***
BVC(trade,v =0.05)	149.00 ***	-206.06 ***	-24.00 ***	23.12 ***
BVC(volume,v =0.05)	136.00 ***	-198.12 ***	-26.00 ***	22.95 ***
Case 2: n = 75 (p = 0.95, q = 0.85)				
True	-12.00 ***	1323.38 ***	145.00 ***	-4.07 **
BVC(time,300s)	-50.00 ***	132.50 ***	27.00 ***	-14.91 ***
BVC(trade,v =0.05)	-52.00 ***	220.93 ***	24.00 ***	-15.17 ***
BVC(volume,v =0.05)	-56.00 ***	198.34 ***	24.00 ***	-13.37 ***
Case 3: p = 0.99 (n = 50, q = 0.85)				
True	-25.00 ***	6171.39 ***	325.00 ***	-3.43
BVC(time,300s)	-102.00 ***	138.00	26.50 ***	6.06
BVC(trade,v =0.05)	-99.00 ***	114.05 ***	12.00 ***	-7.32 *
BVC(volume,v =0.05)	-105.00 ***	89.17 **	7.00 ***	-3.49 *

***,**,* means statistically significant at the 1%, 5% and 10% level, respectively

Table IV
Robustness of the VPIN: the *n* parameter

We provide statistics on the robustness of the VPIN when signaling toxic events as we change n, the number of volume buckets used in each update of the metric. In Panel A, we provide the percentage of toxic events identified by the VPIN using n = 50 that overlap with toxic events identified by the VPIN using either n = 25 or n = 75 instead ("*% matches*"). Conditional on a match, we obtain the proportion of the focal toxic event's duration, both in seconds and in buckets, overlapped. We report the median and interquartile range (in parenthesis) across toxic events ("% time", "% buckets"). A toxic event begins when the relative VPIN (i.e., the cumulated probability of each VPIN reading) crosses bottom-up the VPIN limit p, and it ends when it crosses up-bottom a second threshold q. Results in Panel A correspond to p = 0.99 and q = 0.85. We consider different VPIN implementations. "True" means that we use the actual order imbalance within each volume bucket. "BVC" means we use estimated order imbalances according to the bulk-volume classification method. For the latter case, we provide results with 60, 300 and 1800 seconds time bars, and trade bars (volume bars) of 2%, 5%, and 20% of the average daily number of trades (volume bucket size) over the preceding month. The volume bucket size is 1/50 of the average daily volume over the preceding month. In Panel B, we provide average statistics across all the VPIN specifications considered and across three (p, q) alternatives: (0.99, 0.85), (0.99, 0.9) and (0.95, 0.85). Statistical test are based on the non-parametric Wilcoxon (1945) test for equality of medians.

Panel A: Toxic events with n = 50, p = 0.99 and q = 0.85

	Matches if n = 25			Matches if n = 75		
TCA	% matches	% time	% buckets	% matches	% time	% buckets
True	0.9619	0.6907	0.6669	0.7232 ***	0.9802 ***	0.9372 ***
		(0.42)	(0.38)		(0.16)	(0.21)
BVC(time, 60")	0.9470	0.5161	0.5682	0.6939 ***	0.9755 ***	0.8969 ***
		(0.30)	(0.28)		(0.13)	(0.22)
BVC(time, 300")	0.9512	0.5238	0.5750	0.7085 ***	0.9744 ***	0.8813 ***
		(0.30)	(0.29)		(0.11)	(0.19)
BVC(time, 1800")	0.9036	0.5385	0.5646	0.7008 ***	1.0000 ***	0.8596 ***
		(0.33)	(0.29)		(0.11)	(0.24)
BVC(trade, v = 0.02)	0.9040	0.5346	0.5350	0.6792 ***	0.9896 ***	0.9338 ***
		(0.33)	(0.26)		(0.15)	(0.23)
BVC(trade, v = 0.05)	0.9267	0.5508	0.5556	0.6875 ***	0.9792 ***	0.9363 ***
		(0.34)	(0.29)		(0.16)	(0.22)
BVC(trade, v = 0.2)	0.8585	0.5758	0.5938	0.6725 ***	1.0000 ***	0.8788 ***
		(0.39)	(0.34)		(0.20)	(0.25)
BVC(vol., v = 0.02)	0.9358	0.5486	0.5349	0.7027 ***	0.9845 ***	0.9032 ***
		(0.33)	(0.26)		(0.15)	(0.21)
BVC(vol., v = 0.05)	0.9277	0.5441	0.5294	0.7069 ***	0.9861 ***	0.9020 ***
		(0.35)	(0.26)		(0.15)	(0.22)
BVC(vol., v = 0.2)	0.9014	0.5330	0.5212	0.6871 ***	0.9968 ***	0.9060 ***
		(0.32)	(0.23)		(0.17)	(0.25)

Panel B: Matches across specifications and (p ,q) values

		Matches with n = 50			Matches with n = 75	
n = 25	0.5669	0.8507	0.7987	0.4278	0.8470	0.7910
		(0.23)	(0.26)		(0.24)	(0.31)
		Matches with n = 25			Matches with n = 75	
n = 50	0.9143	0.5430	0.5819	0.7012	0.8743	0.8434
		(0.23)	(0.26)		(0.20)	(0.28)
		Matches with n = 25			Matches with n = 50	
n = 75	0.8816	0.4074	0.4472	0.9349	0.6963	0.7108
		(0.21)	(0.27)		(0.21)	(0.27)

*** means statistilly significant differences in medians at the 1% level between n =25 and n =75

Table V
Robustness of the VPIN-BVC: the bar size

We provide statistics on the robustness of the VPIN-BVC when signaling toxic events as we change the size of the time, trade or volume bar used to apply the BVC. In Panel A, we provide evidence for time bar sizes of 60, 300 and 1800 seconds. In Panel B, we provide evidence for trade bar sizes of 2%, 5%, and 20% of the average daily number of trades over the preceding month. In Panel C, we provide evidence for volume bar sizes of 2%, 5%, and 20% of the volume bucket size. The volume bucket size is 1/50 times the average daily volume over the preceding month. For each bar size, we compute the proportion of toxic events identified by the VPIN-BVC that overlap with toxic events identified alternative bar sizes within the same type. Conditional on a match, we obtain the proportion of the focal toxic event's duration, both in seconds and in buckets, overlapped. We report the median and interquartile range (in parenthesis) across toxic events ("% buckets"). A toxic event begins when the relative VPIN-BVC (i.e., the cumulated probability of each VPIN-BVC reading) crosses bottom-up the VPIN-BVC limit p and it ends when it crosses up-bottom a second threshold q. We choose p = 0.99 and q = 0.85. In computing the VPIN-BVC in eq. [1], we use a rolling window of n=50 volume buckets. Statistical tests are based on the non-parametric Wilcoxon (1945) test for equality of medians. *** (**) means that differences in medians between consecutive bar sizes (e.g., Panel A: 300 sec. vs. 60 sec., 1800 sec. vs. 300 sec.) are statistically different at the 1% (5%) level.

Panel A: Time bars

	60 sec.		*300 sec.*		*1800 sec.*	
	% matches	% buckets	% matches	% buckets	% matches	% buckets
60 sec.			0.6330	0.9021	0.4896	0.7450***
				(0.26)		(0.41)
300 sec.	0.6822	0.8675			0.6302	0.7713***
		(0.27)				(0.38)
1800 sec.	0.5241	0.7593	0.6327	0.7966**		
		(0.39)		(0.34)		

Panel B: Trade bars

	v = 0.02		*v = 0.05*		*v = 0.2*	
v = 0.02			0.6545	0.8833	0.5307	0.7343***
				(0.30)		(0.43)
v = 0.05	0.6493	0.8730			0.5712	0.7407***
		(0.31)				(0.43)
v = 0.2	0.5061	0.7568	0.5468	0.7706		
		(0.43)		(0.41)		

Panel C: Volume bars

	v = 0.02		*v = 0.05*		*v = 0.2*	
v = 0.02			0.8581	0.9581	0.6576	0.8396***
				(0.16)		(0.31)
v = 0.05	0.8571	0.9657			0.6934	0.8685***
		(0.11)				(0.29)
v = 0.2	0.6173	0.9574	0.6497	0.9592		
		(0.17)		(0.26)		

Table VI
Robustness of the VPIN: the TCA

We provide statistics on the robustness of the VPIN when signaling toxic events to changes in the TCA used to estimate order imbalances (OIs) within volume buckets. We use true OIs, that is, OIs computed using the actual direction of trades, and OIs estimated using the BVC algorithm. In applying the BVC, we consider time, trade, and volume bars. We provide statistics for selected bar types/sizes: 60-second time bars, trade bars of size equal to 2% of the average daily number of trades over the preceding month, and volume bars of size equal to 2% of the volume bucket size. The volume bucket size is 1/50 of the average daily volume over the preceding month. For the VPIN-true, we report the cross-sectional median proportion of toxic events that match other toxic events identified by each version of the VPIN-BVC considered. We also include cross sectional median statistics across all the VPIN-BVC implementations we consider in this study: time bars of 60, 300, 600 and 1800 seconds, trade bars of 0.02, 0.05, 0.1, and 0.2 of the average daily number of trades over the preceding month, and volume bars of 0.02, 0.05, 0.1, and 0.2 of the volume bucket size. For the VPIN-BVC, given a type of bar (time, trade or volume), we provide matching statistics with the VPIN-BVC itself but for alternative bar types. Conditional on a match, we obtain the proportion of the focal toxic event's duration, both in seconds and in buckets, overlapped. We report the median and interquartile range (in parenthesis) across toxic events ("% buckets"). A toxic event begins when the relative VPIN-BVC (i.e., the cumulated probability of each VPIN-BVC reading) crosses bottom-up the VPIN-BVC limit p and it ends when it crosses up-bottom a second threshold q. We choose p = 0.99 and q = 0.85. In computing the VPIN in eq. [1], we use a rolling window of n=50 volume buckets.

