High-Frequency Traders and Flash Events: Trading Activity and Liquidity Dynamics Around Mini Flash Crashes*

Christophe Desagre† Floris Laly‡ Paolo Mazza§ Mikael Petitjean¶ ‖

July 27, 2019

Abstract

We investigate the behavior of high-frequency traders (HFTs) and non-high frequency traders (NHFTs) on Nasdaq around a certain type of flash events known as mini flash crashes (MFCs), focusing on both the trading activity and liquidity dynamics of both types of traders before, during and after the crash. To do so, we identify MFCs by replicating Nanex MFC detection algorithm and complement our study with a parallel analysis of extreme price movements (EPMs). We find that (1) HFTs tend to have a virtuous behavior during the crash, whether it is a mini flash crash or an extreme price movement, by significantly reducing their liquidity demand during MFCs and EPMs occuring in the first five and last five minutes of the trading day, (2) the decline in HFTs' liquidity demand during the crash seems more pronounced during periods of known market stress as characterized by extreme hours (opening and closing periods) or when the crash is non systemic (standalone crash) as opposed to when the crash is not anticipated or potentially systemic (simultaneous crashes), (3) the price recovery that follows the crash, whether for a mini flash crash or an extreme price movement, flows from the virtuous behavior of non-high-frequency traders (while high-frequency traders viciously demand liquidity in the direction of the crash during the recovery phase).

JEL classification: G1, G10, G14

Keywords: mini flash crashes, flash events, high-frequency trading, trading activity, liquidity, market stability, market microstructure

*__Preliminary and incomplete. Please do not quote.__

†Louvain School of Management, Université Catholique de Louvain, 151 Chaussée de Binche, 7000 Mons (Belgium). Email: christophe.desagre@uclouvain.be. Phone: +32 (0) 65 323 519.

‡Louvain School of Management, Université Catholique de Louvain, 151 Chaussée de Binche, 7000 Mons (Belgium). Email: floris.laly@uclouvain.be. Phone: +32 (0) 65 323 517. Corresponding author.

§IESEG School of Management (Lille Catholic University) and LEM-CNRS (UMR 8179).

¶Louvain School of Management & CORE, Université Catholique de Louvain, 151 Chaussée de Binche, 7000 Mons (Belgium). Email: mikael.petitjean@uclouvain.be. Phone: +32 (0) 65 323 381.

‖We would like to thank NASDAQ OMX and Frank Hatheway for providing the data.

1 Introduction

High-frequency traders (HFTs) have conquered most of the exchanges around the world and now represent a large portion of both the overall trading activity and liquidity on these exchanges. As a consequence, HFTs now make up about 55% of trading volume in the U.S. equity markets, 66% in treasury markets and up to 80% in foreign-exchange markets (Foucault and Moinas, 2018).

Many benefits have been attributed to the arrival of HFTs in electronic markets such as a decrease in spreads (Stoll, 2014; Jovanovic and Menkveld, 2016; Boehmer, Fong, and Wu, 2018), a decrease in transaction costs (Jones, 2013), a decrease in short-term volatility (Hagströmer and Nordén, 2013; Hasbrouck and Saar, 2013) as well as an improvement in price efficiency (Carrion, 2013; Brogaard, Hendershott, and Riordan, 2014). However, HFTs are also suspected of harming modern automated markets from time to time (Hasbrouck & Saar, 2013) and especially during so called flash events. On the one hand, flash events can take the form of flash crashes, which are sudden, extreme and short-lived price jumps (up or down) that partially or totally self-correct within a maximum of a few dozen of minutes. Many flash crashes have occured in the past few years, whether in equities (2010 and 2015 flash crashes), in treasuries (2014 flash crash), in currencies (2015 flash crash in the U.S. dollar; 2016 flash crash in the pound sterling), in commodities (as reported by Massad, 2015) or more recently in cryptocurrencies. On the other hand, flash events can take the form of mini flash crashes (shorter in duration when compared to flash crashes), which are of particular interest in this paper. Johnson et al (2013) report 18,520 MFCs in the U.S. equity markets alone between January 3, 2006 and February 3, 2011, i.e. about 15 MFCs per day, while Félez-Vinas (2018) reports 2,105 MFCs on the Spanish IBEX35 exchange and 947 MFCs on the Italian FTSE MIB exchange over the period November 2012-April 2013, i.e. about 17 MFCs per day on the Italian exchange and about 8 MFCs per day on the Spanish exchange. Both studies put together reveal that MFCs are a global phenomenon and not just a U.S. phenomenon. It transpires from the literature that MFCs have coincided with the emergence of high-frequency trading. For example, Leal et al. (2016) find that when the market (in their model) is populated with low-frequency traders only, flash crashes do not occur. And while a number of papers examine the role played by HFTs during flash crashes (Kirilenko et al., 2017; Aldrich et al., 2017, Menkveld & Yueshen, n.d.), the literature regarding the role played by HFTs during mini flash crashes, however, is still limited.

Our objective is to fill in this gap by addressing the following questions: (1) Do HFTs trigger mini flash crashes ? (2) Do HFTs exacerbate the crash phase during mini flash crashes ? (3) Do HFTs lead the price recovery right after the crash ?

Within the body of literature focusing on flash crashes, findings regarding the role played by HFTs point to similar conclusions. Kirilenko et al. (2017), in their study of the behavior of HFTs in the E-mini S&P500 futures market during the May 6, 2010 Flash Crash, find that HFTs did not cause the crash (the large automated selling program of a mutual fund later on identified as Waddell & Reed did). Moreover, they find that HFTs did not fundamentally change their trading pattern during the flash crash. However, Kirilenko et al. note that just before the market was paused for 5 seconds, HFTs liquidated 2,000 contracts accumulated earlier (in an already illiquid market), coinciding with significant additional price declines (the most abrupt price decline of the crash). On the contrary, traditional market makers (NHFTs) did not liquidate their accumulated inventory. In that sense, HFTs contributed to the Flash Crash. Their findings are confirmed by the empirical study led by Aldrich et al., (2017). As for Menkveld & Yueshen (n.d.), they also conclude that the crash cannot be attributed to the mutual fund alone and that it is rather the result of the interaction between market participants that degenerated into a flash crash. In a recent working paper focusing on 65 flash crashes identified in 37 CAC40 stocks over the year 2013, Bellia et al. (2018) find that (1) about 70% of flash crashes are triggered by HFTs, (2) HFTs exacerbate the magnitude of the crash at its climax by selling more as the crash unfolds, and (3) HFTs do not contribute to the price recovery but keep selling aggressively.

Within the body of literature dealing with mini flash crashes, Golub and Keane (2011) find that most MFCs occur at the beginning and at the end of the trading session, that is to say during periods of relative stress and that the first price change (the first tick in the series of ten consecutive ticks or more) during the crash is always the largest one. Golub et al. (2012), in a second study, argue that MFCs are caused by market fragmentation, which is contradicted by the findings of Félez-Vinas (2018) who concludes that market fragmentation reduces the number of MFCs and speeds up the recovery phase. Finally, Aquilina et al (2018) identify extreme events that they call "mini flash crashes/rallies" on the UK equity market over the period January 2014-June 2015. Contrary to Nanex (2010), Aquilina et al (2018) define mini flash crashes as large price movements that revert within a short time window and during which there is a high level of traded volume. More specifically, for such an event to be considered a mini flash crash, the authors argue that the price movement has to (1) exceed a pre-determined threshold (three times the average realized variation of the previous 20 trading days), (2) revert at least 50% within a short time frame (less than 30 minutes) and (3) trigger high levels of trading volumes (levels higher than the top 5% percentile of the distribution). Finally, the authors exclude all events with a price change below 1% whose magnitude they consider "too small", which leaves them with a total of 40 flash events whose drop or spike duration can last up to 10 minutes. Contrary to the authors, we believe that flash events whose drop or spike duration is more than a few seconds should not be considered as mini flash crashes but rather as flash crashes. As such, we discard their methodology from the following analysis.

This paper empirically investigates the behavior of HFTs around mini flash crashes, also known as flash equity failures (Nanex LLC, 2010) or ultrafast extreme events (Johnson et al, 2013). As such, mini flash crashes share similarities with extreme price movements (Brogaard et al, 2018) even though they are in fact different phenomena. Indeed, EPMs are not exactly the same as MFCs since their existence is determined ex-post (statistically) based on the 99.9th percentile of the return distribution and their duration (10 second-intervals) exceeds the couple of second time intervals of MFCs. More specifically, we carry out an event study of sub-two-second price jumps[1] on a sample of large cap, medium cap and small cap Nasdaq equities over a two-year period (2008, 2009 and week of February 22-26, 2010).

We define mini flash crashes as sudden, extreme and very short-lived price jumps that partially or totally self-correct within a few seconds. As such, mini flash crashes are smaller versions of flash crashes. However, and as pointed out by Johnson et al (2013), they differ from flash crashes in two aspects. First, mini flash crashes only last for about one second (up to two seconds) instead of many minutes in the case of flash crashes, which does not allow ample time for human intervention. Second, the extremely rapid speed and recovery of most MFCs suggests that they are unlikely caused by exogeneous news arrival.

Figure 1 presents an example of a down mini flash crash that impacted Apple (APPL) stock on January 22, 2008. The crash from top to bottom occurred in 1180 milliseconds (starting at 10:02:24.100 and ending at 10:02:25.280), the (transaction) price collapsing 5.06% within the interval before bouncing back to its initial level.

[1] Our analysis focuses on MFCs whose duration is comprised between 1 and 2 seconds, with a base case of 1.5 seconds.

Figure 1: **Example of a down MFC on APPL - January 22, 2008**
The data are from Tickdata.

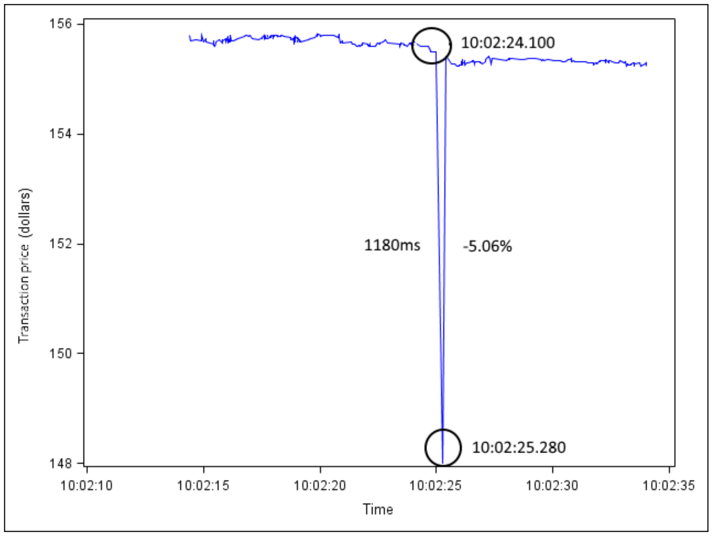
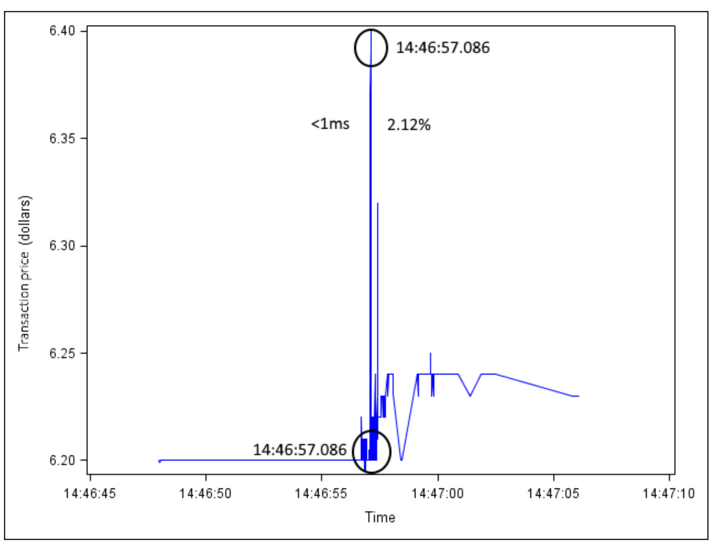


Figure 2 presents an example of an up mini flash crash that impacted Alcoa (AA) stock on March 16, 2009. The crash from bottom to top occurred in less than 1 millisecond (starting at 10:46:57.086 and ending at 10:46:57.086), the (transaction) price jumping 2.12% within this sub-millisecond interval before reverting back to its initial level.

Figure 2: **Example of an up MFC on AA - March 16, 2009**
The data are from Tickdata.



Our paper directly relates to the paper by Brogaard et al. (2018), complementing their initial research. Brogaard et al. (2018) focus on so called extreme price movements or EPMs, which are computed as returns above the 99.9th percentile of the return distribution (with median absolute returns of 0.436% at 10 second-intervals) on the 40 largest stocks of the Nasdaq HFT database. We use the same database in this paper except we use an additional week of data.[2] Their EPMs are based on pre-specified time intervals. The base case interval is 10 seconds, implying that all the identified EPMs have a duration of 10 seconds. In total, they identify 45,200 EPMs at 10 second-intervals on the 40 large cap stocks of the Nasdaq HFT database over the period 2008-2009.

We could question whether Brogaard et al. (2018) really capture extreme price movements since they potentially never identify the top and bottom of the price movements. This is particularly important since price movements can be extremely short-lived (a few milliseconds). Even when they change the time interval from 1s, 5s, 10s, 30s, to 1 minute, they never identify EPMs based on tops and bottoms within each interval.

Why is it potentially misleading? Let us consider a down crash for example. If there is a larger drop within the interval (larger than the drop between the open and close prices of the interval), then Brogaard et al. underestimate the down crash and include some price correction in their EPM. In other words, resiliency may already be occurring during the interval of the EPM. When there is a larger drop just after the close price of the interval, Brogaard et al. also underestimate the down crash and pollute the next interval since the crash has not ended yet. Consequently, while Brogaard et al. (2018) detect extreme price movements endogeneously, using the 10-second returns in the 99.9th percentile according to magnitude as well as according to the Lee and Mykland's (2012) jump-detection methodology, we detect MFCs exogenesouly replicating Nanex MFC detection algorithm (2010) and complementing this methodology with two EPM identification methods (explained in more details in Section 2). As such, we are able to both focus on proven extreme price movements (rather than on approximated ones) and at much higher frequencies (as advocated by Brogaard et al, 2018).

[2]While Brogaard et al. (2018) use data from 2008 and 2009 only, we include the week of February 22-26, 2010.

2 Data, MFC identification and summary statistics

2.1 Data and sample

In this section, we present the way we build our stock sample using both the Nasdaq HFT dataset and a Tickdata dataset. Our sample includes 74 stocks from the Nasdaq HFT dataset (32 large, 30 medium and 12 small cap stocks) to be compared to the original Nasdaq HFT dataset which is composed of 120 stocks (40 large, 40 medium and 40 small cap stocks).

First, we use tick-by-tick data timestamped to the millisecond on trades from Tickdata for 74 stocks (out of the 120 stocks) included in the Nasdaq HFT dataset (see list in Appendix). The data from Tickdata account for all transactions on U.S. stock exchanges[3] for the 74 stocks at our disposal. Second, we use tick-by-tick data timestamped to the millisecond on trades from Nasdaq OMX for the same 74 stocks. The data from Nasdaq OMX account for transactions on Nasdaq and NYSE only. A flag on Nasdaq data enables us to know if the liquidity demander/supplier is a high-frequency trader (HFT) or a non-high frequency trader (NHFT). We gather both datasets so as to get a clear picture regarding (1) the true magnitude of mini flash crashes (which would be underestimated if measured on Nasdaq only) and (2) the trading activity of HFTs (and NHFTs) on Nasdaq around these crashes. We make a clear distinction between MFCs occuring on Nasdaq and MFCs not occuring on Nasdaq. In more details, we consider MFCs for which the proportion of transactions occuring on Nasdaq during the crash represents at least 50% of all transactions on U.S. exchanges thus filtering out MFCs that are not prevalent on Nasdaq. We then test the robustness of our results by replicating our analysis on the full MFC sample (proportion of Nasdaq transactions comprised between 0% and 100%) and on a restricted sample where all transactions occur exclusively on Nasdaq during the crash (proportion of Nasdaq transactions = 100%).