	BVC(time, 60")		BVC(trade, v = 2%)		BVC(volume, v = 2%)		All BVC cases	
TCA	% matches	% buckets	% matches	% buckets	% matches	% buckets	% matches	% buckets
True	0.1107	0.2272	0.1557	0.1214	0.1228	0.4353	0.1192	0.2243
		(0.47)		(0.43)		(0.40)		(0.43)
BVC(time, 60")			0.4495	0.7568	0.3943	0.8360	0.3900	0.7874
				(0.48)		(0.29)		(0.40)
BVC(trade, v = 2%)	0.3741	0.9118			0.2465	0.9200	0.3563	0.9055
		(0.33)				(0.32)		(0.35)
BVC(volume, v = 2%)	0.2036	0.8127	0.2566	0.6709			0.2036	0.6620
		(0.52)		(0.52)				(0.53)

Table VII
Extraordinary illiquidity and volatility after VPIN-limit hits

We study the extremity of the liquidity and volatility levels reached soon after each VPIN-limit violation. A VPIN limit violation happens when the relative VPIN (i.e., the cumulated probability of each VPIN value) crosses bottom-up the VPIN limit p = 0.99 and ends when it crosses up-bottom a second threshold q = 0.85. For each VPIN-limit hit, we compare the post-event relative spreads (*RS*) and realized volatility (*RV*) with the 90th, 95th, and 99th percentiles of the empirical distribution of those same metrics over an event-specific benchmark period. Post-event LOB depth (*Db*) is compared with the 10th, 5th, and 1st percentiles of the same distribution. The benchmark period consists of the 250 days closest in time to the event day with no toxicity and the same tick regime. We split the benchmark days in regular 5-minute intervals, and all the liquidity and volatility measurements within these intervals are standardized using the mean and standard deviation of the metric over the 250 benchmark days for the same interval. The *RS* is the quoted spread divided by the quote midpoint. The *Db* is the average between the accumulated euro-volume at the 5 best bid and offer quotes of the LOB. Liquidity proxies are averaged weighting by time. The *RV* is the standard deviation of the 1-minute price changes within each 5-minute interval. We provide the proportion of extreme illiquidity and volatility levels across events for 3 selected post-event intervals: 1-5, 10-15, and 20-25 minutes. We also provide the cross-sectional average deviation between the standardized metric in the post event interval and the benchmark percentile. Statistical test are based on the non-parametric Wilcoxon (1945) test for equality of medians. Panel A provides results with the VPIN-true; Panel B with the VPIN-BVC with 60-second time bars; Panel C with the VPIN-BVC with trade bars (v = 0.02), and Panel D with the VPIN-BVC with volume bars (v = 0.02).

		RS - percentile			*Db* - percentile			*RV* - percentile		
Interval		90th	95th	99th	10th	5th	1st	90th	95th	99th
Panel A: VPIN-true										
0-5 min.	%	12.48	6.06	2.20	9.72	5.50	1.10	27.45	17.45	8.30
	Dif.	-1.181 ***	-1.782 ***	-3.265 ***	1.828 ***	1.960 ***	2.178 ***	-0.289 ***	-0.984 ***	-2.822 ***
10-15 min.	%	10.07	5.30	1.77	10.07	5.65	1.59	17.73	11.55	4.54
	Dif.	-1.271 ***	-1.870 ***	-3.350 ***	1.693 ***	1.825 ***	2.042 ***	-0.953 ***	-1.647 ***	-3.483 ***
20-25 min.	%	11.27	4.73	0.73	8.55	5.27	1.64	19.07	12.71	3.81
	Dif.	-1.317 ***	-1.915 ***	-3.395 ***	1.719 ***	1.851 ***	2.069 ***	-0.933 ***	-1.626 ***	-3.458 ***
Panel B: VPIN-BVC with 60-second bars										
0-5 min.	%	31.47	22.02	9.22	10.70	6.05	1.86	60.65	50.68	33.22
	Dif.	-0.218 ***	-0.846 ***	-2.406 ***	1.920 ***	2.055 ***	2.273 ***	2.164 ***	1.493 ***	-0.208 ***
10-15 min.	%	22.47	14.90	6.81	11.04	6.88	2.42	47.60	37.75	21.44
	Dif.	-0.610 ***	-1.237 ***	-2.788 ***	1.529 ***	1.664 ***	1.881 ***	0.864	0.192 ***	-1.513 ***
20-25 min.	%	23.12	15.44	6.39	12.09	6.92	2.36	44.32	33.78	18.59
	Dif.	-0.636 ***	-1.262 ***	-2.816 ***	1.503 ***	1.638 ***	1.856 ***	0.623 ***	-0.050 ***	-1.754 ***
Panel C: VPIN-BVC with trade bars (v = 0.02)										
0-5 min.	%	32.65	22.82	10.27	16.26	10.02	3.34	53.57	42.31	27.06
	Dif.	-0.249 ***	-0.868 ***	-2.394 ***	0.843 ***	0.965 ***	1.161 ***	1.467 ***	0.797 ***	-0.916 ***
10-15 min.	%	27.39	18.69	7.91	16.62	9.92	3.16	41.79	33.47	17.80
	Dif.	-0.461 ***	-1.077 ***	-2.603 ***	0.793 ***	0.916 ***	1.113 ***	0.528 ***	-0.142 ***	-1.853 ***
20-25 min.	%	26.08	16.49	7.70	15.94	10.20	3.12	39.33	28.51	15.50
	Dif.	-0.511 ***	-1.127 ***	-2.651 ***	0.813 ***	0.935 ***	1.133 ***	0.211 ***	-0.459 ***	-2.17 ***
Panel D: VPIN-BVC with volume bars (v = 0.02)										
0-5 min.	%	35.87	25.44	10.69	16.92	9.90	2.30	48.19	38.82	22.57
	Dif.	-0.086 ***	-0.734 ***	-2.333 ***	0.951 ***	1.080 ***	1.278 ***	1.044 ***	0.374 ***	-1.353 ***
10-15 min.	%	25.63	16.85	6.77	16.98	10.34	2.67	30.04	21.94	10.61
	Dif.	-0.578 ***	-1.225 ***	-2.825 ***	0.842 ***	0.971 ***	1.169 ***	-0.370 ***	-1.040 ***	-2.766 ***
20-25 min.	%	24.40	15.77	5.97	15.51	9.34	2.86	28.39	21.11	9.27
	Dif.	-0.617 ***	-1.263 ***	-2.860 ***	0.850 ***	0.979 ***	1.178 ***	-0.364 ***	-1.033 ***	-2.759 ***

*** Means the post-event level is different than the benchmark percentile at the 1% level across all toxic events.

Table VIII
VPIN-limit vs. price-limit hits

We provide descriptive statistics on the number of SSE toxic halts (price limit hits). A halt is toxic if it falls within a toxic period according to the VPIN metric. A toxic period begins when the relative VPIN (i.e., cumulated probability of each VPIN value) crosses bottom-up the VPIN limit *p* and ends when it crosses up-bottom a second threshold *q*. We report results for (p = 0.99, q = 0.85) and (p = 0.95, q = 0.85). In computing the VPIN, we use a rolling window of n = 50 volume buckets. The VPIN-true uses the actual direction of trades to estimate order imbalances. The VPIN-BVC uses the BVC algorithm to classify trades. The volume bucket size equals 1/50 times the average daily volume over the preceding month. We report results using the VPIN-BVC with time bars of 60 and 1800 seconds, trade bars of size equal to 2% and 20% of the average daily number of trades over the preceding month, and volume bars of size equal to 2% and 20% of the bucket size. A halt is static (dynamic) if it is triggered by a violation of the static (dynamic) price limit. Static limits are set over the allocation price of the last auction. Dynamic limits are set over the last trade price. We also report the median distance and interquartile range (in parentheses), across the toxic periods comprising at least one trading halt, from the VPIN-limit violation to the closest toxic halt, both in minutes and number of buckets. Statistical test are based on the Wilcoxon (1945) test for equality of medians.