Moreover, we use a window of three trading days around each MFC in our sample, discarding all the other days, so that the final sample period is shorter than the initial reference (Nasdaq) sample period which runs from January 1, 2008 to December 31, 2009, and includes the week of February 22-26, 2010. The Tickdata data contain all trades done on all U.S. stock exchanges around the MFC days of our stock sample and the Nasdaq OMX data contain all trades done on the Nasdaq exchange around the MFC days of our stock sample, ignoring trades that were executed at the opening, closing, during intraday crosses or trades executed in dark pools. The

[3]NYSE American (formerly AMEX and NYSE MKT), NASDAQ OMX BX (Boston), National Stock Exchange (Cincinnati), ISE (International Securities Exchange), DirectEdge A, DirectEdge X, Chicago, NYSE, NYSE Arca, NASDAQ, CBOE, NASDAQ OMX PSX (Philadelphia), BATS Y-Echanges, BATS.

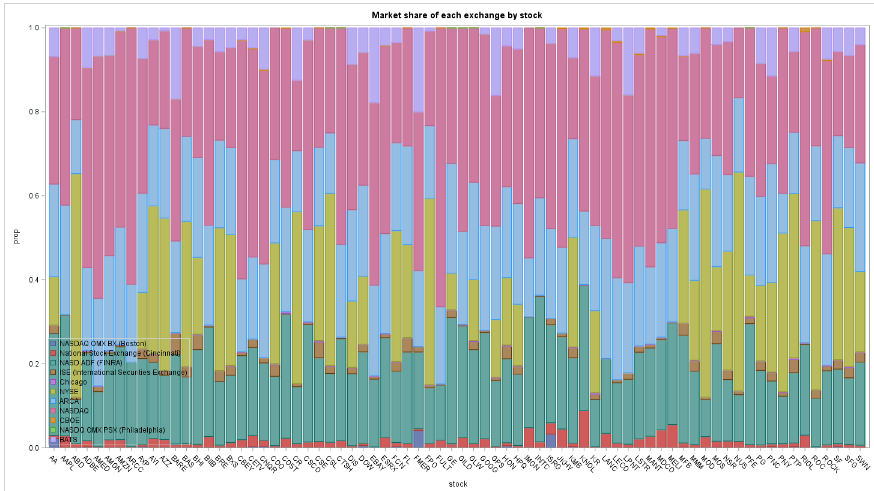
liquidity status of each trade, i.e. the characteristics of the liquidity demander and supplier, is included as well as the type of trader involved (HFT or NHFT).

We use trade prices instead of midquotes in order to take into account the full magnitude of each crash (from top to bottom or from bottom to top) and focus on trading activity from 9:30 a.m to 4:00 p.m. ET so as to take into account the full trading period from the opening to the closing of the U.S. equity market. We later include a dummy variable to take into account "extreme hours" corresponding to the first five and last five minutes of the trading session. The isolation of the first five and last five minutes of trading activity is meant to see if MFCs that could result from the price distortion caused by the increased volatility specific to the opening and closing phases present similar or different characteristics with the rest of our MFC sample.

2.2 Nasdaq market share

Over the reference (Nasdaq) sample period, and when considering the 74 stocks of our sample, Nasdaq is by far the U.S. exchange with the highest market share (36.02%) based on the number of trades, which makes a focus on Nasdaq all the more relevant. The market share of each U.S. exchange is presented in Figure 3.

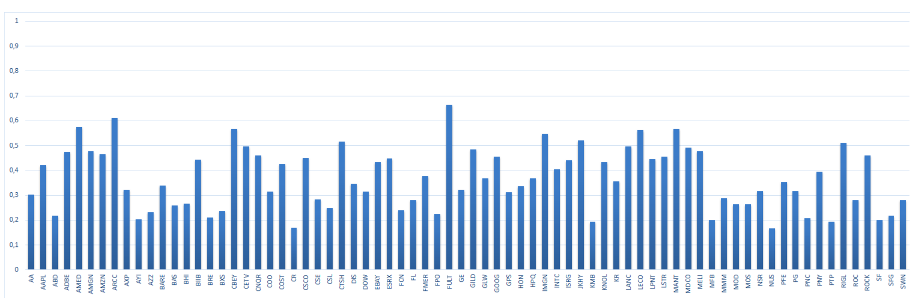
Figure 3: **Market share of each U.S. Stock Exchange over the reference sample period for the 74 stocks of our sample**



The figure plots the market share of each U.S. stock exchange for each one of the 74 sample stocks based on the number of trades. Overall, Nasdaq is the stock exchange with the highest market share over the sample period (36.02%), followed by ARCA (23.96%), Finra NASD ADF (23.82%), NYSE (9.29%), BATS (3.88%), ISE (1.39%), National Stock Exchange (1.28%), NASDAQ OMX BX (0.19%), Chicago (0.15%), CBOE (0.03%) and NASDAQ OMX PSX (0.00%). The data are from Tickdata.

Nasdaq market share based on the number of trades is more than 50% in 10 of the 74 sample
1 stocks. The market share of Nasdaq for each of the 74 sample stocks is presented in Figure 4.

Figure 4: **Nasdaq market share per stock over the reference sample period**



The figure plots the market share of Nasdaq for each one of the 74 sample stocks based on the number of trades. Nasdaq has more than a 50% market share in the following 10 stocks: AMED, ARCC, CBEY, CTSH, FULT, IMGN, JKHY, LECO, MANT, RIGL. The data are from Tickdata.

2.3 Additional comments on the Nasdaq HFT dataset

The HFT dataset we use in this paper is the so-called Nasdaq HFT dataset, provided by Nasdaq OMX to academics under a non-disclosure agreement. The dataset, which consists in a stratified sample of 120 U.S. stocks[4] representing different market capitalization groups (large, medium and small) on two listing venues (Nasdaq and NYSE), is the same as in Brogaard et al. (2018), Brogaard, Hendershott & Riordan (2017), Hirschey (2018), Gerig (2015), Carrion (2013), O'Hara, Yao, Ye (2014) and Brogaard (2010).

A limitation of the Nasdaq HFT database, as pointed out in previous papers, is the fact that while Nasdaq has identified many HFTs within the database, based on different metrics, large integrated firms (acting as HFTs but not only) as well as HFTs routing their orders through large integrated firms have been excluded from the database due to the impossibility for Nasdaq to identify them precisely. As such, the 26 high-frequency trading firms of the database can be considered as "independent proprietary trading firms" (Brogaard, Hendershott & Riordan, 2017) or pure HFTs. Still, the database enables us to zoom on the trading activity of these pure HFTs on Nasdaq around mini flash crashes, keeping in mind Nasdaq is by far the dominant U.S. exchange in the 74 stocks of our sample.

[4]The sample was selected by professors Terrence Hendershott and Ryan Riordan.

2.4 Structure of the Nasdaq stock market

Nasdaq, which was originally an acronym standing for "National Association of Securities Dealers Automated Quotations" (NASDAQ), was founded in 1971 by the National Association of Securities Dealers (NASD) to become the first electronic stock market in the world. It then separated from the NASD and started operating as a national securities exchange in 2006. Over the 2008-2010 period (reference sample period), the Nasdaq stock market had an average 25.3% market share in U.S. equities based on consolidated volume alone and an average 52% total market share based on consolidated volume, internalization and other trade reporting.[5]

Trading on Nasdaq occurs continuously from 9:30 a.m. to 4:00 p.m., Eastern Time. The opening and closing crosses are determined through the use of both an opening and a closing book. To do so, Nasdaq accepts order types that are only executable within the opening or closing books. At 9:30 a.m. ET, the opening cross is initiated so that both the opening book and the Nasdaq continuous book are brought together to create a single Nasdaq opening cross (opening bid and ask quote). The same occurs at 4:00 p.m. ET, the closing cross is initiated so that both the closing book and the Nasdaq continuous book are brought together to create a single Nasdaq closing cross (closing bid and ask quote). The opening cross provides the Nasdaq Official Opening Price (NOOP) and the closing cross provides the Nasdaq Official Closing Price (NOCP). If a stock does not have an opening cross, the NOOP is determined by the first last-sale eligible trade reported at or after 9:30 a.m., when regular trading hours begin. In the same way, if a stock does not have a closing cross, the last last-sale eligible trade reported prior to 4:00 p.m. is used as the NOCP.

The Nasdaq stock market relies on a price-display-time priority model. First, better priced orders are presented for execution so that a buy order at $50 is ranked ahead of a buy order at $49.99. In the same way, a sell order at $49.99 is ranked ahead of a sell order at $50. Second, displayed orders are ranked ahead of hidden orders. Thus, a displayed order entered after a hidden order is ranked ahead of the hidden order all else equal. Third, better timed orders are presented for execution first so that a buy order received at 09:50:00:001 is ranked ahead of a buy order received at 09:50:00:002. Fourth, any price improvement resulting from an order execution is given to the liquidity taker. For example, if a buy order is positioned in the limit order book (LOB) at $50 and a sell order priced at $49.90 arrives in the LOB, the order is executed at $50 and the $0.10 price improvement benefits the liquidity taker (the seller in this case).[6]

[5] U.S. equities market share statistics provided by Nasdaq.
[6] Source: Nasdaq website.

2.5 Identification of mini flash crashes

We identify mini flash crashes by replicating Nanex MFC detection algorithm (2010) and complement our study with a parallel analysis on extreme price movements (EPMs).

First, based on the initial definition of an MFC provided by Nanex (Nanex, 2010), we identify price movements with at least 10 tick movements in the same direction before ticking in the other direction (ignoring trades with no tick change), all within 1.5 seconds[7] (based on variable intervals lower than 1.5 seconds) and with a price change exceeding 0.8%, in the same way as Golub et al. (2012) and Johnson et al. (2013). This method is meant to capture price jumps that meet all the conditions of an MFC (tick rule, time rule and price change rule).

Second, following Brogaard et al. (2018), we remove the tick rule and the price change rule and instead identify price movements exceeding the 99.9th percentile of the absolute return distribution by stock, computed from open to close, all within 1.5 seconds[8] (based on fixed 1.5-second intervals). This method is meant to capture price jumps that are extreme and that still meet one out of the three conditions of an MFC (time rule).

Since returns are computed from open to close using this methodology and since this may not fairly represent the magnitude of the true crash, we provide an alternative third methodology by identifying price movements exceeding the 99.9th percentile of the absolute return distribution by stock, computed from high to low or from low to high depending on the direction of the crash, so as to take into account the true crash, all within 1.5 seconds[9] (based on fixed 1.5-second intervals). This method is again meant to capture price jumps that are extreme and that still meet one out of the three conditions of an MFC (time rule), while also taking into account the true magnitude of the crash.

Finally, all three methods are computed using alternative time intervals: 1 second and 2 seconds respectively (versus a base case of 1.5 seconds), thus following Brogaard et al. (2018), who use 1-second intervals as a robustness check to capture EPMs on the U.S. equity market, Nanex (2010), who use 1.5-second intervals to capture MFCs on the U.S. equity market, and Félez-Vinas (2018), who uses a maximum of 2-second intervals to identify MFCs on the Spanish and Italian equity markets (Spanish IBEX and Italian FTSE MIB indices).

[7] We use a variable sub-1.5-second interval here.
[8] We use a fixed 1.5-second interval here.
[9] We use a fixed 1.5-second interval here.

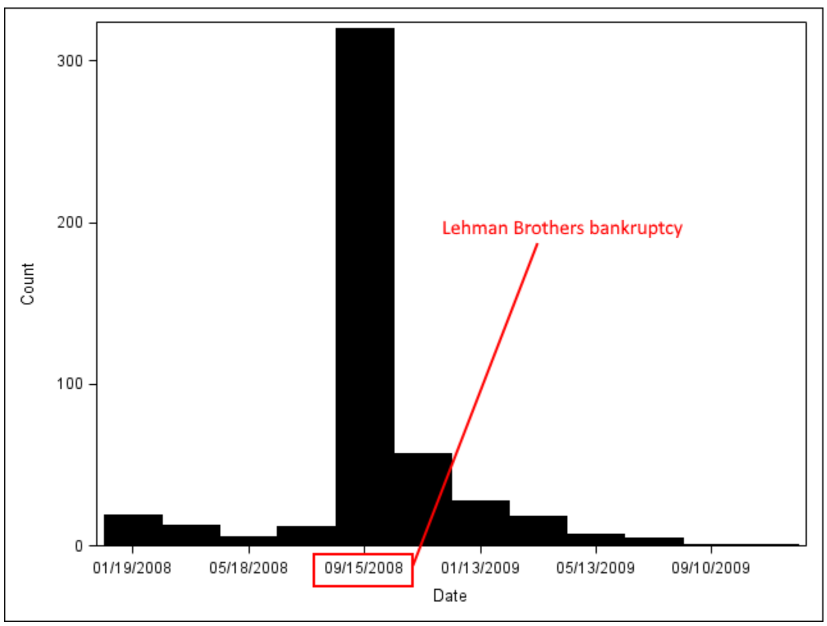
2.6 Descriptive statistics

We present descriptive statistics based on the three different identification methods used in this paper for MFCs, open-close EPMs and high-low EPMs.

2.6.1 Daily and intraday distribution of mini flash crashes

Figure 5 reports the daily distribution of MFCs following the Nanex identification method. Most MFCs occur around the bankruptcy of Lehman Brothers on September 15, 2008. We count 44 MFCs in the week and 316 MFCs in the month following the news of the bankruptcy respectively thus representing 9.03% and 64,89% of all MFCs in the sample based on the Nanex identification method.

Figure 5: Daily distribution of MFCs (Nanex)

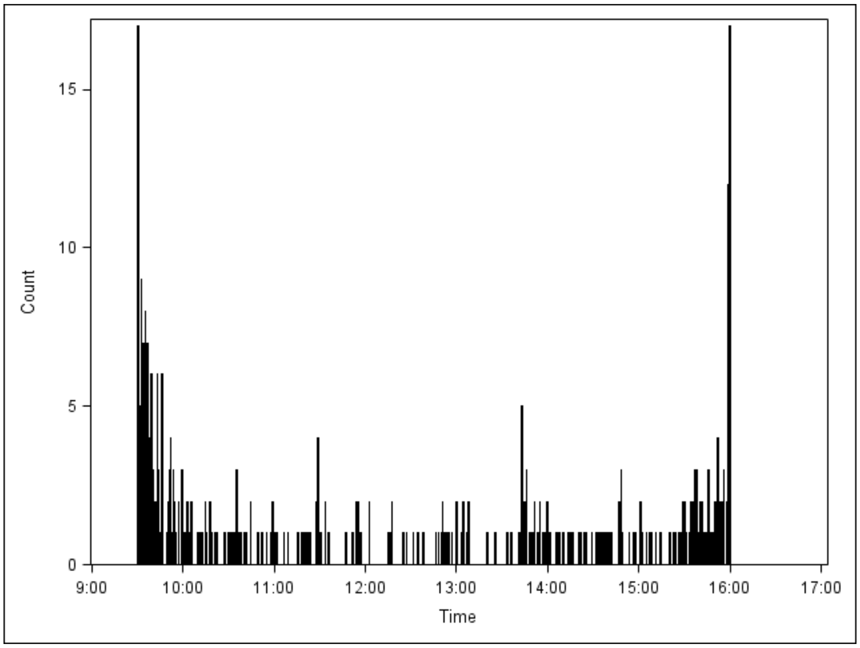


The figure plots the daily distribution of MFCs over the sample period following the Nanex identification method. The data are from Tickdata.