	Trading halts					Toxic periods according to VPIN					
VPIN	Toxic	%	Static halts(%)	Dynamic halts(%)		Periods	Covering halts(%)	Distance (minutes)	iqr.	Distance (buckets)	iqr.
p = 0.99 and q = 0.85											
True	301	4.5	5.37	3.80	***	578	18.86	907.8	(2298.4)	83	(202)
BVC(time, 60")	519	7.7	8.56	7.04		1395	20.22	80.5	(237.4)	33	(63)
BVC(time, 1800")	574	8.5	11.28	6.38	***	1307	23.26	99.7	(288.6)	34	(52)
BVC(trd., v = 0.02)	667	9.9	11.82	8.41	**	1708	19.67	142.0	(408.2)	32	(60)
BVC(trd., v = 0.2)	680	10.1	12.47	8.25	***	1957	19.42	107.9	(322.2)	24	(44)
BVC(vol., v = 0.02)	494	7.3	5.54	8.73	*	1621	14.74	199.6	(456.7)	41	(66)
BVC(vol., v = 0.2)	623	9.3	7.99	10.26	**	1764	15.87	148.4	(402.9)	30	(51)
p = 0.95 and q = 0.85											
True	727	10.8	13.12	9.02	***	2233	14.24	462.9	(1459.5)	67	(124)
BVC(time, 60")	1340	19.9	23.00	17.48	***	6148	13.13	81.9	(226.7)	27	(46)
BVC(time, 1800")	1514	22.5	28.30	17.96	***	5645	15.16	94.8	(281.0)	27	(47)
BVC(trd., v = 0.02)	1483	22.0	25.01	19.70	**	6445	12.79	140.1	(408.3)	29	(48)
BVC(trd., v = 0.2)	1546	23.0	27.49	19.44	***	6953	13.10	136.1	(333.4)	25	(39)
BVC(vol., v = 0.02)	1173	17.4	14.37	19.78		6177	11.01	160.1	(406.0)	33	(50)
BVC(vol., v = 0.2)	1397	20.8	19.71	21.57		6825	11.14	146.1	(364.4)	27	(41)

***,**,* means that the hypothesis that the proportions of toxic static and dynamic halts are equal is rejected at the 1%, 5% and 10% level, respectively.

Table IX
Toxic halts vs. non-toxic halts

We test the null that liquidity and volatility around SSE toxic and non-toxic halts are alike. We use 6,734 single-stock trading halts for 45 stocks. We distinguish between static halts (Panels A to D) and dynamic halts (Panels E to H). Static halts are triggered by violations of the static price limit, set around the allocation price of the last auction. Dynamic halts are triggered by a violation of the dynamic price limits, set around the last trade price. There are 2,943 static halts and 3,791 dynamic halts within the sample. A halt is toxic if it falls within a toxic event according to the VPIN. A toxic event begins when the relative VPIN (i.e., the cumulated probability of the VPIN values) crosses bottom-up the VPIN limit p = 0.99 (VPIN-limit violation) and ends when the VPIN crosses up-bottom a second threshold q = 0.85. The VPIN-true uses actual order imbalances, whilst the VPIN-BVC uses order imbalance estimates based on the BVC algorithm. We apply BVC using time bars of 60 seconds, and trade and volume bars of size v = 0.02. The relative spread (*RS*) is the quoted spread divided by the quote midpoint weighted by time. The LOB depth (*Db*) is the accumulated displayed depth at the five best ask and bid quotes in euros, also weighted by time. Realized volatility (*RV*) is the standard deviation of the 1-minute price changes. For each metric and 5-minute interval, we provide the average difference between non-toxic and toxic halts. We use the non-parametric rank sum test of Wilcoxon (1945) to test the null that those differences equal zero. We study the 30-minute window around the halt. All the observations are standardized by subtracting the mean and dividing by the standard deviation of the corresponding variable for the exact same time interval across 250 non-toxic non-halt benchmark days.

Interval	Rspread	Dbook	Volat.
Panel A: Static halts and VPIN-true			
-3	0.170	-0.244 ***	-0.308 **
-2	0.226	-0.200 ***	-0.219
-1	0.171	-0.234 ***	-0.412 *
1	0.080	-0.244 **	-0.187
2	0.129	-0.231 ***	0.035
3	0.139	-0.231 *	-0.238
Panel B: Static halts and VPIN-BVC time bars (60-sec.)			
-3	-0.889 ***	-0.070	-2.020 ***
-2	-1.064 ***	-0.082	-2.816 ***
-1	-1.008 ***	-0.105	-2.773 ***
1	-1.196 ***	-0.199	-4.940 ***
2	-1.510 ***	-0.037	-2.417 ***
3	-1.016 ***	-0.014	-1.565 ***
Panel C: Static halts and VPIN-BVC trade bars (v = 0.02)			
-3	-0.721 ***	0.181 ***	-1.518 ***
-2	-1.001 ***	0.198 ***	-1.753 ***
-1	-1.165 ***	0.196 ***	-2.536 ***
1	-1.260 ***	0.045 **	-3.659 ***
2	-1.456 ***	0.150 ***	-1.804 ***
3	-1.200 ***	0.148 ***	-1.624 ***
Panel D: Static halts and VPIN-BVC volume bars (v = 0.02)			
-3	-1.969 ***	0.290 ***	-1.285 **
-2	-1.988 ***	0.256 ***	-1.318 **
-1	-1.791 ***	0.328 ***	-1.924 ***
1	-1.945 ***	0.177 ***	-4.616 ***
2	-1.672 ***	0.189 ***	-1.783 ***
3	-1.565 ***	0.192 ***	-0.536

***,**,* mean statistically significant difference at the 1%, 5% and 10% level, respectively.

Table IX (Cont.)
Toxic halts vs. non-toxic halts

Interval	Rspread	Dbook	Volat.
Panel E: Dynamic halts and VPIN-true			
-3	-0.928 **	0.050 *	-0.724
-2	-1.101 *	0.033 *	-0.089
-1	-1.044 **	0.113 **	-0.253
1	-0.928 **	-0.043	-1.531 **
2	-1.062 **	0.003 *	-0.717
3	-0.621	-0.038 *	-1.441 *
Panel F: Dynamic halts and VPIN-BVC time bars (60-sec.)			
-3	-2.245 ***	0.001	-2.267 ***
-2	-2.288 ***	-0.116	-3.139 ***
-1	-3.400 ***	0.046	-3.949 ***
1	-2.770 ***	-0.087	-4.221 ***
2	-2.697 ***	-0.071	-2.531 ***
3	-2.422 ***	0.058	-2.121 ***
Panel G: Dynamic halts and VPIN-BVC trade bars (v = 0.02)			
-3	-2.022 ***	0.142 ***	-3.023 ***
-2	-1.806 ***	0.157 ***	-2.958 ***
-1	-2.350 ***	0.188 **	-4.757 ***
1	-2.639 ***	0.033	-4.934 ***
2	-1.936 ***	0.011	-3.007 ***
3	-1.861 ***	0.099	-2.106 ***
Panel H: VPIN-BVC volume bars (v = 0.02)			
-3	-2.951 ***	0.249 ***	-1.313 *
-2	-2.955 ***	0.190 ***	-2.378 ***
-1	-2.944 ***	0.212 ***	-2.579 **
1	-3.276 ***	0.171 ***	-3.528 ***
2	-3.264 ***	0.100 ***	-0.772
3	-3.141 ***	0.202 ***	-1.260 ***

***,**,* mean statistically significant difference at the 1%, 5% and 10% level, respectively.

Table X
Robustness of VPIN in signaling toxic halts

We study the robustness of the VPIN approach in identifying trading halts that are driven by extreme order flow toxicity (adverse selection costs). A halt is toxic if it falls within a toxic event according to the VPIN. We provide the percentage of toxic halts identified by a given ("focal") VPIN parameterization that are still classified as toxic if we change the TCA (true vs. BVC), or the bar type (time, volume, trade) or size when applying the BVC algorithm. We use 6,734 single-stock trading halts for 45 stocks. A toxic event begins when the relative VPIN (i.e., the cumulated probability of the VPIN values) crosses bottom-up the VPIN limit p = 0.99 (VPIN-limit violation) and ends when the VPIN crosses up-bottom a second threshold q = 0.85. The VPIN-true uses actual order imbalances, whilst the VPIN-BVC uses order imbalance estimates based on the BVC algorithm. We apply BVC using time bars of 60 and 1800 seconds, and trade and volume bars of size v = 0.02 and 0.2.