Figure 6 reports the intraday distribution of MFCs following the Nanex identification method. Most MFCs occur at the beginning and at the end of the trading day, which is consistent with previous studies (Golub and Keane, 2011; Brogaard et al., 2018). In more details, more than a quarter of MFCs (27.31%) occur in the first and last five minutes of the trading day (15,81% of

MFCs occur in the first five minutes while 11,50% of MFCs occur in the last five minutes) and more than half of MFCs (55,44%) occur in the first and last half hour of the trading day (33,26% of MFCs occur in the first half hour and 22,18% of MFCs occur in the last half hour) so that the overall intraday distribution is U-shaped.

Figure 6: Intraday distribution of MFCs (Nanex)



The figure plots the intraday distribution of MFCs over the sample period following the Nanex identification method. The data are from Tickdata.

2.6.2 Mini flash crashes

We first investigate the general characteristics of our MFC sample. We carry out a similar investigation on our EPM samples.

Table 1 reports the summary statistics for the full sample (Panel A), the sample of mini flash crashes (MFCs) following the Nanex identification method (Panel B), the sample of open-close EPMs following Brogaard et al. (2018) (Panel C) and the sample of high-low EPMs, which is an alternative identification method we propose (Panel D) using a variable sub-1.5-second interval for Panel B and a fixed 1.5-second interval for Panels A, C and D.

Panel B reports the descriptive statistics for the sample of 510 MFCs following the Nanex identification method. As expected, the absolute return, trading activity (as measured by total trades, share volume and dollar volume), and spread (as measured by quoted spread and relative spread) are substantially larger during MFCs than during the average 1.5-second interval of the full sample (Panel A). The mean absolute MFC return is 1.668% while the full sample mean absolute return is 0.0142%. As such, the mean absolute MFC return is more than 117 times larger than the mean absolute full sample return. Trading activity also appears to be materially higher during MFCs. Indeed, while about 4 trades are executed on average per 1.5 second within the full sample (Panel A) we note that about 104 trades are executed on average per 1.5 second during MFCs, i.e. there are on average 26 times more trades per 1.5-second interval during MFCs. In the same way, share volume and dollar volume are 63 times and 60 times higher respectively during MFCs based on the mean. Indeed, while 87,000 shares ($36,595.73) are traded on average per 1.5 second-interval over the full sample, 5,446,660 shares ($2,191,106) are traded on average per 1.5 second-interval during MFCs. Moreover, the quoted spread is almost 19 times (1.5 times) higher and the relative spread is more than 3 times (1.5 times) higher during MFCs based on the mean (median) when compared to the full sample. Finally, the liquidity picture as represented by depth and dollar depth would let us think that liquidity is slightly increased during MFCs, however the test of means indicates that one cannot reject the hypothesis that both means are equal since the difference between the MFC sample mean and the full sample mean is not statistically significant.

The difference between the mean and median in our sample vs the mean and median in Brogaard et al. (2018) is striking. Indeed, while our sample takes into account sub-1.5-second MFCs (vs 10-second EPMs in Brogaard et al.), the mean (median) absolute MFC return is about 3.5 times (2.6 times) the mean (median) absolute extreme price movement return in Brogaard et al. (2018). Moreover, the mean (median) quoted spread during the MFCs of our sample is 17 times (2 times) higher than the mean (median) quoted spread in Brogaard et al. (2018) and the mean (median) relative spread is 5.6 times (1.4 times) the mean (median) relative spread in Brogaard et al. (2018). As such, we proceed further and compare our variable-1.5-second MFC sample to our fixed 1.5-second open-close EPM sample (Panels B and C of Table 1). We observe that the mean (median) absolute return of the MFC sample is about 3 times (2.8 times) the mean (median) absolute return of the open-close EPM sample and we note that trading activity is far more intense during MFCs than during EPMs with total trades (on all U.S. stock exchanges) during the crash being 3.3 times (5.7 times) higher during MFCs than during open-close EPMs and total trades on Nasdaq during the crash being 3.1 times (6.25 times) higher during MFCs than during open-close EPMs. Moreover, we note that the proportion of HFTs in activity is higher during MFCs than during EPMs, the proportion of HFT trades, HFT shares and HFT volume representing 61.95% (66.67%), 54.92% and 54.93% respectively during MFCs versus

51.64% (50.05%), 48.63% and 48.63% respectively during open-close EPMs based on the mean (median). As for share and dollar volume, they are 3.8 times and 4.3 times higher respectively based on the mean during MFCs when compared to open-close EPMs. Last but not least, we observe that the quoted spread is 4.9 times higher during MFCs when compared to open-close EPMs based on the mean while the relative spread is 1.1 times higher during MFCs based on the mean but 50% lower based on the median.

Consistent with previous studies (Nanex, 2010; Golub and al., 2012; Johnson et al., 2013; Brogaard et al., 2018), we find that the proportions of down and up MFCs are very close in the sample based on the Nanex identification method, down and up MFCs representing 47.84% and 52.16% of MFCs respectively.

We note that some MFCs within our sample occur in less than 1 millisecond (presented as 0 ms in the table), which is in line with the fact the fastest HFTs act within 5 ms while other relatively fast traders act at speed levels of 50 ms to 150 ms (Scholtus et al., 2014).

We perform a hypothesis test for difference of means in order to check whether the trading activity statistics (total trades, share volume, dollar volume) and the liquidity statistics (quoted spread, relative spread, depth, dollar depth) are statistically different between the MFC and EPM samples and the full sample. The results are included in Table 1.

Table 1: Summary statistics of MFCs, open-close EPMs, high-low EPMs

	Minimum	Median	Mean	Maximum	Std Dev
Panel A: Full sample					
	Minimum	Median	Mean	Maximum	Std Dev
Absolute return, %	0	0	0.0142	68.40	0.0590
Total trades (all U.S. exchanges)	0	0	3.86	1,205	11.48
Total trades (Nasdaq)	0	0	1.57	880	5.45
Proportion HFT trades (Nasdaq)	0	0	0.1821	1	0.3611
Proportion HFT volume (Nasdaq)	0	0	0.1784	1	0.3591
Proportion HFT volume (Nasdaq)	0	0	0.1784	1	0.3591
Share volume	0	0	870.45	28,368,232	10,705.63
Dollar volume	0	0	36,595.73	1,458,142,722	456,960.03
Depth	2	6	23.67	100,816	99.32
Dollar Depth	6.34	152.30	363.93	2,225,817	1805.47
Quoted spread, $	0.01	0.02	0.04	113.91	0.11
Relative spread, %	0.002	0.06	0.12	56	0.77
N	29,390,400				
Panel B: MFC Sample					
	Minimum	Median	Mean	Maximum	Std Dev
Crash duration, ms (MFC specific)	0	60	211	1487	338
Total tick change (MFC specific)	10	12	13.3	34	3.84
Absolute return, %	0.8011	1.118	1.668	35.29	2.31
Total trades (all U.S. exchanges)	3	68	104.34***	883	110.76
Total trades (Nasdaq)	0	25	44.64***	672	60.32
Proportion HFT trades (Nasdaq)	0	0.6667	0.6195***	1	0.2824
Proportion HFT shares (Nasdaq)	0	0.5667	0.5492***	1	0.2955
Proportion HFT volume (Nasdaq)	0	0.5661	0.5493***	1	0.2954
Share volume	500	20,972	54,466***	3,138,737	160,530
Dollar volume	13,343	729,535	2,191,106***	146,022,000	8,962,257
Depth	2	7.57	26.62	1,505.41	91.87
Dollar depth	20.42	177.16	519.87	36,063	2,317.95
Quoted spread, $	0.01	0.0267	0.7558**	113.91	7.43
Relative spread, %	0.0093	0.0874	0.4294**	42.89	2.72
N	510				
Panel C: open-close EPM sample					
	Minimum	Median	Mean	Maximum	Std Dev
Absolute return, %	0.1358	0.3951	0.5750***	68.40	1.48
Total trades (all U.S. exchanges)	0	12	31.45***	1,054	55.28
Total trades (Nasdaq)	0	4	14.28***	672	28.95
Proportion HFT trades (Nasdaq)	0	0.6	0.5164***	1	0.40
Proportion HFT shares (Nasdaq)	0	0.5	0.4863***	1	0.4024
Proportion HFT volume (Nasdaq)	0	0.5005	0.4863***	1	0.4024
Share volume	0	2,000	14,437***	9,871,210	117,456
Dollar volume	0	65,351	503,162***	265,820,000	3,504,283
Depth	2	5.09	13.17***	1,824	45.34
Dollar depth	6.58	121.68	239.10***	75,122	897.78
Quoted spread, $	0.01	0.0507	0.1545***	113.91	1.3465
Relative spread, %	0,0024	0,1728	0,3935***	56.21	0.9465
N	29,424				
Panel D: high-low EPM sample					
	Minimum	Median	Mean	Maximum	Std Dev
Absolute return, %	0.1357	0.4566	0.5476***	30.35	0.5679
Total trades (all U.S. exchanges)	0	13	32.38***	1,054	56.10
Total trades (Nasdaq)	0	4	14.76***	672	29.80
Proportion HFT trades (Nasdaq)	0	0.6	0.5146***	1	0.3979
Proportion HFT shares (Nasdaq)	0	0.5	0.4833***	1	0.3993
Proportion HFT volume (Nasdaq)	0	0.5002	0.4833***	1	0.3993
Share volume	0	2,100	13,646***	28,368,232	200,069.6
Dollar volume	0	67,902	531,484***	1,458,142,722	9,886,679
Depth	2	4.79	12.32***	1,824	45.94
Dollar depth	6.58	117.57	232.05***	75,122	890.87
Quoted spread, $	0.01	0.0652	0.1723***	113.91	1.36
Relative spread, %	0,0024	0,2222	0,4355***	42.89	0.84
N	29,427				

The table reports descriptive statistics for the full sample (Panel A), the sample of mini flash crashes (MFCs) following the Nanex identification method (Panel B), the sample of open-close EPMs following Brogaard et al. (2018) identification method (Panel C) and the sample of high-low EPMs (Panel D), which is a proposed alternative method to Brogaard et al. (2018). We use a sub-1.5-second variable interval for Panel B and a fixed 1.5-second interval for Panels A, C and D. All data are from Tickdata except Total trades and Proportion of HFT trades, HFT shares and HFT volume which are from Nasdaq. The mean of Absolute return, Total trades, Depth, Dollar volume, Share volume, Quoted spread and Relative spread is computed in two steps. First, we compute the P50 by stock so as to obtain one observation by stock. Second, we compute the mean of P50 across the 74 stocks of our sample. As an example, the mean of Total trades in panel A is the mean across stocks of the median number of trades within a 1.5-second interval. Note that Share volume represents round lots of 100 share units and that absolute returns in Panel B are returns computed over the MFC interval and not over the 1.5-second interval.The table also reports univariate tests for means differences. Asterisks ***, ** and * indicate statistical significance at the 1%, 5% and 10% significance levels.

2.6.3 Mini flash crashes by market capitalization

Among the 120 Nasdaq stocks of the Nasdaq HFT database, 74 stocks[10] suffer at least one mini flash crash over the sample period based on the Nanex identification method. The original dataset being made of 40 large cap stocks, 40 midcap stocks and 40 small cap stocks, we are able to determine the proportion of stocks impacted by MFCs in each subsample (Table 2). We observe that large cap and midcap stocks are the ones that mainly suffer MFCs with 80% of the large cap subsample and 75% of the midcap subsample being hit by MFCs over the sample period. The proportion of small cap stocks suffering MFCs is far lower with 30% of the small cap subsample being hit by MFCs over the sample period. Within our sample of 74 stocks impacted by MFCs, about 43% are large cap stocks, about 41% are midcap stocks and about 16% are small cap stocks (Table 2).

Table 2: MFC stocks by market capitalization

Market Cap	Number of stocks suffering at least one MFC	Percentage of stocks
Large	32	43.24%
Medium	30	40.54%
Small	12	16.22 %
Total	74	100.00%

The table reports the number of stocks suffering at least one MFC by market capitalization following the Nanex identification method (2010).

We find a total of 510 MFCs over the sample period. Among these MFCs, 83% impact large cap stocks, about 15% impact midcap stocks and about 2% only impact small cap stocks based on the Nanex identification method (Table 3). As such, we observe that the overwhelming majority of MFCs occur on large cap stocks (sometimes within the same day, the same hour or even within the same minute), while this does not prevent medium and small cap stocks from also being impacted by MFCs (though to a smaller extent). As emphasised in the literature, MFCs mostly occur on the most liquid assets.

Table 3: MFCs by market capitalization

Market Cap	Number of MFCs	Percentage of stocks
Large	423	82.94%
Medium	76	14.90%
Small	11	02.16%
Total	510	100.00%

The table reports MFCs by market capitalization following the Nanex identification method (2010).

[10] We only focus on these 74 stocks in the study.

We note several interesting characteristics when focusing on market capitalization (Table 4). First, we note that large, medium and small cap stocks are all stricken by lightning fast MFCs (MFCs with a crash duration < 1ms). However, there does not seem to exist any pattern related to crash duration since crash duration rather seems random within the different market cap groups. Second, we note that the total tick change during MFCs (consecutive down ticks during down crashes and consecutive up ticks during up crashes), from the start of the crash to the end of the crash, is very similar for each market cap category with a mean comprised between 12 and 14 tick movements and a median comprised between 11.5 and 12 tick movements over the crash period. Third, the lower the market capitalization of the stock, the higher the absolute return during MFCs, with mean (median) absolute returns of 1.81% (1.09%), 1.94% (1.30%) and 2.03% (1.44%) for large, mid and small cap stocks respectively. Fourth, based on the proportion of HFT trades on Nasdaq, it appears that HFTs are more active on large cap stocks (mean of 0.66 and median of 0.73) than on midcap stocks (mean of 0.48 and median of 0.49) and small cap stocks (mean of 0.39 and median of 0.28) during MFCs. The pattern is similar when considering the proportion of HFT shares and HFT volume during the crash. Finally, we note that the relative spread observed during MFCs is on average lower on large cap stocks (0.41%) than on midcap (0.50%) and small cap stocks (0.61%), which would tend to indicate that the lower the market cap the higher the impact of MFCs on relative bid-ask spreads.

We also report summary statistics on open-close EPMs (Table 5) and high-low EPMs (Table 6) in order to compare MFC characteristics to EPM characteristics. We note that the proportion of HFT activity within the different market capitalization groups tends to decrease far more for EPMs than for MFCs with a proportion of HFT trades, HFT shares and HFT volume that even falls to 0% for medium and small cap stocks during EPMs based on the median versus about 40% and 25% for medium and small cap stocks respectively during MFCs. As a consequence, one may extrapolate that HFTs cannot be responsible for extreme price movements occuring in medium and small cap stocks since they do not play any active role in them. On the contrary, HFTs keep playing an active role in medium and small cap stocks during MFCs, even though their activity is reduced when compared to large cap stocks.

Table 4: Summary statistics of MFCs by market cap

Large cap					
	Minimum	Median	Mean	Maximum	Std Dev
Crash duration, ms (MFC specific)	0	60	203	1,487	323
Total tick change (MFC specific)	10	12	13.50	34	4.01
Absolute return, %	0.8011	1.0823	1.6136	35.29	2.36
Total trades (all U.S. exchanges)	3	77	116.50	883	117.03
Total trades (Nasdaq)	0	30.50	50.86	672	65.17
Proportion HFT trades (Nasdaq)	0	0.7273	0.6575	1	0.2696
Proportion HFT shares (Nasdaq)	0	0.6129	0.5776	1	0.2891
Proportion HFT volume (Nasdaq)	0	0.6128	0.5776	1	0.2891
Share volume	500	26,119	62,316	3,138,737	175,007
Dollar volume	14,973	934,890	2,557,532	146,022,200	9,799,648
Depth	2	8.35	27.37	1,505	95.48
Dollar depth	49.56	195.22	548.38	36,063	2,431
Quoted spread, $	0.01	0.0227	0.8843	113.91	8.15
Relative spread, %	0.0093	0.0077	0.4117	42.89	2.96
N	423				
Medium cap					
	Minimum	Median	Mean	Maximum	Std Dev
Crash duration, ms (MFC specific)	0	60	265	1,457	421
Total tick change (MFC specific)	10	11.5	12.29	25	2.74
Absolute return, %	0.8104	1.2891	1.9158	14.32	2.16
Total trades (all U.S. exchanges)	3	37.50	45.53	225	34.50
Total trades (Nasdaq)	0	14	19.92	130	21.63
Proportion HFT trades (Nasdaq)	0	0.4900	0.4820	1	0.2780
Proportion HFT shares (Nasdaq)	0	0.3742	0.4498	1	0.2933
Proportion HFT volume (Nasdaq)	0	0.3745	0.4496	1	0.2934
Share volume	1,100	9,786	15,170	113,328	17,627
Dollar volume	13,343	275,121	403,301	2,877,091	499,764
Depth	2	5.41	24.61	483	76.85
Dollar depth	20.42	91.83	415.78	11,576	1,775
Quoted spread, $	0.0126	0.0674	0.1305	0.995	0.1815
Relative spread, %	0.0476	0.2783	0.5017	6.90	0.8637
N	76				
Small cap					
	Minimum	Median	Mean	Maximum	Std Dev
Crash duration, ms (MFC specific)	0	30	114	826	240
Total tick change (MFC specific)	10	12	12.73	19	2.72
Absolute return, %	0.8092	1.44	2.03	4.83	1.21
Total trades (all U.S. exchanges)	20	34	43.36	132	31.54
Total trades (Nasdaq)	4	10	22.18	105	29.38
Proportion HFT trades (Nasdaq)	0	0.2762	0.3891	0.967	0.3162
Proportion HFT shares (Nasdaq)	0	0.25	0.3583	0.9867	0.3104
Proportion HFT volume (Nasdaq)	0	0.2488	0.3588	0.9855	0.3099
Share volume	3,400	14,400	24,107	139,574	38,950
Dollar volume	38,462	230,889	452,469	2,818,876	791,671
Depth	2.06	5.63	11.62	51	16.03
Dollar depth	23.57	66.69	140.76	505.71	181
Quoted spread, $	0.0320	0.0772	0.1142	0.6017	0.1637
Relative spread, %	0.1464	0.3281	0.6113	2.88	0.7721
N	11				

The table reports descriptive statistics for the subsamples of mini flash crashes (MFCs) according to market capitalization (large, medium, small) following the Nanex identification method. All data are from Tickdata except Total trades and Proportion of HFT trades, HFT shares and HFT volume which are from Nasdaq. The mean of Absolute return, Total trades, Depth, Dollar volume, Share volume, Quoted spread and Relative spread is computed in two steps. First, we compute the P50 by stock so as to obtain one observation by stock. Second, we compute the mean of P50 across the 74 stocks of our sample. As an example, the mean of Total trades in panel A is the mean across stocks of the median number of trades within a 1.5-second interval.

Table 5: Summary statistics of open-close EPMs by market cap

Large cap					
	Minimum	Median	Mean	Maximum	Std Dev
Absolute return, %	0.1545	0.3774	0.5514	68.40	1.5135
Total trades (all U.S. exchanges)	0	26	48.72	1,054	66.23
Total trades (Nasdaq)	0	12	24.07	672	36.24
Proportion HFT trades (Nasdaq)	0	0.7778	0.6902	1	0.3054
Proportion HFT shares (Nasdaq)	0	0.7205	0.6426	1	0.3274
Proportion HFT volume (Nasdaq)	0	0.7206	0.6426	1	0.3274
Share volume	0	4,779	23,251	9,871,210	152,506
Dollar volume	0	165,345	822,141	265,818,454	4,539,566
Depth	2	6.05	17.72	1,824	55.33
Dollar depth	24.70	159.04	326.35	75,122	1,113
Quoted spread, $	0.01	0.0311	0.1373	113.91	1.7421
Relative spread, %	0.0024	0.0920	0.1773	42.89	0.6536
N	17,270				
Medium cap					
	Minimum	Median	Mean	Maximum	Std Dev
Absolute return, %	0.1358	0.4081	0.5978	68.15	1.3790
Total trades (all U.S. exchanges)	0	4	7.63	225	12.27
Total trades (Nasdaq)	0	1	3.19	130	6.02
Proportion HFT trades (Nasdaq)	0	0	0.3402	1	0.4082
Proportion HFT shares (Nasdaq)	0	0	0.3281	1	0.4098
Proportion HFT volume (Nasdaq)	0	0	0.3281	1	0.4098
Share volume	0	500	2,117	452,584	9,423
Dollar volume	0	13,848	56,508	17,789,123	332,403
Depth	2	4.27	7.05	1,051	26.28
Dollar depth	10.90	93.07	127.96	21,089	453.57
Quoted spread, $	0.01	0.1100	0.1901	8.19	
Relative spread, %	0.0226	0.3937	0.6258	18.40	0.8145
N	9,607				
Small cap					
	Minimum	Median	Mean	Maximum	Std Dev
Absolute return, %	0.2155	0.4582	0.6497	59.34	1.6138
Total trades (all U.S. exchanges)	0	2	4.23	352	9.93
Total trades (Nasdaq)	0	1	1.94	349	8.57
Proportion HFT trades (Nasdaq)	0	0	0.2218	1	0.3724
Proportion HFT shares (Nasdaq)	0	0	0.2200	1	0.3762
Proportion HFT volume (Nasdaq)	0	0	0.2200	1	0.3762
Share volume	0	300	1,144	336,765	8,224
Dollar volume	0	4,414	25,044	9,277,725	220,098
Depth	2	3.95	5.48	181	7.37
Dollar depth	6.58	53.36	67.08	1,152	60.55
Quoted spread, $	0.01	0.09	0.1376	4.30	0.1886
Relative spread, %	0.0398	0.5485	0.9821	0.5621	0.0202
N	2,547				

The table reports descriptive statistics for the subsamples of open-close EPMs according to market capitalization (large, medium, small) following Brogaard et al (2018). All data are from Tickdata except Total trades and Proportion of HFT trades, HFT shares and HFT volume which are from Nasdaq. The mean of Absolute return, Total trades, Depth, Dollar volume, Share volume, Quoted spread and Relative spread is computed in two steps. First, we compute the P50 by stock so as to obtain one observation by stock. Second, we compute the mean of P50 across the 74 stocks of our sample. As an example, the mean of Total trades in panel A is the mean across stocks of the median number of trades within a 1.5-second interval.

Table 6: Summary statistics of high-low EPMs by market cap

Large cap					
	Minimum	Median	Mean	Maximum	Std Dev
Absolute return, %	0.1752	0.4465	0.5128	30.35	
Total trades (all U.S. exchanges)	0	27	49.82	1,054	67.14
Total trades (Nasdaq)	0	12	24.82	672	37.39
Proportion HFT trades (Nasdaq)	0	0.7692	0.6834	1	0.3062
Proportion HFT shares (Nasdaq)	0	0.7065	0.6340	1	0.3277
Proportion HFT volume (Nasdaq)	0	0.7065	0.6340	1	0.3277
Share volume	0	4,900	21,833	28,368,232	260,700
Dollar volume	0	172,724	869,266	1,458,142,722	12,891,615
Depth	2	5.53	16.47	1,824	56.45
Dollar depth	26.67	151.22	315.72	75,122	1,102
Quoted spread, $	0.0100	0.0400	0.1570	113.91	1.7594
Relative spread, %	0.0024	0.115	0.2118	42.89	0.6555
N	17,272				
Medium cap					
	Minimum	Median	Mean	Maximum	Std Dev
Absolute return, %	0.1357	0.4630	0.5851	17.70	0.4974
Total trades (all U.S. exchanges)	0	4	8.41	225	13.53
Total trades (Nasdaq)	0	1	3.50	150	6.64
Proportion HFT trades (Nasdaq)	0	0	0.3457	1	0.4069
Proportion HFT shares (Nasdaq)	0	0	0.3332	1	0.4085
Proportion HFT volume (Nasdaq)	0	0	0.3332	1	0.4085
Share volume	0	600	2,189	452,584	9,156
Dollar volume	0	14,951	57,847	17,789,123	326,388
Depth	2	4.13	6.67	1,051	25.42
Dollar depth	10.90	92.98	124.84	21,089	466
Quoted spread, $	0.0100	0.1259	0.2064	10.81	0.3129
Relative spread, %	0.0317	0.4468	0.6797	18.40	0.8337
N	9,607				
Small cap					
	Minimum	Median	Mean	Maximum	Std Dev
Absolute return, %	0.2269	0.5073	0.6423	9.50	0.52
Total trades (all U.S. exchanges)	0	2	4.62	352	10.25
Total trades (Nasdaq)	0	1	2.08	349	8.67
Proportion HFT trades (Nasdaq)	0	0	0.2293	1	0.3731
Proportion HFT shares (Nasdaq)	0	0	0.2258	1	0.3754
Proportion HFT volume (Nasdaq)	0	0	0.2258	1	0.3753
Share volume	0	300	1,351	630,200	13,109
Dollar volume	0	4,626	27,588	9,277,725	249,206
Depth	2	3.93	5.47	181	7.13
Dollar depth	6.58	55.54	69.16	1,152	62.39
Quoted spread, $	0.0113	0.1062	0.1474	1.9971	0.1478
Relative spread, %	0.0398	0.6109	1.03	9.07	1.34
N	2,548				

The table reports descriptive statistics for the subsamples of high-low EPMs according to market capitalization (large, medium, small). All data are from Tickdata except Total trades and Proportion of HFT trades, HFT shares and HFT volume which are from Nasdaq. The mean of Absolute return, Total trades, Depth, Dollar volume, Share volume, Quoted spread and Relative spread is computed in two steps. First, we compute the P50 by stock so as to obtain one observation by stock. Second, we compute the mean of P50 across the 74 stocks of our sample. As an example, the mean of Total trades in panel A is the mean across stocks of the median number of trades within a 1.5-second interval.

2.6.4 Mini flash crashes by sector

We then investigate the distribution of MFCs by sector. To do so, we use the GICS sector classification. We then proceed with a similar analysis on EPMs.

Within our sample of 74 stocks, we note that 10 out of 11 sectors (GICS classification) are impacted by MFCs, based on the Nanex identification method. The three sectors which are the most impacted by MFCs over the sample period are information technology (22.55%), financials (18.43%) and industrials (17.45%), representing altogether 58.43% of sectors impacted by MFCs in our sample (Table 7).

Table 7: MFCs by sector

Ranking	GICS Sector	Proportion HFT trades (median)	Number of MFCs	Proportion
1	Information Technology	0.7188	115	22.55%
2	Financials	0.6250	94	18.43%
3	Industrials	0.8333	89	17.45%
4	Materials	0.7304	61	11.96%
5	Healthcare	0.5000	59	11.57%
6	Consumer Discretionary	0.5376	47	09.22%
7	Consumer Staples	0.6363	20	03.92%
8	Energy	0.6281	18	03.53%
9	Real Estate	0.5000	5	00.98%
10	Utilities	0.3571	2	00.39%
11	Telecommunications Services	0.0000	0	00.00%
	Total		510	100.00%

The table reports MFCs by sector following the Nanex identification method.

When looking more closely at the first three sectors impacted by MFCs (Table 8) we note that, in the same way as for market capitalizations, there does not seem to exist any pattern related to crash duration. As such, crash duration does not seem to be related to either company size or the sector to which the stock belongs to.

Table 8: Summary statistics of MFCs by sector

	Minimum	Median	Mean	Maximum	Std Dev
Information Technology					
Crash duration, ms (MFC specific)	0	3	150	1,314	272
Total tick change (MFC specific)	10	13	14.64	35	5.19
Absolute return, %	0.815	1.24	3.14	36.84	6.34
Total trades (all U.S. exchanges)	8	95	130	883	114.69
Total trades (Nasdaq)	1	52.5	78.90	672	88.5
Total HFT trades (Nasdaq)	0	0.72	0.62	1	0.28
Share volume	1,700	31,996	55,442	376,297	64,400
Dollar volume	20,642	1,509,295	3,914,375	120,167,804	12,336,819
Depth	2	9.87	17.21	134.01	20.75
Dollar depth	36.88	235.64	503.87	9,116	1,010
Quoted spread, $	0.01	0.02	3.10	113.92	15.40
Relative spread, %	0.0093	0.0737	1.17	42.89	5.59
N	110				
Industrials					
	Minimum	Median	Mean	Maximum	Std Dev
Crash duration, ms (MFC specific)	0	3	80	1,202	163
Total tick change (MFC specific)	10	12	13.22	28	3.59
Absolute return, %	0.809	1.05	1.22	4.21	0.55
Total trades (all U.S. exchanges)	15	92	164.48	815	170.35
Total trades (Nasdaq)	2	18	54.42	427	73.89
Total HFT trades (Nasdaq)	0.06	0.83	0.75	1	0.24
Share volume	3,000	39,600	96,398	1,073,971	155,753
Dollar volume	85,167	942,925	2,390,986	29,489,243	4,107,374
Depth	2.77	17.79	67.83	1,505	191.59
Dollar depth	34.52	248.95	988.11	36,063	3,976
Quoted spread, $	0.01	0.01	0.05	0.95	0.11
Relative spread, %	0.0235	0.0595	0.1582	2.28	0.2697
N	92				
Financials					
	Minimum	Median	Mean	Maximum	Std Dev
Crash duration, ms (MFC specific)	0	60	183	1,440	312
Total tick change (MFC specific)	10	12	12.49	28	2.81
Absolute return, %	0.801	1.06	1.28	5.42	0.61
Total trades (all U.S. exchanges)	3	46	70	432	70.57
Total trades (Nasdaq)	0	9	22.10	173	32
Total HFT trades (Nasdaq)	0	0.63	0.58	1	0.31
Share volume	500	13,903	30,500	275,790	45,323
Dollar volume	14,973	401,948	897,355	8,089,938	1,425,109
Depth	2.68	6.92	10.58	111	13.22
Dollar depth	20.42	132.22	195.24	958.85	173.21
Quoted spread, $	0.01	0.03	0.05	0.42	0.06
Relative spread, %	0.0283	0.0982	0.1481	0.6321	0.1292
N	87				

The table reports descriptive statistics for the three sectors which are the most impacted by MFCs in our sample. All data are from Tickdata except Total trades and Total HFT trades which are from Nasdaq. The mean of Absolute return, Total trades, Depth, Dollar volume, Share volume, Quoted spread and Relative spread is computed in two steps. First, we compute the P50 by stock so as to obtain one observation by stock. Second, we compute the mean of P50 across the 74 stocks of our sample. As an example, the mean of Total trades in panel A is the mean across stocks of the median number of trades within a 1.5-second interval.

Comparing MFCs by sector to open-close EPMs and high-low EPMs by sector, we observe that the ranking is the same so that the sectors that suffer the highest number of price jumps, whether represented by mini flash crashes or extreme price movements, are the same, i.e. information technology, financials and industrials. However, we observe that HFTs are not active (based on the proportion of HFT trades) in the telecommunication services and utilities sectors during EPMs so that EPMs can be observed even though HFTs are not involved in any trading activity. On the contrary, MFCs are never observed in our sample when HFTs are not involved in any trading activity (the minimum median proportion of HFT trades observed is 0.35 for MFCs in the utilities sector while it is 0.00 for EPMs in the real estate and telecommunication services sectors).