VPIN parameterization	Matched toxic events (%)						
	[1]	[2]	[3]	[4]	[5]	[6]	[7]
True [1]		0.05	0.06	0.05	0.09	0.07	0.11
BVC(time, 60") [2]	0.03		0.52	0.54	0.41	0.43	0.49
BVC(time, 1800") [3]	0.03	0.47		0.53	0.49	0.17	0.33
BVC(trd., v = 0.02) [4]	0.02	0.42	0.45		0.51	0.25	0.39
BVC(trd., v = 0.2) [5]	0.04	0.31	0.41	0.50		0.18	0.31
BVC(vol., v = 0.02) [6]	0.04	0.46	0.20	0.34	0.24		0.73
BVC(vol., v = 0.2) [7]	0.05	0.41	0.30	0.41	0.33	0.58	

Table XI
Extreme illiquidity and volatility around trading halts

We study how often liquidity and volatility around price-limit violations (trading halts) are out of the ordinary. For each halt, we consider the 1-hour window around the halt, which we split into 5-minute intervals. The relative spread (*RS*) is the quoted spread divided by the quote midpoint. The LOB depth (*Db*) is the average between the accumulated euro-volume at the 5 best ask and the 5 best bid quotes of the LOB. Both *RS* and *Db* are averaged per interval weighting by time. The realized volatility (*RV*) is the standard deviation of the 1-minute price changes within each interval. Each observation is standardized by subtracting the mean and dividing by the standard deviation of the same variable across the stock-event-specific benchmark days and for the same 5-minute interval. The benchmark days consists of the 250 days closest in time to the event day with no halts and the same tick regime. We split the benchmark days into regular 5-minute intervals, compute the liquidity and volatility proxies and standardize the values per interval. Finally, we obtain the 1st, 5th, 10th, 90th, 95th and 99th percentiles of the empirical distribution of each variable. For selected intervals, we report the percentage of halts with *RS* and *RV* (*Db*) above (below) the RHS (LHS) tail percentiles. We also provide the cross-sectional average difference between the standardized metric and each percentile's value. Statistical tests are based on the non-parametric Wilcoxon (1945) test for equality of medians. We distinguish between static (Panel A) and dynamic (Panel B) halts. Static halts are triggered by violations of the static price limit, set around the allocation price of the last auction. Dynamic halts are triggered by a violation of the dynamic price limits, set around the last trade price.

Panel A: Static halts

		RS - percentile			*Db* - percentile			*RV* - percentile		
		90th	95th	99th	10th	5th	1st	90th	95th	99th
[-15 10)	%	21.93	14.71	5.68	15.53	9.37	2.98	42.79	31.98	16.71
	Dif.	-0.680 ***	-1.281 ***	-2.762 ***	0.818 ***	0.965 ***	1.209 ***	0.450 ***	-0.215 ***	-1.910 ***
[-5 0)	%	23.63	16.92	6.82	15.86	9.73	2.64	63.55	52.39	31.57
	Dif.	-0.558 ***	-1.160 ***	-2.641 ***	0.838 ***	0.984 ***	1.228 ***	2.241 ***	1.576 ***	-0.119 ***
(0 5]	%	33.91	24.70	11.72	14.88	9.07	2.58	75.26	66.05	47.41
	Dif.	0.037 ***	-0.565 ***	-2.046 ***	0.872 ***	1.018 ***	1.262 ***	3.983 ***	3.318 ***	1.621 ***
(10 15]	%	29.26	20.01	9.14	16.89	9.99	2.85	51.06	40.87	22.80
	Dif.	-0.354 ***	-0.957 ***	-2.438 ***	0.749 ***	0.895 ***	1.139 ***	0.971	0.305 ***	-1.392 ***

Panel B: Dynamic halts

		RS 90th	*RS* 95th	*RS* 99th	*Db* 10th	*Db* 5th	*Db* 1st	*RV* 90th	*RV* 95th	*RV* 99th
[-15 10)	%	31.75	23.51	11.84	21.92	12.74	3.87	30.68	22.37	12.08
	Dif.	-0.057 ***	-0.685 ***	-2.216 ***	0.554 ***	0.690 ***	0.912 ***	0.036 ***	-0.666 ***	-2.456 ***
[-5 0)	%	44.23	35.13	21.17	22.50	13.28	4.35	45.70	38.38	27.53
	Dif.	0.903 ***	0.274 ***	-1.261 ***	0.604 ***	0.739 ***	0.960 ***	2.275 ***	1.573 ***	-0.218 ***
(0 5]	%	57.27	46.61	27.28	18.17	10.31	2.90	70.47	63.28	46.00
	Dif.	1.498 ***	0.870 ***	-0.658 ***	0.696 ***	0.832 ***	1.054 ***	4.101 ***	3.399 ***	1.610 ***
(10 15]	%	39.59	30.81	15.83	20.76	12.79	3.72	37.83	29.08	16.07
	Dif.	0.359 ***	-0.269 ***	-1.797 ***	0.590 ***	0.726 ***	0.948 ***	0.403 ***	-0.299 ***	-2.088 ***

*** means that the metric is different than the benchmark percentile at the 1% level across all toxic events.

Table XII
Restarting the VPIN

We study the sensitivity of the VPIN when detecting toxic events to the starting point of the series. We consider two starting points: VPIN02 (VPIN09) refers to the VPIN with starting point January 2002 (2009). For each case, we obtain the number of toxic events and toxic halts located by the VPIN. For toxic events, we provide the total number of events for the sample ("All") and the average number of toxic events across stocks ("Avg."). We use non-parametric tests to compare VPIN02 vs VPIN09 in terms of the later statistic. For toxic halts, we provide the number (and %) of toxic halts for the sample ("All"), and the average number (and %) of toxic halts per stock ("Avg."). We use non-parametric tests to compare VPIN02 vs VPIN09 in terms of the later statistic. Finally, we report the proportion of VPIN02-based toxic halts corroborated by VPIN09 ("VPIN02(%)"), and the proportion of VPIN09-based toxic halts corroborated by VPIN02 ("VPIN09(%)"). A toxic event begins when the relative VPIN (i.e., the cumulated probability of each VPIN observation) crosses bottom-up the VPIN limit $p = 0.99$ (VPIN-limit violation) and ends when the VPIN crosses up-bottom another threshold $q = 0.85$. VPIN-true uses the actual order imbalances, whilst VPIN-BVC uses estimated order imbalances according to the BVC algorithm. We apply BVC using time bars of 60 and 1800 seconds, and trade and volume bars of size v = 0.02 and 0.2. For trade bars, v is the proportion of the average daily number of trades over the preceding month. For volume bars, v is the proportion of the volume bucket size.

	Toxic events				Toxic halts								Matched toxic halts (%)	
VPIN	VPIN02		VPIN09		VPIN02				VPIN09				VPIN02	VPIN09
	All	Avg.	All	Avg.	All	%	Avg.	%	All	%	Avg.	%	All	All
True	157	3.49	217	4.82	74	4.87	1.64	5.62	65	4.35	1.44	5.92	59.46	67.69
BVC(time, 60")	349	7.76	591	13.13 ***	56	3.69	1.24	5.14	83	5.55	1.84 **	7.01 **	89.29	60.24
BVC(time, 1800")	358	7.96	639	14.20 ***	72	4.74	1.60	7.78	106	7.09	2.36 **	9.84 *	97.22	66.04
BVC(trd., v = 0.02)	318	7.07	831	18.47 ***	81	5.34	1.80	6.74	150	10.03	3.33 ***	11.42 **	95.06	51.33
BVC(trd., v = 0.2)	427	9.49	937	20.82 ***	93	6.13	2.07	7.55	172	11.51	3.82 ***	12.04 ***	87.10	47.09
BVC(vol., v = 0.02)	142	3.16	733	16.29 ***	18	1.19	0.40	0.94	54	3.61	1.20 ***	3.27 ***	100.00	33.33
BVC(vol., v = 0.2)	259	5.76	828	18.40 ***	45	2.96	1.00	2.63	103	6.89	2.29 ***	6.62 ***	95.56	41.75

***, **, * means that the average across stocks with the VPIN09 is different than with VPIN02

Table XIII
Static and dynamic price variations within toxic periods

This table provides summary statistics (cross-sectional mean and standard deviation) on the maximum dynamic (Panel A) and static (Panel B) price variations within toxic periods according to the VPIN. We also provide the distribution of dynamic (Panel A) and static (Panel B) price variations with respect to SSE pre-stablished categories of dynamic and static price limits. A dynamic price variation is the relative change in prices with respect to the previous transaction price. A static price variation is the relative change in prices with respect to the static price, that is, the allocation price of the last auction completed. We also provide the average 99th percentile and maximum of both dynamic and static price variations across benchmark periods. Each benchmark period is toxic-event-specific, and it is given by the 250 days closest to the toxic event with no remarkable toxicity, no trading halts, and the same tick regime. A toxic event begins when the relative VPIN (i.e., the cumulated probability of each VPIN observation) crosses bottom-up the VPIN limit $p = 0.99$ (VPIN-limit violation) and ends when the VPIN crosses up-bottom another threshold $q = 0.85$. We consider different VPIN implementations. VPIN-true uses the actual order imbalances, whilst VPIN-BVC uses estimated order imbalances according to the BVC algorithm. We apply BVC using time bars of 60 and 300 seconds, and trade and volume bars of size v = 0.02 and 0.05. For trade bars, v is the proportion of the average daily number of trades over the preceding month. For volume bars, v is the proportion of the volume bucket size.