Table 9: Open-close EPMs by sector

Ranking	GICS Sector	Proportion HFT trades (median)	Number of EPMs	Proportion
1	Information Technology	0.7164	5,453	18.54%
2	Financials	0.5000	5,137	17.46%
3	Industrials	0.4667	4,465	15.17%
4	Healthcare	0.5384	3,687	12.53%
5	Materials	0.8181	3,246	11.03%
6	Consumer Discretionary	0.5000	3,076	10.45%
7	Consumer Staples	0.6558	1,830	06.22%
8	Energy	0.8383	1,202	04.09%
9	Real Estate	0.0000	734	02.49%
10	Telecommunications Services	0.0000	344	01.17%
11	Utilities	0.0000	250	00.85%
	Total		29,424	100.00%

The table reports open-close EPMs by sector.

Table 10: High-low EPMs by sector

Ranking	GICS Sector	Proportion HFT trades (median)	Number of EPMs	Proportion
1	Information Technology	0.6923	5,456	18.54%
2	Financials	0.5000	5,137	17.46%
3	Industrials	0.5000	4,465	15.17%
4	Healthcare	0.5555	3,687	12.53%
5	Materials	0.8125	3,246	11.03%
6	Consumer Discretionary	0.5000	3,076	10.45%
7	Consumer Staples	0.6000	1,829	06.22%
8	Energy	0.8000	1,203	04.09%
9	Real Estate	0.0000	734	02.49%
10	Telecommunications Services	0.0000	344	01.17%
11	Utilities	0.0000	250	0.85%
	Total		29,427	100.00%

The table reports high-low EPMs by sector.

2.6.5 Mini flash crashes by U.S. Exchange

We finally investigate the distribution of trades among the different U.S. stock exchanges during MFCs (Table 11). Based on the Nanex identification method, trades during MFCs occur predominantly on three exchanges: NYSE (40.67%), Nasdaq (29.43%) and Arca (25.01%). In total, 95.11% of MFC trades occur on these three exchanges together while only 4.89% of MFC trades occur on other exchanges. We note that up MFC trades occur mainly on NYSE (46.55%), followed by Nasdaq (26.18%) and Arca (21.87%) while down MFC trades are more evenly spread between NYSE (34.27%), Nasdaq (32.98%) and Arca (28.44%).

Table 11: Proportion of MFC trades by U.S. Exchange

All MFCs		
Ranking	Exchange	Proportion
1	NYSE	40.67%
2	Nasdaq	29.43%
3	Arca	25.01%
4	NASD FINRA	03.12%
5	ISE	00.91%
6	Cincinnati	00.40%
7	Bats	00.35%
8	Others	00.11%
Total		100.00%
Down MFCs		
Ranking	Exchange	Proportion
1	NYSE	34.27%
2	Nasdaq	32.98%
3	Arca	28.44%
4	NASD FINRA	02.52%
5	ISE	0.88%
6	Cincinnati	0.41%
7	Bats	0.45%
8	Others	0.05%
Total		100.00%
Up MFCs		
Ranking	Exchange	Proportion
1	NYSE	46.55%
2	Nasdaq	26.18%
3	Arca	21.87%
4	NASD FINRA	03.67%
5	ISE	0.94%
6	Cincinnati	0.39%
7	Bats	0.26%
8	Others	0.14%
Total		100.00%

The table reports the proportion of MFCs by U.S. exchange following the Nanex identification method. The data are from Tickdata.

The main takeaway of section 2.6 is that MFCs embody a far bigger shock (in terms of absolute return) than EPMs while they are also characterized by both enhanced trading activity (in terms of total trades, dollar volume and share volume) and enhanced liquidity dynamics (in terms of depth, dollar depth, quoted spread and relative spread) when compared to EPMs. Moreover, while HFTs are always in activity during MFCs (based on the proportion of HFT trades, HFT shares and HFT volume) they are sometimes inactive during EPMs, which by definition may question their specific role in both types of crashes.

3 Trading activity and liquidity measures

We both measure trading activity and liquidity through several metrics that we describe below.

3.1 Trading activity measures

We measure trading activity through three different metrics: number of trades (*Total Trades*), dollar volume (*Dollar Volume*) and share volume (*Share Volume*).

The first metric, *Total Trades*, measures the total number of trades per stock per interval ($\leq$1.5 seconds in Panel B; =1.5 seconds in Panels A, C and D). The second metric, *Dollar Volume*, measures the average dollar volume per stock per interval. The third metric, *Share Volume*, measures the average share volume per stock per interval.

We also measure the trading activity of HFTs through three different metrics: proportion of HFT trades (*Proportion HFT trades*), proportion of HFT shares (*Proportion HFT shares*), Proportion of HFT volume (*Proportion HFT volume*).

3.2 Liquidity measures

We measure the order book liquidity through four different metrics: depth at best prices (*Depth*), dollar depth at best prices (*Dollar Depth*), bid-ask spread (*Quoted Spread*) and relative bid-ask spread (*Relative Spread*).

We define *Depth* as:

$$Depth_{i,t} = BestAskSize_{i,t} + BestBidSize_{i,t} \tag{3.1}$$

where $BestBidSize_{i,t}$ and $BestAskSize_{i,t}$ correspond to the share volume resting at best prices on both sides of the order book.

We define *Dollar Depth* as:

$$DollarDepth_{i,t} = BestAskPrice_{i,t}BestAskSize_{i,t} + BestBidPrice_{i,t}BestBidSize_{i,t} \tag{3.2}$$

where $BestBidSize_{i,t}$ and $BestAskSize_{i,t}$ correspond to the share volume resting at best prices on both sides of the order book.

We define *Quoted Spread* as:

$$QuotedSpread = BestAskPrice_{i,t} - BestBidPrice_{i,t} \tag{3.3}$$

where $BestBidPrice_{i,t}$ and $BestAskPrice_{i,t}$ represent the best bid and ask prices of the order book respectively, and where i represents the stock and t represents the time of the quote update.

We define *Relative Spread* as:

$$RelativeSpread = (BestAskPrice_{i,t} - BestBidPrice_{i,t})/Midquote \tag{3.4}$$

where $BestBidPrice_{i,t}$ and $BestAskPrice_{i,t}$ represent the best bid and ask prices of the order book respectively, and where i represents the stock and t represents the time of the quote update. The Midquote is defined as the average of the best bid and best ask prices.

3.3 Correlation matrix

We present the correlation matrix of the main variables in Table 12. We observe a very low positive correlation between MFC and both open-close EPMs and high-low EPMs.

Table 12: Correlation matrix of the main variables

	MFC	EPM^{OC}	EPM^{HL}	HFT^{NET}	$NHFT^{NET}$	$Depth$	$DollarDepth$	SV	DV	QS	RS
MFC	1.0000	0.05695	0.06805	-0.00108	0.00108	0.00012	0.00036	0.02085	0.01964	0.02611	0.00165
EPM^{OC}	0.05695	1.0000	0.63953	0.00398	-0.00398	-0.00335	-0.00219	0.04012	0.03232	0.03157	0.01115
EPM^{HL}	0.06805	0.63953	1.0000	0.00152	-0.00152	-0.00362	-0.00232	0.03778	0.03429	0.03653	0.01288
HFT^{NET}	-0.00108	0.00398	0.00152	1.0000	-1.0000	-0.02459	-0.01019	-0.01218	0.01238	0.00546	-0.00078
$NHFT^{NET}$	0.00108	-0.00398	-0.00152	-1.0000	1.000	0.02459	0.01019	0.01218	-0.01238	-0.00546	0.00078
$Depth$	0.00012	-0.00335	-0.00362	-0.02459	0.02459	1.0000	0.87774	0.10009	0.04150	-0.05123	-0.01454
$DollarDepth$	0.00036	-0.00219	-0.00232	-0.01019	0.01019	0.87774	1.0000	0.08335	0.04994	-0.01195	-0.01424
SV	0.02085	0.04012	0.03778	-0.01218	0.01218	0.10009	0.08335	1.0000	0.74330	-0.01424	-0.00708
DV	0.01964	0.03232	0.03429	0.01238	-0.01238	0.04150	0.04994	0.74330	1.0000	0.01272	-0.00738
QS	0.02611	0.03157	0.03653	0.00546	-0.00546	-0.05123	-0.01195	-0.01424	0.01272	1.0000	0.64193
RS	0.00165	0.01115	0.01288	-0.00078	0.00078	-0.01454	-0.01424	-0.00708	-0.00738	0.64193	1.0000

The table reports the correlation matrix of the main variables.

3.4 Methodology

We proceed in three steps. First we capture HFT and NHFT trading activity around MFCs via a measure of trade imbalance which enables us to determine the role played by both market participants during the phase preceeding the crash (pre-crash), during the crash (crash), as well as during the recovery phase (post-crash).

Second, we use an autoregressive model to study the behavior of HFTs during the crash. The regressions are run on an MFC subsample whose proportion of transactions on Nasdaq is set to 0.5, meaning that at least 50% of transactions during the crash occur on Nasdaq. We standardize all non-dummy variables at the stock level and we run the regressions with stock fixed effects.

Third, we run probit regressions to measure the probability of having a crash as a function of lagged values of HFT^{NET}, absolute return, share volume and relative spread. We standardize all non-dummy variables at the stock level. Results are presented in the following section.

We run parallel regressions on open-close EPMs and high-low EPMs so as to compare MFCs to EPMs.

4 Empirical results

4.1 HFT activity around mini flash crashes

We capture HFT and NHFT trading activity around mini flash crashes via a measure of trade imbalance following Brogaard et al. (2018). We compute trade imbalance for both HFTs and NHFTs in the following way:

$$(N)HFT^{D} = (N)HFT^{D^{+}} - (N)HFT^{D^{-}} \tag{4.1}$$

$$(N)HFT^{S} = (N)HFT^{S^{+}} - (N)HFT^{S^{-}} \tag{4.2}$$

where *(N)*HFT^D represents the liquidity demanded by (N)HFTs, where *(N)*HFT^S represents the liquidity supplied by (N)HFTs, where *(N)*HFT^{D^+} and *(N)*HFT^{S^+} represent the liquidity demanded and supplied in the direction of the MFC (down for a down crash and up for an up crash), and where *(N)*HFT^{D^-} and *(N)*HFT^{S^-} represent the liquidity demanded and supplied in the opposite direction of the MFC (up for a down crash and down for an up crash).

Net imbalance (HFT^{NET} and $NHFT^{NET}$) informs us on the direction of net trading activity vis-à-vis the MFC direction. A positive net imbalance implies trading activity in the direction of the MFC on an aggregated basis (vicious behavior). On the contrary, a negative net imbalance implies trading activity in the opposite direction of the MFC on an aggregated basis (virtuous behavior).

We compute net imbalance of both types of traders at times t_{-2}, t_{-1}, $t_{(crash)}$, t_{+1}, t_{+2} for MFCs, open-close EPMs and high-low EPMs using fixed 1.5-second intervals. In particular, we focus on time intervals t_{-1}, $t_{(crash)}$, t_{+1}, which represent the pre-crash, crash and post-crash phases respectively. Time intervals t_{-2} and t_{+2} are included in the analysis so as to see if the pattern observed at t_{-1} and t_{+1} are persistent.

4.2 Trade imbalance around MFCs

Table 13: Trade imbalance around MFCs - 1.5-second interval

	t_{-2}	t_{-1}	$t_{(crash)}$	t_{+1}	t_{+2}
HFT^{NET}	169.1*	176.8	-169.9	184.5*	-67.4383
HFT^{D}	-7.3680	121.9	224.0*	130.9	-34.5130
HFT^{S}	176.5**	54.9208	-394.0***	53.5415	-32.9253
$NHFT^{NET}$	-169.1*	-176.8	169.9	-184.5*	67.4383
$NHFT^{D}$	391.0	300.8**	3376.0***	28.2120	760.6
$NHFT^{S}$	-560.1**	-477.6***	-3206.1***	-212.7*	-693.2

The table reports trade (share volume) imbalance around mini flash crashes (MFCs) computed from Nasdaq. Time interval t is the sub-1.5-second interval corresponding to the crash. We also report the trade imbalance figures for 1 second and 2 seconds prior to the crash ($t_{\_2}$, $t_{\_1}$) and for 1 second and 2 seconds following the crash (t_{+1}, t_{+2}). HFT^{NET} ($NHFT^{NET}$) is the difference between HFT^{D} and HFT^{S} ($NHFT^{D}$ and $NHFT^{S}$). $HFT^{NET} = -NHFT^{NET}$ and vice versa. Asterisks ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels.

4.2.1 Before the crash

Just before the crash, at time interval $t_{\_1}$, we note that HFT^{S} and HFT^{D} are positive ($HFT^{S} = 54.9208$ and $HFT^{D} = 121.9$), which would imply HFTs trading in the direction of the crash (triggering the crash?). We note for example that HFTs demand 2.2 times more liquidity in the direction of the crash to come than they supply (121.9/54.9208). As for NHFTs, their liquidity demand in the direction of the crash to come only represents 0.63 times their liquidity supply in the opposite direction of the crash. However, HFT^S and HFT^D are not statistically significant at time interval t_{-1}, which implies that one cannot reject the null hypothesis according to which:

- the number of shares demanded by HFTs in the direction of the crash (HFT^D) is zero in the population;
- the number of shares supplied by HFTs in the direction of the crash (HFT^S) is zero in the population;
- the total number of shares (both demanded and supplied) by HFTs in the direction of the crash (HFT^{NET}) is zero in the population.

NHFTs appear to be the ones that are (significantly) active during this phase, NHFTs being the ones that consume significantly more liquidity in the direction of the future MFC ($NHFT^D$=300.8**) while supplying significantly more liquidity in the opposite direction of the future MFC ($NHFT^S$=-477.6***). In net values however, no statistical conclusion can be reached at this stage.

Moreover, we note that the consumption of liquidity in the direction of the crash by NHFTs is 2.5 times (=300.8/121.9) higher than the consumption of liquidity in the direction of the crash by HFTs. Overall, NHFTs supply 8.7 times (=477.6/54.9208) more liquidity than HFTs during this time interval. While NHFTs supply liquidity in the opposite direction of the crash (-477.6***), we note that HFTs supply liquidity in the direction of the crash (54.9208). Statistically, however, one cannot reject the null hypothesis that HFT^S=0 and HFT^D=0.

Time interval t_{-2} is included in the analysis so as to check if the trading pattern that we observe at t_{-1} is persistent over time. The behaviors of both types of traders, HFTs and NHFTs, are more or less similar from t_{-2} to t_{-1}, especially for NHFTs. That said, we note a higher statistical significance for HFTs.

4.2.2 During the crash

During the crash, at time t, we note that HFT^{NET} ($NHFT^{NET}$) is negative (positive), which would imply NHFTs (as a group) generating the trade imbalance and HFTs (as a group) counteracting on this imbalance. However, since HFT^{NET} and $NHFT^{NET}$ are not statistically significant at time interval t, we cannot provide any definitive statistical conclusion regarding the global activity of HFTs and NHFTs in the direction of the crash.

In more details, we note that NHFTs are the ones that are particularly active during the crash. They consume 11.2 times (=3376/300.8) more liquidity in the direction of the crash than during the prior time interval (to be compared to the fact that they only supply 6.7 times more liquidity in the opposite direction of the crash). NHFTs would thus tend to exacerbate the crash. As for HFTs, they appear to change their behavior in terms of liquidity supply, from supplying liquidity in the direction of the crash at time interval t_{-1} to supplying liquidity in the opposite direction of the crash at time interval t, with a 817% (=(394+54.9208)/54.9208) liquidity supply change in the opposite direction of the crash. The liquidity demand change is smaller with an increase lower than 100% when compared to the prior time interval and statistically significant at 10% only. One can thus suspect that HFTs, due to their shift in liquidity supply during the crash, reduce the magnitude of the crash.

4.2.3 After the crash

In line with the findings of Bellia et al. (2018), NHFTs appear to be the ones that contribute to resiliency (even though at a 10% level of significance only), helping the stock price to recover. They mainly offer liquidity in the opposite direction of the crash (to the bid for a down crash and to the ask for an up crash). In a way, NHFTs seem to adopt the same behavior as HFTs during the crash, mimicking the liquidity supply of HFTs in the prior time interval. In the phase that immediately follows the crash, HFTs (as a group) appear to be the ones generating the trade imbalance while NHFTs (as a group) appear to be the ones counteracting on this imbalance and driving the price recovery.

We then proceed with a similar analysis on both open-close EPMs and high-low EPMs to see if similar patterns can be observed during the pre-crash, crash and post-crash phases of extreme price movements.

4.3 Trade imbalance around open-close EPMs

Table 14: Trade imbalance around open-close EPMs - 1.5-second interval

	t_{-2}	t_{-1}	$t_{(crash)}$	t_{+1}	t_{+2}
HFT^{NET}	-18.1346**	-17.3686	23.3090	58.5018***	-6.0885
HFT^{D}	-24.4327**	3.9137	129.7***	49.0000***	-15.2832
HFT^{S}	6.2981	-21.2824**	-106.4***	9.5018*	9.1947
$NHFT^{NET}$	18.1346**	17.3686	-23.3090	-58.5018***	6.0885
$NHFT^{D}$	-6.8413	57.8667***	209.1***	6.8021	63.8673
$NHFT^{S}$	24.9760	-40.4980**	-232.4***	-65.3039***	-57.7788

The table reports trade (share volume) imbalance around open-close extreme price movements (open-close EPMs) computed from Nasdaq. Time interval t is the 1.5-second interval corresponding to the crash. We also report the trade imbalance figures for 1 second and 2 seconds prior to the crash (t_{-2}, t_{-1}) and for 1 second and 2 seconds following the crash (t_{+1}, t_{+2}). HFT^{NET} ($NHFT^{NET}$) is the difference between HFT^{D} and HFT^{S} ($NHFT^{D}$ and $NHFT^{S}$). $HFT^{NET} = -NHFT^{NET}$ and vice versa. Asterisks ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels.

4.3.1 Before the crash

Just before the crash, at time interval t_{-1}, we note that HFT^{S} is significantly negative while HFT^{D} is not significantly positive ($HFT^{S} = -21.2824{*}{*}$ and $HFT^{D} = 3.9137$), which would imply HFTs supplying liquidity in the opposite direction of the crash (not triggering the EPM). Indeed, one cannot reject the null hypothesis according to which the number of shares demanded by HFTs in the direction of the crash (HFT^{D}) is zero in the population. We note for example that HFTs supply 5.4 times more liquidity in the opposite direction of the crash to come than they demand (-21.2824/3.9137).

NHFTs appear to be the ones that are (significantly) active during this phase, in the same way as during MFCs. While NHFTs supply about twice (=40.4980/21.2824) more liquidity in the opposite direction of the crash than HFTs during this time interval, they are the ones who consume significantly more liquidity in the direction of the future EPM ($NHFT^{D}$=57.8667***) while supplying significantly less liquidity in the opposite direction of the future EPM ($NHFT^{S}$=-40.4980***). In net values however, no statistical conclusion can be reached.

Time interval t_{-2} is included in the analysis so as to check if the trading pattern that we observe at t_{-1} is persistent over time. The behaviors of both types of traders, HFTs and NHFTs, are more or less similar from t_{-2} to t_{-1} when we focus on net values only. They are more heterogeneous when we focus on demand and supply only.

4.3.2 During the crash

During the crash, at time t, we note that HFT^{NET} ($NHFT^{NET}$) is positive (negative), which would imply HFTs (as a group) exacerbating the liquidity imbalance during the crash phase and NHFTs (as a group) counteracting on this liquidity imbalance. However, since HFT^{NET} and $NHFT^{NET}$ are not statistically significant at time interval t, we cannot provide any definitive statistical conclusion regarding the global activity of HFTs and NHFTs in the direction of the crash.

In more details and similar to MFCs, we note that NHFTs are the ones that are particularly active during the crash. They consume about 4 times more liquidity in the direction of the crash (=209.1/57.8667) than during the prior time interval. As for HFTs, they appear to maintain their behavior in terms of liquidity supply, from supplying liquidity in the opposite direction of the crash at time interval t_{-1} (HFT^{S}=-21.2824**) to supplying even more liquidity in the opposite direction of the crash at time interval t (HFT^{S}=-106.4***). However, they appear to change their behavior in terms of liquidity demand, from demanding little (potentially zero) liquidity at time interval t_{-1} to significantly demanding liquidity in the direction of the crash at time interval t (HFT^{D}=129.7***).

4.3.3 After the crash

Similarly to MFCs, NHFTs appear to be the ones that contribute to resiliency (at a 1% level of significance) during open-close EPMs, helping the stock price to recover quickly. They mainly offer liquidity in the opposite direction of the crash (to the bid for a down crash; to the ask for an up crash) ($NHFT^{NET}$=-58.5018***) while HFTs keep demanding liquidity in the direction of the crash (HFT^{NET}=58.5018***).

4.4 Trade imbalance around high-low EPMs

Table 15: Trade imbalance around high-low EPMs - 1.5-second interval

	t_{-2}	t_{-1}	$t_{(crash)}$	t_{+1}	t_{+2}
HFT^{NET}	-19.8450**	-17.6575*	3.8072	56.6719***	-2.8843
HFT^{D}	-12.9600	-0.7559	100.3***	43.9688***	-11.3194
HFT^{S}	-6.8850	-16.9016**	-96.5151***	12.7031*	8.4352
$NHFT^{NET}$	19.8450**	17.6575*	-3.8072	-56.6719***	2.8843
$NHFT^{D}$	-5.2800	55.5945***	219.3***	-3.6289	76.8843
$NHFT^{S}$	25.1250	-37.9370*	-223.1***	-53.0430***	-74.0000

The table reports trade (share volume) imbalance around high-low extreme price movements (high-low EPMs) computed from Nasdaq. Time interval t is the 1.5-second interval corresponding to the crash. We also report the trade imbalance figures for 1 second and 2 seconds prior to the crash (t_{-2}, t_{-1}) and for 1 second and 2 seconds following the crash (t_{+1}, t_{+2}). HFT^{NET} ($NHFT^{NET}$) is the difference between HFT^{D} and HFT^{S} ($NHFT^{D}$ and $NHFT^{S}$). $HFT^{NET} = -NHFT^{NET}$ and vice versa. Asterisks ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels.

4.4.1 Before the crash

Just before the crash, at time interval t_{-1}, we note that HFT^{S} and HFT^{D} are negative ($HFT^{S} = -16.9016$ and $HFT^{D} = -0.7559$), which would imply HFTs trading in the opposite direction of the crash. We note for example that HFTs supply 22 times more liquidity in the opposite direction of the crash to come than they demand (-16.9016/-0.7559). We note that HFT^{S} is statistically significant at the 5% level at time interval t_{-1}, while HFT^{D} is not statistically significant which implies that one cannot reject the null hypothesis according to which the number of shares demanded by HFTs in the direction of the crash (HFT^{D}) is zero in the population.

NHFTs appear to be the ones that are (significantly) active during this phase, NHFTs being the ones that consume significantly more liquidity in the direction of the future MFC ($NHFT^{D}$=55.5945***) while supplying significantly less liquidity in the opposite direction of the future MFC ($NHFT^{S}$=-37.9370***). In net values, NHFTs appear to be the ones triggering the crash ($NHFT^{NET}$=17.6575) with $NHFT^{NET}$ being statistically significant at the 10% level, while HFTs appear to be the ones counteracting on the liquidity imbalance generated by NHFTs (HFT^{NET}=-17.6575) with HFT^{NET} being statistically significant at the 10% level.

Moreover, we note that the consumption of liquidity in the direction of the crash by NHFTs is about 74 times (=55.5945/0.7559) higher than the consumption of liquidity in the direction of the crash by HFTs. However, NHFTs supply about 2.2 times (=37.9370/16.9016) more liquidity than HFTs during this time interval. We note that HFTs and NHFTs both supply liquidity in the opposite direction of the crash (-16.9016** for HFTs and -37.9370* for NHFTs).

4.4.2 During the crash

During the crash, at time t, we note that HFT^{NET} ($NHFT^{NET}$) is positive (negative), which would imply HFTs (as a group) exacerbating the liquidity imbalance during the crash phase and NHFTs (as a group) counteracting on this liquidity imbalance. However, since HFT^{NET} and $NHFT^{NET}$ are not statistically significant at time interval t, we cannot provide any definitive statistical conclusion regarding the global activity of HFTs and NHFTs in the direction of the crash.

In more details, we note that NHFTs are the ones that are particularly active during the crash, in the same way as during open-close EPMs and MFCs. They consume about 4 times more liquidity in the direction of the crash (=219.3/55.5945) than during the prior time interval. As for HFTs, they appear to maintain their behavior in terms of liquidity supply, from supplying liquidity in the opposite direction of the crash at time interval t_{-1} (HFT^S=-16.9016**) to supplying even more liquidity in the opposite direction of the crash at time interval t (HFT^S=-96.5151***). Moreover, they appear to change their behavior in terms of liquidity demand, from demanding liquidity in the opposite direction of the crash at time interval t_{-1} (HFT^D=-0.7559) to demanding more liquidity than they supply in the direction of the crash at time interval t (HFT^D=100.3***).

4.4.3 After the crash

NHFTs appear once again to be the ones that contribute to resiliency (at a 1% level of significance), helping the stock price to recover quickly. They mainly offer liquidity in the opposite direction of the crash (to the bid for a down crash; to the ask for an up crash) ($NHFT^{NET}$=-56.6719***) while HFTs keep demanding liquidity in the direction of the crash (HFT^{NET}=56.6719***).

The main takeway of Section 4.2 is probably that the price recovery that follows the crash, whether for mini flash crashes or extreme price movements, flows from the virtuous behavior of non-high-frequency traders (while high-frequency traders viciously demand liquidity in the direction of the crash during the recovery phase). As such, non-high-frequency traders can be thanked for bringing about resiliency.

4.5 HFT activity during mini flash crashes

We now focus on the behavior of HFTs during the crash per se. To do so, we use an autoregressive model.

4.5.1 Mini flash crashes

We first test the behavior of HFTs during mini flash crashes in several multivariate regressions, taking into account the type of MFC (standalone vs simultaneous) as well as the time of occurrence of the MFC (extreme hours vs rest of the day). Our central multivariate regression is the following:

$$HFT^{NET} = \alpha + \beta_1 1MFC_{it} + \beta_2 AbsRet_{it} + \beta_3 SV_{it} + \beta_4 RS_{it} + Lags_{kit-\sigma}\gamma_{k\sigma} + \epsilon_{it} \qquad (4.3)$$

where HFT^{NET} is the difference between HFT^{D} and HFT^{S}; $1MFC_{it}$ is a dummy variable equal to one if the 1.5 second interval t in stock i is identified as an MFC and is equal to zero otherwise, $AbsRet_{it}$ is the absolute return, SV is the share volume and RS is the relative spread. $Lags_{kit-\sigma}\gamma_{k\sigma}$ is a vector of σ lags for the dependent and all of the independent variables of the regression, with $\sigma \epsilon$ {1,2,...,10} and the variables indexed with a subscript k. All the non-dummy variables are standardized at the stock level and we include stock fixed effects.

Instead of considering all MFCs, we consider MFCs for which the proportion of transactions occuring on Nasdaq during the crash represents at least 50% of all transactions on U.S. stock exchanges. Indeed, transactions may occur on different exchanges during mini flash crashes and they do sometimes occur outside of the Nasdaq exchange. By doing so, we filter out MFCs that are not prevalent on Nasdaq, which enables us to focus on the activity of HFTs on Nasdaq (for which we possess information) during crashes that partially or totally occur on Nasdaq. As a robustness check however, we also run the regressions on (1) the full MFC sample, i.e. MFCs for which the proportion of transactions on Nasdaq is comprised between 0% and 100%, thus taking into account all MFCs, including MFCs where no transaction is observed on Nasdaq, and on (2) an MFC subsample in which the proportion of transactions on Nasdaq is equal to 100%, thus taking into account MFCs where all transactions during the crash occur on Nasdaq exclusively.[11]

Table 16 reports the coefficients of the multivariate regressions. First, we focus on all MFCs (regression 1) and do not discriminate MFCs depending on their type or the time of occurence

[11]Results are available upon request.

of the crash. First, the positive estimated coefficient on the AbsRet variable indicates that HFTs tend to demand liquidity in the direction of the return. Second, the negative estimated coefficient of the 1_{MFC} dummy variable suggests that HFTs reduce their liquidity demand during mini flash crashes and trade in the opposite direction of the crash, which can be considered a virtuous behavior. In more details, HFTs reduce their liquidity demand by 4.41 standard deviations on average during the MFCs of our sample. Moreover, the coefficients of the control variables indicate that HFTs demand more liquidity when share volume is high (which is in line with the literature) and provide more liquidity when spreads widen (which is also in line with the literature). HFT^{NET} being equal to $-NHFT^{NET}$ (opposite picture), the positive estimated coefficient on $-NHFT^{NET}$ also suggests that NHFTs increase their liquidity demand by 4.41 standard deviations on average during the MFCs of our sample so that they trade in the direction of the crash, which can be considered a vicious behaviour. That being said, one cannot rule out the reduction of liquidity demand by HFTs being equal to zero in the population due to the lack of statistical significance of the 1_{MFC} dummy variable's coefficient.

We then focus on two types of MFCs: standalone MFCs on the one hand, which occur on their own, and simultaneous MFCs on the other hand, which occur on several stocks within the same minute (regression 2). The negative estimated coefficients of the $1MFC{-}STANDALONE$ and $1MFC{-}SIMULTANEOUS$ dummy variables suggest that the reduction in liquidity demand from HFTs is more pronouced during standalone mini flash crashes than during simultaneous mini flash crashes, which in turn suggests that the liquidity demand reduction from HFTs may be constrained by the nature of the shock faced by HFTs, HFTs reducing their liquidity demand more during isolated shocks than during simultaneous (and potentially systemic) shocks. Again, one cannot rule out the reduction of liquidity demand by HFTs being equal to zero in the population due to the lack of statistical significance of the $1MFC{-}STANDALONE$ and $1MFC{-}SIMULTANEOUS$ dummy variables' coefficients.

Finally, we focus on MFCs depending on the time of occurence of the MFC (regression 3). We use a dummy variable $1_{MFC-EXTREME-HOURS}$ to capture MFCs that occur from 9:30 to 9:35 and from 15:55 to 16:00 exclusively. The estimated coefficients of the $1_{MFC-EXTREME-HOURS}$ dummy variable suggest that HFTs significantly reduce their liquidity demand by 5.79 standard deviations on average during the MFCs that occur in the first five and last minutes of the trading day, which implies that the decline in liquidity demand is potentially more pronounced during periods of known market stress (opening and closing hours). We note that the coefficient is statistically significant at the 5% level.

Table 16: Net HFT activity during MFCs

Proportion of transactions on Nasdaq $\geq$ 50%			
	(1)	(2)	(3)
1_{MFC}	-441.44		
$1_{MFC-STANDALONE}$		-161.2607	
$1_{MFC-SIMULTANEOUS}$		-60.1194	
$1_{MFC-EXTREME-HOURS}$			-579.05**
$AbsRet$	6.3262***	6.3849**	6.3216***
SV	0.4617***	0.4928***	0.4807***
RS	-4.3111***	-4.4710***	-4.3032***
$Adj.R^2$	0.0104	0.0104	0.0104
N	29,285,112	29,283,943	29,285,112

The table reports the estimated coefficients of the following regression: $HFT^{NET} = \alpha + \beta_1 1MFC_{it} + \beta_2 AbsRet_{it} + \beta_3 SV_{it} + \beta_4 RS_{it} + Lags_{kit-\sigma}\gamma_{k\sigma} + \epsilon_{it}$, where HFT^{NET} is the difference between HFT^D and HFT^S; $1MFC_{it}$ is a dummy variable equal to one if the 1.5 second interval t in stock i is identified as an MFC and is equal to zero otherwise; $1MFC - STANDALONE$ is a dummy that captures MFCs that occur on their own, $1MFC - SIMULTANEOUS$ is a dummy that captures MFCs that occur on several sample stocks within the same minute, $1MFC - EXTREME - HOURS$ is a dummy that captures MFCs that occur in the first five and last five minutes of the trading day, $AbsRet$ is the absolute return, SV is the share volume and RS is the relative spread. The regressions are run on an MFC subsample whose proportion of transactions on Nasdaq is set to 0.5, meaning that at least 50% of transactions during the crash occur on Nasdaq. All non-dummy variables are standardized at the stock level and regressions are run with stock fixed effects. Asterisks ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels.