Panel A: Price changes with respect to the dynamic price

	Tox periods				Distribution of max. dynamic price change (%) w.r.t. stadardized SSE dynamic limits			
TCA	Avg.(max)	Std.(max)	99th	Max.	<=1% (min)	<=2%	<=3%	<=4%
True	0.0082	0.0047	0.0025 ***	0.0211 ***	71.73	98.72	99.57	100.00
BVC(time, 60")	0.0072	0.0051	0.0024 ***	0.0201 ***	79.13	97.56	99.64	99.91
BVC(time, 300")	0.0073	0.0054	0.0024 ***	0.0203 ***	79.98	96.66	99.39	99.90
BVC(trade, v = 0.02)	0.0083	0.0058	0.0023 ***	0.0211 ***	73.43	95.61	99.05	99.93
BVC(trade, v = 0.05)	0.0082	0.0055	0.0023 ***	0.0212 ***	72.89	96.25	99.50	99.93
BVC(vol., v = 0.02)	0.0073	0.0048	0.0022 ***	0.0183 ***	78.55	98.04	99.64	99.93
BVC(vol., v = 0.05)	0.0074	0.0049	0.0023 ***	0.0189 ***	78.20	97.90	99.64	99.93

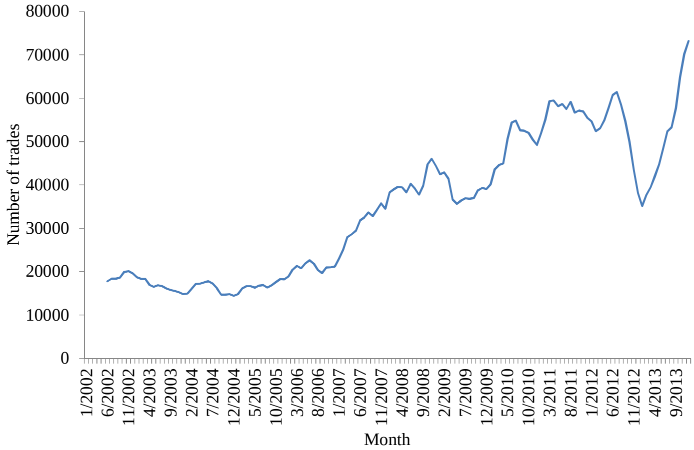
Panel B: Price changes with respect to the static price

	Tox periods				Distribution of max. static price change (%) w.r.t. standardized SSE static limits			
TCA	Avg. Max.	Std. Max.	99th	Max.	<=4% (min)	<=5%	<=7%	<=8%
True	0.0279	0.0196	0.0052 ***	0.0601 ***	82.01	89.29	95.72	97.43
BVC(time, 60")	0.0283	0.0175	0.0048 ***	0.0552 ***	81.03	91.42	97.56	98.64
BVC(time, 300")	0.0305	0.0185	0.0049 ***	0.0554 ***	78.16	87.87	96.46	97.98
BVC(trade, v = 0.02)	0.0290	0.0181	0.0047 ***	0.0596 ***	79.80	89.82	96.41	97.88
BVC(trade, v = 0.05)	0.0299	0.0180	0.0047 ***	0.0598 ***	78.51	88.75	96.40	98.05
BVC(vol., v = 0.02)	0.0226	0.0138	0.0043 ***	0.0513 ***	91.67	97.17	99.20	99.57
BVC(vol., v = 0.05)	0.0227	0.0141	0.0043 ***	0.0521 ***	91.09	96.52	98.99	99.64

*** means statistically different from the cross-sectional average maximum price variation within toxic events a tthe 1% level.

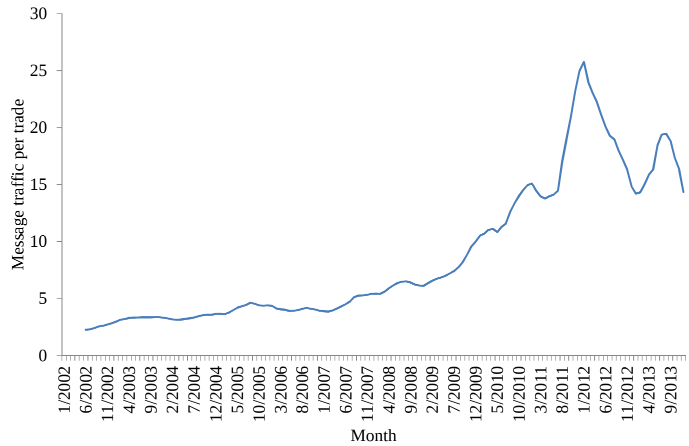
Figure 1
Trades, message traffic, and liquidity

We plot cross-sectional average monthly statistics on trading activity and liquidity for our sample of 45 SIBE-listed stocks from January 2002 to December 2013. We smooth each time series by computing 6-month moving averages. Message traffic (Fig. 1.c) is the sum of order submissions and cancellations. Relative spread (Fig. 1.d) is the quoted spread divided by the quote midpoint. Book depth (Fig. 1.e) is the average between the accumulated displayed depths at the five best offer and bid quotes. Fig. 1.f plots the average of the absolute distances from the fifth best offer and bid quotes to the quote midpoint. All liquidity measures (Figs. 1.d to 1.f) are weighted by time.



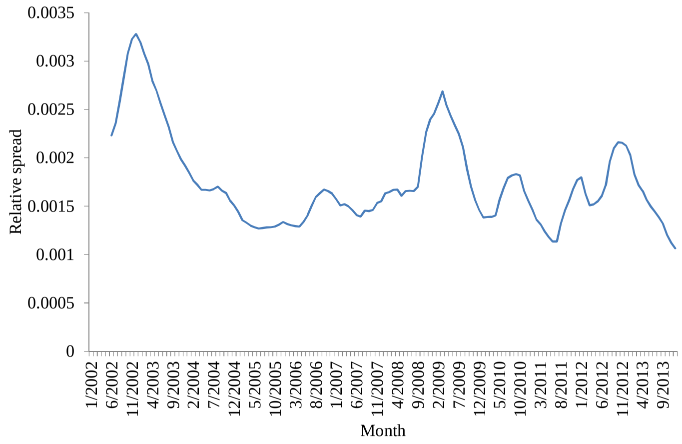
(1.a) Number of trades

(1.b) Trade size

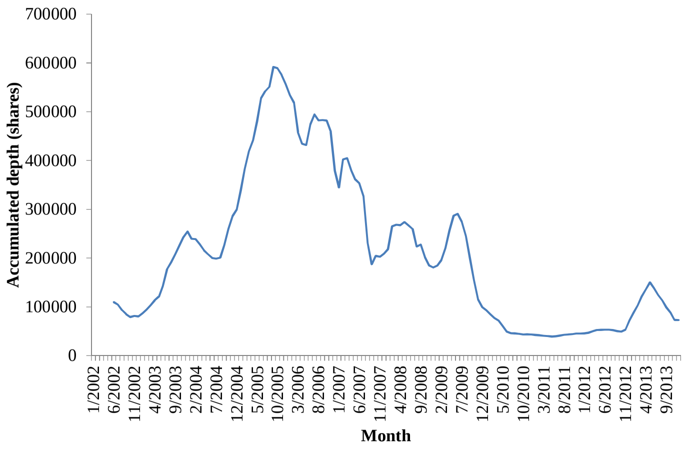


(1.c) Message traffic per trade

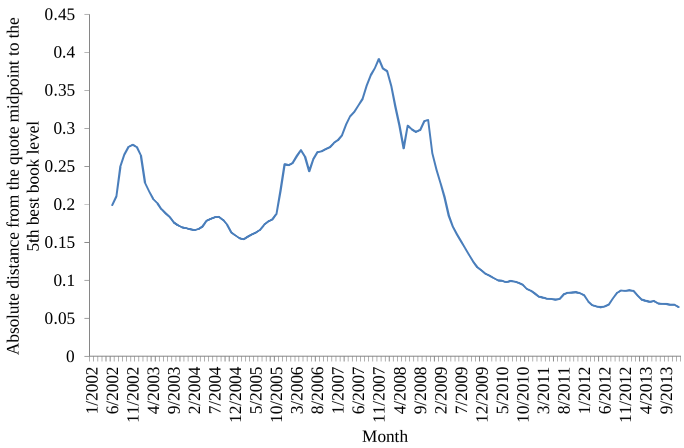
Figure 1 (Cont.)
Trades, message traffic, and liquidity



(1.d) Relative spread



(1.e) Accumulated limit order book depth at the five best quotes



(1.f) Distance from the 5 best quote of the book to the quote midpoint

Figure 2
Static and Dynamic Ranges

We plot the monthly average static and dynamic range (in %) for our sample of 45 Spanish SIBE-listed stocks from January 2002 to December 2013. Static ranges stablish the maximum absolute variation around the static price. The static price is the allocation price of the last auction completed, that can be an opening auction, a closing auction, or a reopening auction after a trading halt. Dynamic ranges set the maximum absolute variation around the last trade price. The static and dynamic ranges for a particular stock are chosen depending on its most recent historical volatility.