4.5.2 Open-close extreme price movements

In a similar process, we then test the behavior of HFTs during open-close extreme price movements in several multivariate regressions, taking into account the type of EPM (standalone vs simultaneous) and the time of occurrence of the EPM (opening-closing hours vs rest of the day). Our central multivariate regression is the following:

$$HFT^{NET} = \alpha + \beta_1 1EPM^{OC}_{it} + \beta_2 AbsRet_{it} + \beta_3 SV_{it} + \beta_4 RS_{it} + Lags_{kit-\sigma}\gamma_{k\sigma} + \epsilon_{it} \tag{4.4}$$

where HFT^{NET} is the difference between HFT^D and HFT^S; $1EPM^{OC}_{it}$ is a dummy variable equal to one if the 1.5 second interval t in stock i is identified as an EPM^{OC} and is equal to zero otherwise; $AbsRet_{it}$ is the absolute return, SV is the share volume and RS is the relative spread. $Lags_{kit-\sigma}\gamma_{k\sigma}$ is a vector of σ lags for the dependent and all of the independent variables of the regression, with $\sigma \in \{1,2,...,10\}$ and the variables indexed with a subscript k. All non-dummy variables are standardized at the stock level and we include stock fixed effects. Table 17 reports the coefficients of the multivariate regressions.

First, the positive estimated coefficient on the AbsRet variable (in all three regressions) indicates that HFTs tend to demand liquidity in the direction of the return. Second, the negative estimated coefficient of the $1EPM^{OC}$, $1EPM^{OC}-STANDALONE$ and $1EPM^{OC}-SIMULTANEOUS$ dummy variables suggest that HFTs reduce their liquidity demand during open-close EPMs and trade in the opposite direction of the crash during the crash, which can be considered a virtuous behavior. In more details, HFTs reduce their liquidity demand, on average, by respectively 0.45 standard deviation during typical open-close EPMs, by 0.46 standard deviation during standalone EPMs and by 0.44 standard deviation during simultaneous EPMs. In addition, the coefficients of the control variables indicate that HFTs demand more liquidity when share volume is high (which is in line with the literature) and provide more liquidity when spreads widen (which is also in line with the literature). HFT^{NET} being equal to $-NHFT^{NET}$ (opposite picture), the estimated coefficients also suggest that NHFTs increase their liquidity demand by 0.45, 0.46 and 0.44 standard deviations respectively so that they trade in the direction of the crash during the open-close EPM, which can be considered a vicious behaviour. We note that all the EPM^{OC} dummy variables are significant at the 1% level.

Finally, the coefficient of the $1_{MFC-EXTREME-HOURS}$ dummy variable suggests that HFTs reduce their liquidity demand by 1.04 standard deviation on average during the open-close EPMs that occur in the first five and last minutes of the trading day, which implies that the decline in liquidity demand is more pronounced during periods of known market stress (opening and closing hours), similar to MFCs. We note that the coefficient of $1_{EPM^{OC}-EXTREME-HOURS}$ is also highly statistically significant.

Table 17: Net HFT activity during open-close EPMs

Proportion of transactions on Nasdaq $\geq 50\%$			
	(1)	(2)	(3)
$1EPM^{OC}$	-45.24***		
$1EPM^{OC}-STANDALONE$		-45.7326***	
$1EPM^{OC}-SIMULTANEOUS$		-44.0923***	
$1EPM^{OC}-EXTREME-HOURS$			-103.7688***
$1EPM^{OC}-Q1$			
$1EPM^{OC}-Q2$			
$1EPM^{OC}-Q3$			
$1EPM^{OC}-Q4$			
$AbsRet$	6.53***	6.9847***	6.42590***
SV	0.4275**	0.4444***	0.4561***
RS	-4.2845***	-4.2516***	-4.2696***
$Adj.R^2$	0.0103	0.0104	0.0104
N	29,285,112	29,285,112	29,285,112

The table reports the estimated coefficients of the following regression: $HFT^{NET} = \alpha + \beta_1 1EPM_{it}^{OC} + \beta_2 AbsRet_{it} + \beta_3 SV_{it} + \beta_4 RS_{it} + Lags_{kit-\sigma}\gamma_{k\sigma} + \epsilon_{it}$, where HFT^{NET} is the difference between HFT^D and HFT^S; $1EPM_{it}^{OC}$ is a dummy variable equal to one if the 1.5 second interval t in stock i is identified as an open-close EPM and is equal to zero otherwise; $1EPM^{OC} - STANDALONE$ is a dummy that captures EPMs that occur on their own, $1EPM^{HL} - SIMULTANEOUS$ is a dummy that captures EPMs that occur on several sample stocks within the same interval, $1MFC - EXTREME - HOURS$ is a dummy that captures MFCs that occur in the first five and last five minutes of the trading day, $EPM^{HL} - Q1$, $EPM^{HL} - Q2$, $1EPM^{HL} - Q3$ and $1EPM^{HL} - Q4$ are dummies that capture EPMs by quartile, from the smallest to the largest, $AbsRet$ is the absolute return, SV is the share volume and RS is the relative spread. The regressions are run on an EPM subsample whose proportion of transactions on Nasdaq is set to 0.5, meaning that at least 50% of transactions during the crash occur on Nasdaq. All non-dummy variables are standardized at the stock level and regressions are run with stock fixed effects. Asterisks ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels.

4.5.3 High-low extreme price movements

We finally test the behavior of HFTs during high-low extreme price movements computed from high to low or from low to high in the same several multivariate regressions, taking into account the type of EPM (standalone vs simultaneous) and the time of occurrence of the EPM (opening-closing hours vs rest of the day). Our central multivariate regression is the following:

$$HFT^{NET} = \alpha + \beta_1 1EPM_{it}^{HL} + \beta_2 AbsRet_{it} + \beta_3 SV_{it} + \beta_4 RS_{it} + Lags_{kit-\sigma}\gamma_{k\sigma} + \epsilon_{it} \qquad (4.5)$$

where HFT^{NET} is the difference between HFT^{D} and HFT^{S}; $1EPM_{it}^{HL}$ is a dummy variable equal to one if the 1.5 second interval t in stock i is identified as a high-low EPM and is equal to zero otherwise; $AbsRet_{it}$ is the absolute return, SV is the share volume and RS is the relative spread. $Lags_{kit-\sigma}\gamma_{k\sigma}$ is a vector of σ lags for the dependent and all of the independent variables of the regression, with $\sigma \in \{1,2,...,10\}$ and the variables indexed with a subscript k. All non-dummy variables are standardized at the stock level and we include stock fixed effects. Table 18 reports the coefficients of the multivariate regressions.

First, the positive estimated coefficient on the AbsRet variable (in all three regressions) indicates one more time that HFTs tend to demand liquidity in the direction of the return. Second, the negative estimated coefficients of the $1EPM^{HL}$, $1EPM^{HL}{-}STANDALONE$ and the $1EPM^{HL}{-}SIMULTANEOUS$ dummy variables suggest that HFTs reduce their liquidity demand during high-low EPMs and trade in the opposite direction of the crash during the crash, characterizing again a somewhat virtuous behavior. In more details, HFTs reduce their liquidity demand, on average, by respectively 0.45 standard deviation during typical high-low EPMs, by 0.56 standard deviation during standalone EPMs and by 0.43 standard deviation during simultaneous EPMs. In addition, the coefficients of the control variables indicate that HFTs demand more liquidity when share volume is high (which is in line with the literature) and provide more liquidity when spreads widen (which is also in line with the literature). HFT^{NET} being equal to $-NHFT^{NET}$ (opposite picture), the estimated coefficients also suggest that NHFTs increase their liquidity demand by 0.45, 0.56 and 0.43 standard deviations respectively so that they trade in the direction of the crash during the high-low EPM, characterizing a somewhat vicious behaviour. We note that all the EPM^{HL} dummy variables are significant at the 1% level.

Finally, the coefficient of the $1_{MFC-EXTREME-HOURS}$ dummy variable suggests that HFTs reduce their liquidity demand by 0.73 standard deviation on average during the high-low EPMs that occur in the first five and last minutes of the trading day, which implies that the decline in liquidity demand is more pronounced during periods of known market stress (opening and closing hours), similar to MFCs and open-close EPMs. We note that the coefficient of $1_{EPM^{HL}-EXTREME-HOURS}$ is again highly statistically significant.

Table 18: Net HFT activity during high-low EPMs

Proportion of transactions on Nasdaq $\geq$ 50%			
	(1)	(2)	(3)
$1EPM^{HL}$	-45.11***		
$1EPM^{HL}-STANDALONE$		-55.5082***	
$1EPM^{HL}-SIMULTANEOUS$		-43.2787***	
$1EPM^{HL}-EXTREME-HOURS$			-73.4342***
$1EPM^{HL}-Q1$			
$1EPM^{HL}-Q2$			
$1EPM^{HL}-Q3$			
$1EPM^{HL}-Q4$			
$AbsRet$	4.3337***	4.7840***	4.2468***
SV	0.3499**	0.3448**	0.3839**
RS	-4.1882***	-4.1581***	-4.1832***
$Adj.R^2$	0.0096	0.0097	0.0097
N	29,292,021	29,292,021	29,292,021

The table reports the estimated coefficients of the following regression: $HFT^{NET} = \alpha + \beta_1 1EPM_{it}^{HL} + \beta_2 AbsRet_{it} + \beta_3 SV_{it} + \beta_4 RS_{it} + Lags_{kit-\sigma}\gamma_{k\sigma} + \epsilon_{it}$, where HFT^{NET} is the difference between HFT^D and HFT^S; $1EPM_{it}^{HL}$ is a dummy variable equal to one if the 1.5 second interval t in stock i is identified as a high-low EPM and is equal to zero otherwise; $1EPM^{HL} - STANDALONE$ is a dummy that captures EPMs that occur on their own, $1EPM^{HL} - SIMULTANEOUS$ is a dummy that captures EPMs that occur on several sample stocks within the same interval, $1MFC - EXTREME - HOURS$ is a dummy that captures MFCs that occur in the first five and last five minutes of the trading day, $EPM^{HL} - Q1$, $EPM^{HL} - Q2$, $1EPM^{HL} - Q3$ and $1EPM^{HL} - Q4$ are dummies that capture EPMs by quartile, from the smallest to the largest, $AbsRet$ is the absolute return, SV is the share volume and RS is the relative spread. The regressions are run on an EPM subsample whose proportion of transactions on Nasdaq is set to 0.5, meaning that at least 50% of transactions during the crash occur on Nasdaq. All non-dummy variables are standardized at the stock level and regressions are run with stock fixed effects. Asterisks ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels.

The main takeaway of Section 4.5 is that HFTs tend to have a virtuous behavior during the crash, whether it is a mini flash crash or an extreme price movement. First, HFTs significantly reduce their liquidity demand during MFCs and EPMs (whether computed from open to close or from high to low) occuring in the first five and last five minutes of the trading day. Second, the decline in HFTs' liquidity demand during the crash seems more pronounced during periods of known market stress as characterized by extreme hours (opening and closing periods) or when the crash is non systemic (standalone crash). On the contrary, when the crash is not anticipated or potentially systemic (simultaneous crashes occuring at the same time), the reduction in liquidity demand by HFTs is less pronounced.

4.6 HFT trading activity and probability of future mini flash crashes

We eventually model the probability of having a mini flash crash as a function of lagged values of HFT^{NET}, absolute return, share volume and relative spread as in Brogaard et al. (2018) through the use of a probit model (dependent variable = MFC). We then use the same probit model to estimate the probability of having an open-close extreme price movement (dependent variable = open-close EPM) and a high-low extreme price movement (dependent variable = high-low EPM) as a function of lagged values of HFT^{NET}, absolute return, share volume and relative spread so as to put the different results into perspective.

4.7 Mini flash crashes

We first model the probability of having a mini flash crash as a function of lagged values of HFT^{NET}, absolute return, share volume and relative spread:

$$Prob(MFC = 1)_{it} = \alpha + \beta_1 HFT^{NET}{}_{it-1} + \beta_2 AbsRet_{it-1} + \beta_3 SV_{it-1} + \beta_4 RS_{it-1} + \epsilon_{it} \quad (4.6)$$

where the dependent variable is equal to one if the 1.5-second interval t contains a mini flash crash and zero otherwise. All the independent variables are lagged by one interval and all the variables are standardized at the stock level. HFT^{NET} is the difference between HFT^{D} and HFT^{S} for the 26 HFTs in our dataset, $AbsRet$ is the absolute return, SV is the share volume and RS is the relative spread.

Results are presented in Table 19. HFT^{NET} is statistically significant in the "Standalone" and "Extreme hours" probit regressions (at 5% and 10% respectively). The negative sign of the HFT^{NET} variable in the "Standalone" column (i.e. for MFCs that occur on their own) indicates that an increase in HFT^{NET} at time $t-1$ makes the probability of an MFC less likely. In the same way, the negative sign of the HFT^{NET} variable in the "Extreme hours" column (for MFCs that occur in the first five and last five minutes of the trading day) indicates that an increase in HFT^{NET} at time $t-1$ also makes the probability of an MFC less likely. As such, the probability of having an MFC decreases when HFT^{NET} increases in the pre-crash interval.

Table 19: Net HFT activity and probability of future mini flash crashes

	All	Standalone	Simultaneous	Extreme hours
Intercept	-4.1556***	-4.3165***	-4.3358***	-2.0283***
	(0.00)	(0.00)	(0.00)	(0.00)
HFT^{NET}_{t-1}	-0.00330	-0.00008**	0.00006	-0.00016***
Controls	Yes	Yes	Yes	Yes
$Pseudo-R^2$	0.0235	0.0127	0.0475	0.0388
N	29,302,068	29,302,081	29,302,081	29,301,958

The table reports the estimated coefficients of a probit model regarding the probability of having a mini flash crash in the future:
Prob (MFC = 1)$_{it} = \alpha + \beta_1 HFT^{NET}{}_{it-1} + \beta_2 AbsRet_{it-1} + \beta_3 SV_{it-1} + \beta_4 RS_{it-1} + \epsilon_{it}$,
where the dependent variable is equal to one if the 1.5-second interval t contains a mini flash crash and zero otherwise. All the independent variables are lagged by one interval. HFT^{NET} is the difference between HFT^D and HFT^S for the 26 HFTs in our dataset, *AbsRet* is the absolute return, *SV* is the share volume and *RS* is the relative spread. All non-dummy variables are standardized at the stock level. P-values are presented in parentheses and asterisks ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels.

4.8 Open-close extreme price movements

We repeat the procedure and model the probability of having an open-close extreme price movement as a function of lagged values of HFT^{NET}, absolute return, share volume and relative spread:

$$Prob(EPM^{OC} = 1)_{it} = \alpha + \beta_1 HFT^{NET}{}_{it-1} + \beta_2 AbsRet_{it-1} + \beta_3 SV_{it-1} + \beta_4 RS_{it-1} + \epsilon_{it} \quad (4.7)$$

where the dependent variable is equal to one if the 1.5-second interval t contains an open-close extreme price movement and zero otherwise. All the independent variables are lagged by one interval and all the variables are standardized at the stock level. HFT^{NET} is the difference between HFT^D and HFT^S for the 26 HFTs in our dataset, *AbsRet* is the absolute return, *SV* is the share volume and *RS* is the relative spread.