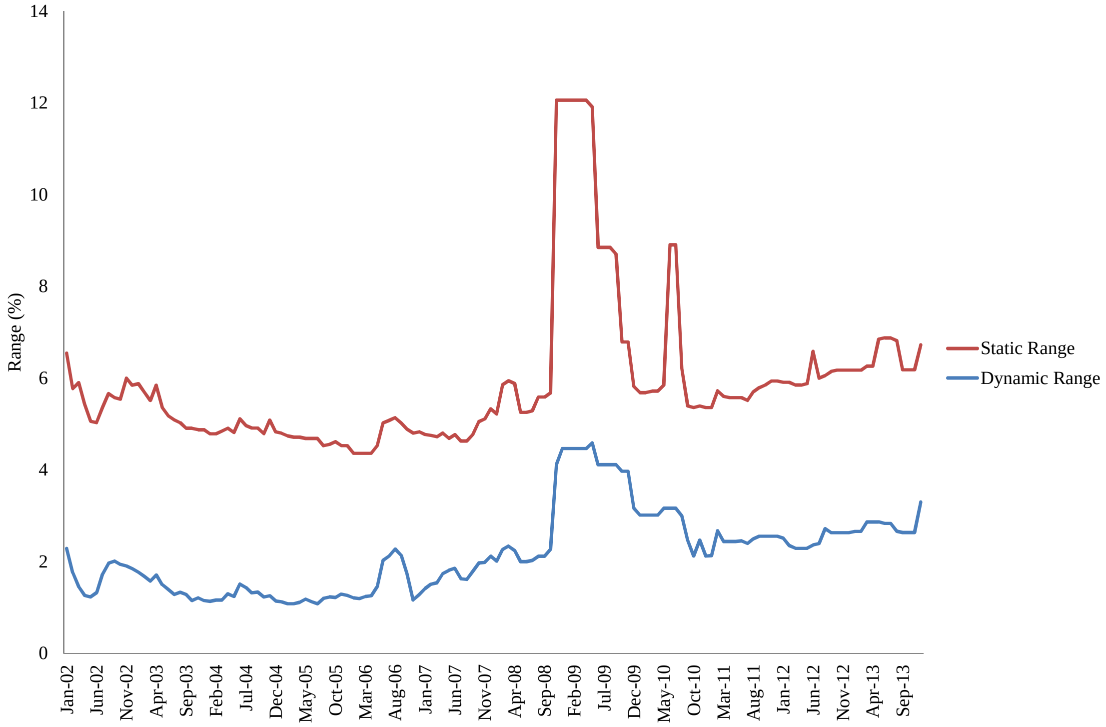


Figure 3
Trading halts per stock-year

We plot the cross-sectional average number of trading halts (volatility auctions) per stock for our sample of 45 Spanish SIBE-listed stocks from January 2002 to December 2013. We distinguish between static and dynamic halts. A static halt is triggered by a violation of the static price limits, which stablish the maximum variation permitted around the static price. The static price is the allocation price of the last auction completed, that can be an opening auction, a closing auction, or a reopening auction after a trading halt. A dynamic halt is triggered by a violation of the dynamic price limits, which set the maximum variation allowed around the last trade price. The static and dynamic price limits for a particular stock are chosen depending on its most recent historical volatility.

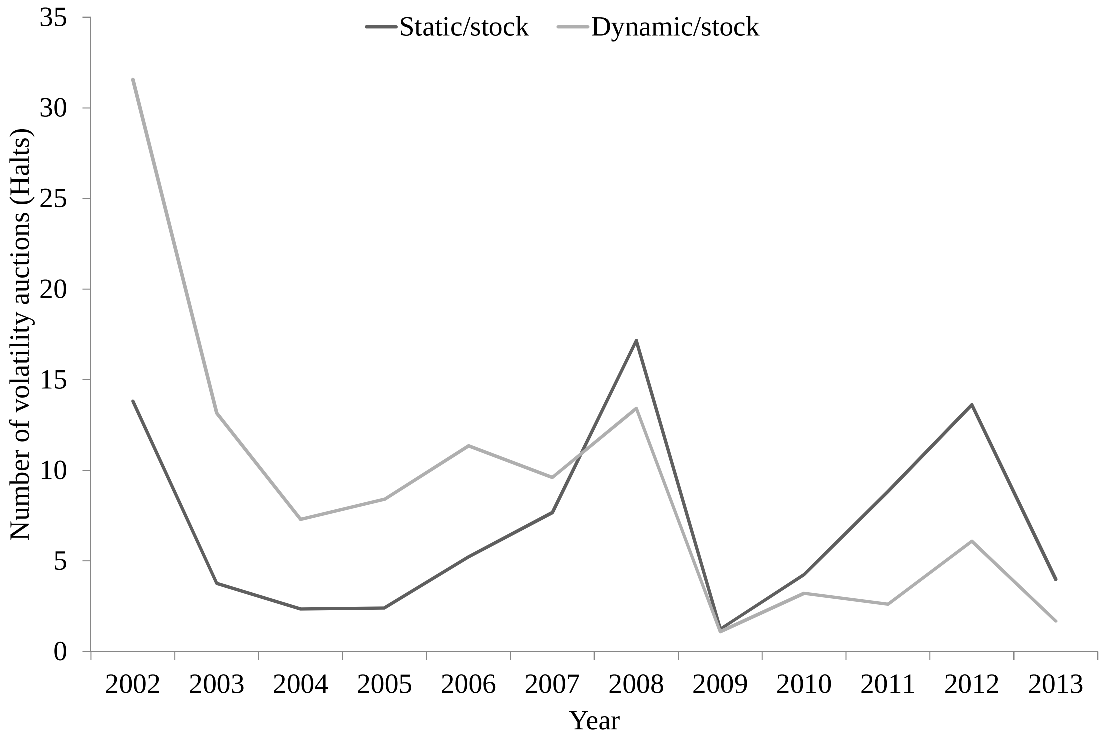


Figure 4
Number of toxic events per year

We plot the cross-sectional average number of toxic periods according to VPIN for our sample of 45 Spanish SIBE-listed stocks from January 2002 to December 2013. A toxic event begins when the relative VPIN (i.e., the cumulated probability of the VPIN) crosses bottom-up the 0.99 limit (*p*) and ends when it crosses up-bottom 0.85 limit (*q*). We select n = 50 (number of volume buckets needed to update the VPIN). We consider different parameterizations of the VPIN. “True” means we use the true direction of trades to obtain the order imbalance estimates. “BVC” means we use the bulk-volume classification method to assign direction to pre-aggregated volume. We consider time bars of 60 and 1800 seconds; trade bars (volume bars) of 2% and 20% of the average daily number of trades (volume bucket size) over the preceding month. The volume bucket size is 1/50 times the average daily volume over the preceding month.

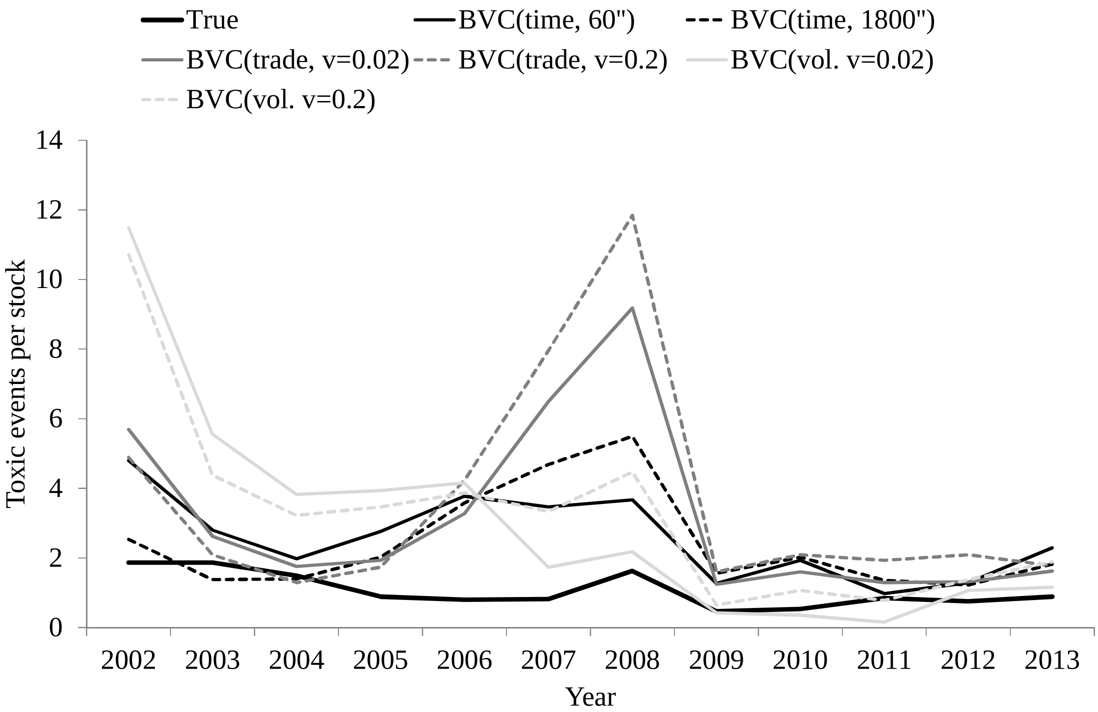


Figure 5
Intraday distribution of toxic events and trading halts

We plot the standardized number of estimated toxic events using VPIN and actual trading halts (volatility auctions) for our sample of 45 Spanish SIBE-listed stocks from January 2002 to December 2013. A toxic event begins when the relative VPIN (i.e., the cumulated probability of the VPIN) crosses bottom-up the 0.95 limit (p) and ends when it crosses up-bottom the 0.85 limit (q). We classify each toxic event according to the time the limit hit happens. We use different parameterizations of the VPIN. “True” means we use the true direction of trades. “BVC” means we use the bulk-volume classification method to assign direction to pre-aggregated volume. Volume is pre-aggregated in time bars of 60 and 1800 seconds or trade bars (volume bars) of 2% and 20% of the average daily number of trades (volume bucket size) over the preceding month. The volume bucket size is 1/50 times the average daily volume over the preceding month. We select n = 50 (number of volume buckets needed to update VPIN). Actual trading halts (volatility auctions) triggered by price limits are also classified according to the time of the limit hit. We distinguish between static and dynamic halts. Static halts are triggered by a violation of the static price limits, which stablish the maximum variation permitted around the static price. The static price is the allocation price of the last auction completed, that can be an opening auction, a closing auction, or a reopening auction after a trading halt. A dynamic halt is triggered by a violation of the dynamic price limits, which set the maximum variation allowed around the last trade price. The static and dynamic price limits for a particular stock are chosen depending on its most recent historical volatility.

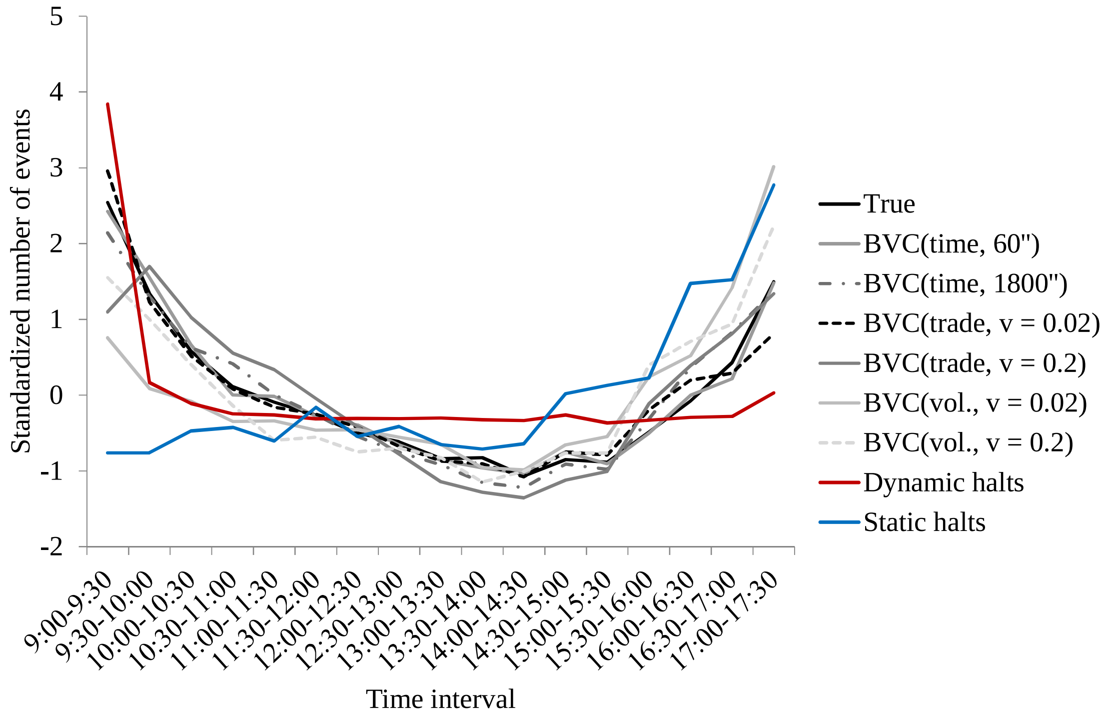
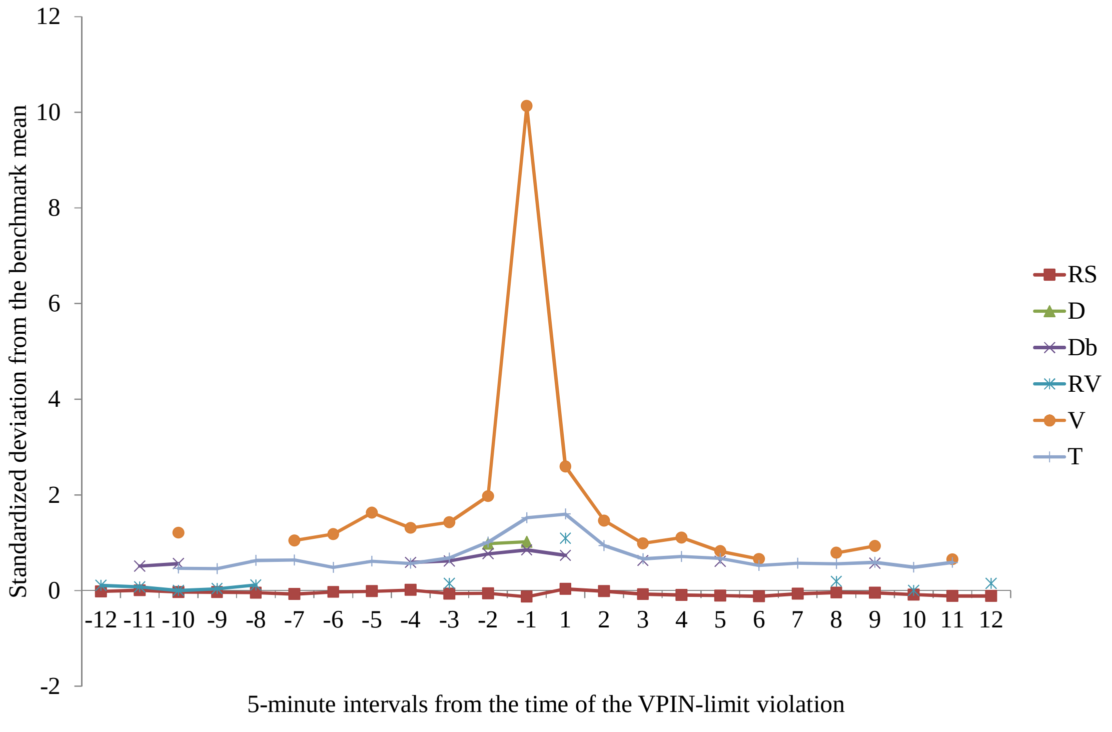


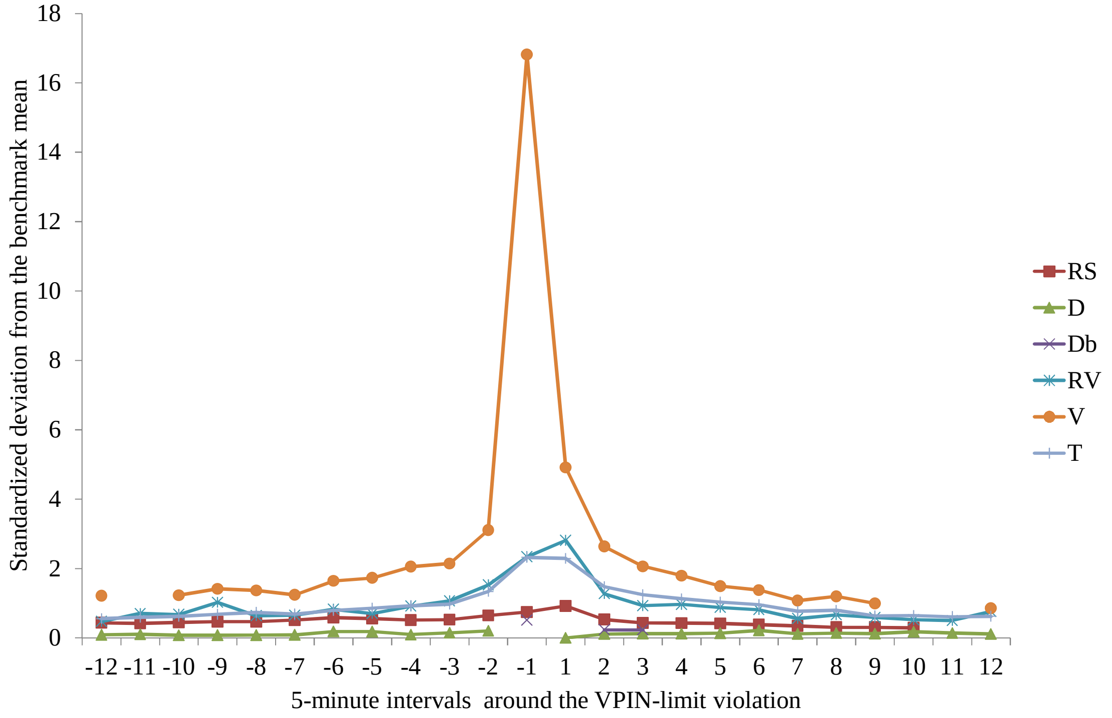
Figure 6
Characterization of VPIN-limit violations

We plot the standardized liquidity, volatility and trading activity around VPIN-limit violations (toxic events) for our sample of 45 Spanish SIBE-listed stocks from January 2002 to December 2013. A toxic event begins when the relative VPIN (i.e., the cumulated probability of the VPIN) crosses bottom-up the 0.99 limit (*p*) ("VPIN-limit violation") and ends when the VPIN crosses up-bottom the 0.85 limit (*q*). We plot our findings for 24 5-minute intervals centered on the time of the VPIN-limit violation. In Figure 6.a, we use the VPIN-true, where "true" means that we use the actual direction of trades to compute order imbalances. In Figure 6.b., we use the VPIN-BVC, where "BVC" means that we use the bulk-volume classification method to assign direction to pre-aggregated volume. In particular, we apply BVC using volume bars of size equal to 2% of the volume bucket size over the preceding month. The volume bucket size is 1/50 times the average daily volume over the preceding month. We measure liquidity by the relative spread (*RS*), the average displayed depth at the best ask and bid quotes in euros (*D*), and the average accumulated displayed depth at the five best ask and bid quotes in euros (*Db*), all weighted by time. We measure realized volatility as the standard deviation of the 1-minute changes in prices (*RV*). Finally, we use the volume in shares (*V*) and the number of trades (*T*) as proxies for trading activity. For each event, we take the closest 250 non-toxic days as the benchmark period. Each observation is standardized by subtracting the mean and dividing by the standard deviation of the same variable for the exact same time interval across the benchmark days. We plot averages across all VPIN-limit violations, but only if they are statistically different from zero according to the Wilcoxon (1945) signed rank sum test.



a) VPIN-true

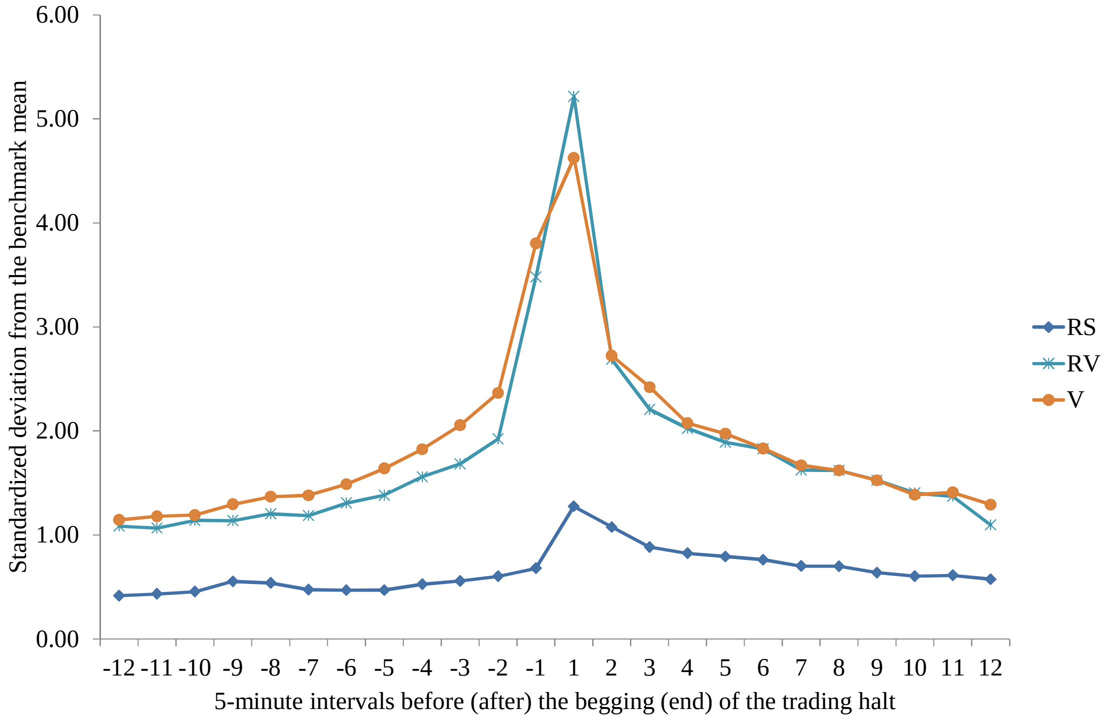
Figure 6 (Cont.)
Characterization of VPIN-limit violations



b) VPIN-BVC (volume bars)

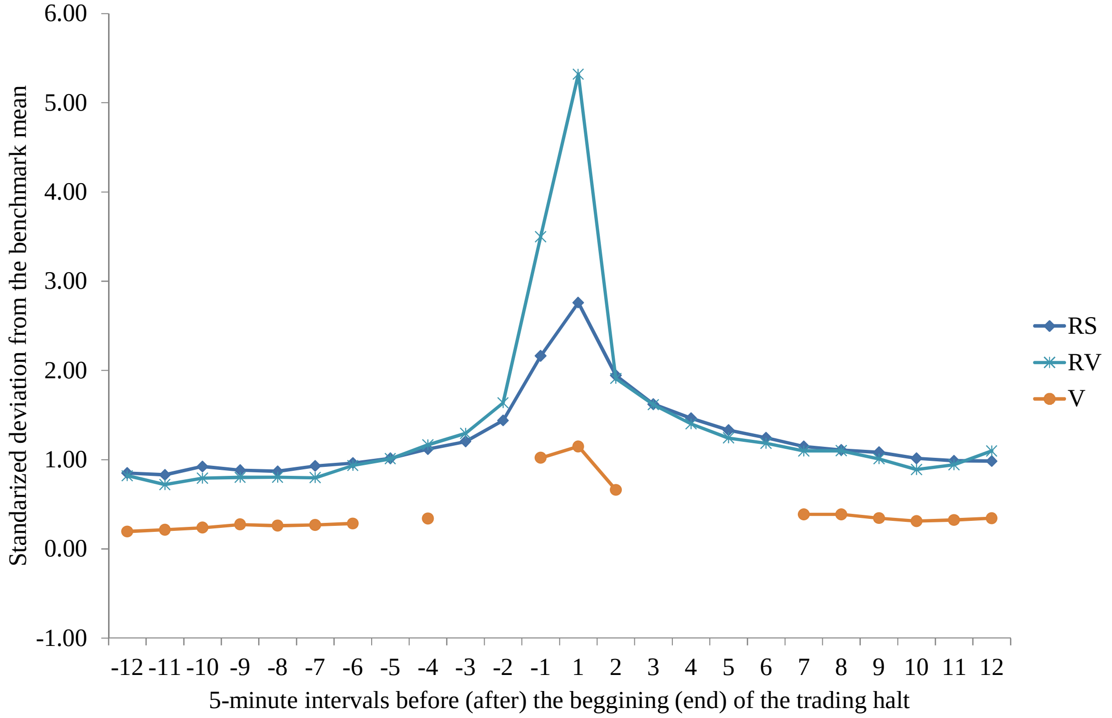
Figure 7
Characterization of SSE trading halts

We plot the average abnormal relative spread (RS), realized volatility (RV) and volume (V) around SSE single-stock trading halts (volatility auctions). Our sample consists of 45 Spanish SIBE-listed stocks from January 2002 to December 2013. We distinguish between static (Figure 7.a) and dynamic (Figure 7.b) halts. A static halt is triggered by a violation of the static price limits, the maximum variation permitted around the static price. The static price is the allocation price of the last auction completed. A dynamic halt is triggered by a violation of the dynamic price limits, the maximum variation allowed around the last trade price. We consider 12 5-minute intervals before the price limit hit and 12 5-minute intervals after the resumption of the continuous session. *RS* is the bid-ask spread divided by the quote midpoint; *RV* is the standard deviation of the 1-minute changes in prices, and *V* is measured in shares. For each event, we take the 250 days closest in time with the same tick regime and no trading halts as the benchmark period. Each observation is standardized by subtracting the mean and dividing by the standard deviation of the same variable for the exact same time interval across the 250 benchmark days. We plot averages across all trading halts when significantly different from zero according to the Wilcoxon (1945) test.



a) Static halts

Figure 7 (Cont.)
Characterization of SSE trading halts



b) Dynamic halts