Results for the "All" probit regression are presented in Table 20. HFT^{NET} is statistically significant at the 5% level. Contrary to MFCs, we find that the sign of the HFT^{NET} variable is positive, which indicates that an increase in HFT^{NET} at time $t-1$ makes the probability of an open-close EPM more likely.

Table 20: Net HFT activity and probability of future open-close EPMs

	All	Standalone	Simultaneous	Extreme hours
Intercept	-3.14597***			
	(0.00)			
HFT^{NET}_{t-1}	0.00507**			
Controls	Yes			
$Pseudo-R^2$	0.0677			
N	29,302,068			

The table reports the estimated coefficients of a probit model regarding the probability of having an open-close extreme price movement in the future:
Prob (EPM$^{OC}=1)_{it} = \alpha + \beta_1 HFT^{NET}{}_{it-1} + \beta_2 AbsRet_{it-1} + \beta_3 SV_{it-1} + \beta_4 RS_{it-1} + \epsilon_{it}$,
where the dependent variable is equal to one if the 1.5-second interval t contains an open-close extreme price movement and zero otherwise. All the independent variables are lagged by one interval. HFT^{NET} is the difference between HFT^D and HFT^S for the 26 HFTs in our dataset, *AbsRet* is the absolute return, *SV* is the share volume and *RS* is the relative spread. All non-dummy variables are standardized at the stock level. P-values are presented in parentheses and asterisks ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels.

4.9 High-low extreme price movements

We finally model the probability of having a high-low extreme price movement as a function of lagged values of HFT^{NET}, absolute return, share volume and relative spread:

$$Prob(EPM^{HL}=1)_{it} = \alpha + \beta_1 HFT^{NET}{}_{it-1} + \beta_2 AbsRet_{it-1} + \beta_3 SV_{it-1} + \beta_4 RS_{it-1} + \epsilon_{it} \quad (4.8)$$

where the dependent variable is equal to one if the 1.5-second interval t contains a high-low extreme price movement and zero otherwise. All the independent variables are lagged by one interval and all the variables are standardized at the stock level. HFT^{NET} is the difference between HFT^D and HFT^S for the 26 HFTs in our dataset, *AbsRet* is the absolute return, *SV* is the share volume and *RS* is the relative spread.

Results for the "All" probit regression are presented in Table 21. HFT^{NET} is statistically significant at the 5% level. Contrary to open-close EPMs, we find that the sign of the HFT^{NET} variable is negative for high-low EPMs, which indicates that an increase in HFT^{NET} at time $t-1$ makes the probability of a high-low EPM less likely.

Table 21: Net HFT activity and probability of future high-low EPMs

	All	Standalone	Simultaneous	Extreme hours
Intercept	-3.21526***			
	(0.00)			
HFT^{NET}_{t-1}	-0.00245**			
Controls	Yes			
$Pseudo-R^2$	0.1849			
N	29,302,803			

The table reports the estimated coefficients of a probit model regarding the probability of having a high-low extreme price movement in the future:
Prob $(\text{EPM}^{HL}=1)_{it} = \alpha + \beta_1 HFT^{NET}{}_{it-1} + \beta_2 AbsRet_{it-1} + \beta_3 SV_{it-1} + \beta_4 RS_{it-1} + \epsilon_{it}$,
where the dependent variable is equal to one if the 1.5-second interval t contains a high-low extreme price movement and zero otherwise. All the independent variables are lagged by one interval. HFT^{NET} is the difference between HFT^D and HFT^S for the 26 HFTs in our dataset, *AbsRet* is the absolute return, *SV* is the share volume and *RS* is the relative spread. All non-dummy variables are standardized at the stock level. P-values are presented in parentheses and asterisks ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels.

The results of our probit regressions are rather ambiguous at this stage and quite difficult to interpret.

5 Robustness checks

We perform robustness checks at two levels.

First, we run parallel analyses on mini flash crashes and extreme price movements (both open-close EPMs and high-low EPMs) throughout the paper so as to (1) highlight the similarities and differences of each type of crash and (2) cross check our results. We note that even though mini flash crashes are somewhat different from extreme price movements, our findings related to mini flash crashes are most of the time corroborated by similar findings related to extreme price movements and vice versa.

Second, we compute the three identification methods (MFCs, open-close EPMs, high-low EPMs) using alternative time intervals. Indeed, while our base time interval is 1.5 second and more precisely a flexible 1.5-second interval for mini flash crashes and a fixed 1.5-second interval for extreme price movements, we also perform a similar analysis using alternative time intervals of 1 second and 2 seconds respectively. Whichever time interval we use, our results remain very similar.[12]

[12] Results for 1-second and 2-second intervals are available upon request.

6 Conclusion

We identify MFCs by replicating Nanex MFC detection algorithm (2010) and we complement our study with a parallel analysis of extreme price movements (EPMs), providing an alternative way to measure extreme price movements. In more detail, we identify 510 MFCs, 29,424 open-close EPMs and 29,427 high-low EPMs over a two-year period, representing about one mini flash crash and fifty-eight extreme price movements per day on average. While we observe that some EPMs occur without HFTs being active (a fact that could clear HFTs from any accusation left), we also observe that HFTs are always active on aggregate during MFCs.

To the question 'Do HFTs trigger mini flash crashes ?' and based on the trade imbalance measure computed right before the crash, we find that HFTs demand twice more liquidity in the direction of the crash to come than they supply. Moreover, we find that NHFTs supply liquidity in the opposite direction of the crash while HFTs supply liquidity in the direction of the crash to come. However, we cannot reject the null hypothesis according to which the net liquidity demanded by HFTs in the direction of the crash to come is actually zero in the population due to the non-significance of the coefficient.

To the question 'Do HFTs exacerbate the crash ?' and based on the autoregressive model, we find that HFTs tend to reduce their liquidity demand during the crash and trade in the opposite direction of the crash. However, one cannot rule out the reduction of liquidity demand by HFTs being equal to zero in the population due to the non-significance of the MFC coefficient. When focusing on extreme price movements (both open close EPMs and high-low EPMs) rather than on mini flash crashes, we find that HFTs do reduce their liquidity demand during the crash at a significance level of 1% thus confirming the findings by Brogaard et al (2018).

Finally, to the question 'Do HFTs lead the price recovery right after the crash?' and based on the trade imbalance measure computed right after the crash, the answer is no. On the contrary, NHFTs appear to be the ones that contribute to the resiliency of stock prices after the crash (even though at a 10% significance level only) thus driving the price recovery. On the contrary, HFTs keep demanding liquidity in the direction of the crash during the post-crash phase. This, we think, offers new insight about liquidity consumption by HFTs during the recovery phase of mini flash crashes and should be studied further.

References

Aldrich, E., Grundfest, J. & Laughlin, G. (2017), 'The flash crash: a new deconstruction', *Working paper* .

Aquilina, M., Eyles, B., Shao, J. & Ysusi, C. (2018), 'How do participants behave during flash events? evidence from the uk equity market', *FCA Research Note* .

Bellia, M., Christensen, K., Kolokolov, A., Pelizzon, L. & Renò, R. (2018), High-frequency trading during flash crashes: Walk of fame or hall of shame?, *in* 'Paris December 2018 Finance Meeting EUROFIDAI-AFFI'.

Boehmer, E., Fong, K. Y. & Wu, J. (2018), Algorithmic trading and market quality: International evidence, *in* 'AFA 2013 San Diego Meetings Paper'.

Brogaard, J., Carrion, A., Moyaert, T., Riordan, R., Shkilko, A. & Sokolov, K. (2018), 'High frequency trading and extreme price movements', *Journal of Financial Economics* **128**(2), 253–265.

Brogaard, J., Hendershott, T. & Riordan, R. (2014), 'High-frequency trading and price discovery', *The Review of Financial Studies* **27**(8), 2267–2306.

Brogaard, J., Hendershott, T. & Riordan, R. (2017), 'High frequency trading and the 2008 short-sale ban', *Journal of Financial Economics* **124**(1), 22–42.

Brogaard, J. et al. (2010), 'High frequency trading and its impact on market quality', *Northwestern University Kellogg School of Management Working Paper* **66**.

Carrion, A. (2013), 'Very fast money: High-frequency trading on the nasdaq', *Journal of Financial Markets* **16**(4), 680–711.

Félez-Vinas, E. (2018), Market fragmentation, mini flash crashes and liquidity, Working paper presented at the FMA European Conference, Kristiansand, Norway.

Foucault, T., Moinas, S. et al. (2018), Is trading fast dangerous?, Technical report, Toulouse School of Economics (TSE).

Gerig, A. (2015), 'High-frequency trading synchronizes prices in financial markets', *Working paper* .

Golub, A. & Keane, J. (2011), 'Mini flash crashes', *Working paper* .

Golub, A., Keane, J. & Poon, S.-H. (2012), 'High frequency trading and mini flash crashes', *Working paper* .

Hagströmer, B. & Nordén, L. (2013), 'The diversity of high-frequency traders', *Journal of Financial Markets* **16**(4), 741–770.

Hasbrouck, J. & Saar, G. (2013), 'Low-latency trading', *Journal of Financial Markets* **16**(4), 646–679.

Hirschey, N. (2018), 'Do high-frequency traders anticipate buying and selling pressure?', *Working paper* .

Johnson, N., Zhao, G., Hunsader, E., Qi, H., Johnson, N., Meng, J. & Tivnan, B. (2013), 'Abrupt rise of new machine ecology beyond human response time', *Scientific reports* **3**, 2627.

Jones, C. M. (2013), 'What do we know about high-frequency trading?', *Columbia Business School Research Paper* (13-11).

Jovanovic, B. & Menkveld, A. J. (2016), 'Middlemen in limit order markets', *Available at SSRN 1624329* .

Kirilenko, A., Kyle, A. S., Samadi, M. & Tuzun, T. (2017), 'The flash crash: High-frequency trading in an electronic market', *The Journal of Finance* **72**(3), 967–998.

Leal, S. J., Napoletano, M., Roventini, A. & Fagiolo, G. (2016), 'Rock around the clock: An agent-based model of low-and high-frequency trading', *Journal of Evolutionary Economics* **26**(1), 49–76.

Lee, S. S. & Mykland, P. A. (2012), 'Jumps in equilibrium prices and market microstructure noise', *Journal of Econometrics* **168**(2), 396–406.

Massad, T. (2015), 'Remarks of chairman timothy massad before the conference on the evolving structure of the u.s. treasury market'.

Menkveld, A. J. & Yueshen, B. (n.d.), 'The flash crash: A cautionary tale about highly fragmented markets', *Working paper* .

Nanex (2010), Flash equity failures, Technical report, Nanex, LLC. http://www.nanex.net/Research/MiniFlashCrashDef/MiniFlashCrashDef.html.

O'Hara, M., Yao, C. & Ye, M. (2014), 'What's not there: Odd lots and market data', *Journal of Finance* **69**, 2199–2236.

Scholtus, M., van Dijk, D. & Frijns, B. (2014), ‘Speed, algorithmic trading, and market quality
1 around macroeconomic news announcements’, *Journal of Banking & Finance* **38**, 89–105.

Stoll, H. R. (2014), ‘High speed equities trading: 1993–2012’, *Asia-Pacific Journal of Financial
2 Studies* **43**(6), 767–797.

7 Appendix : additional tables

Additional tables are presented in this section.

Table 22 reports descriptive statistics for the MFC subsamples following the Nanex identification method and depending on the proportion of transactions on Nasdaq.

Table 23 reports the ticker, company name, market cap, sector (GICS classification) as well as the number of MFCs regarding the 74 sample stocks.

Table 22: Summary statistics of MFC subsamples (Nanex)

Panel A: Full sample					
	Min	Med	Mean	Max	Std Dev
Absolute return, %	0	0.0076	0.0228	19.64	0.0439
Total trades (all U.S. exchanges)	1	4	8.96	1205	16.13
Total trades (Nasdaq)	1	3	5.99	880	9.32
Total HFT trades (Nasdaq)	0	0.88	0.69	1	0.38
Share volume	1	600	2018.76	28,368,232	16,232
Dollar volume	1	22,501	84,874	1,458,142,722	692,954
Depth	2	7.07	31.24	100,816	109.44
Dollar Depth	6.34	195.98	467.59	2,225,817	1907.77
Quoted spread, $	0.01	0.0171	0.0346	113.91	0.0815
Relative spread, %	0.002	0.052	0.083	56.21	0.12
N	29,304,647				
Panel B: MFC subsample with proportion of transactions on Nasdaq ≥ 0%					
	Min	Med	Mean	Max	Std Dev
Crash duration, ms (MFC specific)	0	46	164	1487	286
Total tick change (MFC specific)	10	12	13.38	35	3.96
Absolute return, %	0.80	1.12	1.83	36.84	3.26
Total trades (all U.S. exchanges)	3	68	104.34	883	110.76
Total trades (Nasdaq)	0	25	44.64	672	60.32
Total HFT trades (Nasdaq)	0	0.67	0.62	1	0.28
Share volume	500	20,972	54,466	3,138,737	160,530
Dollar volume	13,343	729,535	2,191,106	146,022,200	8,962,257
Depth	2	7.57	26.62	1,505.41	91.87
Dollar depth	20.42	177.16	519.87	36,063	2,317.95
Quoted spread, $	0.01	0.03	0.76	113.91	7.43
Relative spread, %	0.009	0.09	0.43	42.89	2.72
N	510				
Panel C: MFC subsample with proportion of transactions on Nasdaq ≥ 50%					
	Min	Med	Mean	Max	Std Dev
Crash duration, ms (MFC specific)	0	125	274	1487	344
Total tick change (MFC specific)	10	12	13.87	35	4.56
Absolute return, %	0.80	1.17	2.82	36.84	5.62
Total trades (all U.S. exchanges)	3	62	104.84	709	113.93
Total trades (Nasdaq)	1	38	65	672	77.44
Total HFT trades (Nasdaq)	0	0.53	0.54	1	0.28
Share volume	500	15,900	65,348.79	3,138,737	258,730.39
Dollar volume	14,973	661,288	3,746,384	146,022,200	15,471,513
Depth	2	5.82	40.48	1505	156.49
Dollar depth	43.65	174.05	956.07	36,063	4,113.16
Quoted spread, $	0.01	0.0456	2.46	113.91	13.57
Relative spread, %	0.01	0.14	1.12	42.89	4.94
N	149				
Panel D: MFC subsample with proportion of transactions on Nasdaq = 100%					
	Min	Med	Mean	Max	Std Dev
Crash duration, ms (MFC specific)	0	60	173	1263	254
Total tick change (MFC specific)	10	12	13.58	35	4.90
Absolute return, %	0.806	1.38	4.53	36.84	8.28
Total trades (all U.S. exchanges)	5	64	125.30	580	127.04
Total trades (Nasdaq)	4	44	72.41	264	65.23
Total HFT trades (Nasdaq)	0.0988	0.5769	0.5616	1	0.268
Share volume	600	17,429	138,340	3,138,737	483,529
Dollar volume	29,677	955,624	8,484,162	146,022,200	27,965,515
Depth	2	5.84	44.30	576.41	118.65
Dollar depth	56.86	264.43	884.73	11,576.25	2,263.35
Quoted spread, $	0.01	0.0501	5.26	113.91	19.96
Relative spread, %	0.019	0.1579	2.28	42.89	7.34
N	44				

The table reports descriptive statistics for the MFC subsamples following the Nanex identification method and depending on the proportion of transactions on Nasdaq.

Table 23: Sample stocks summary statistics

Ticker	Company Name	Market Cap	Sector (GICS)	Nb of MFCs
AA	Alcoa Corp	Large	Materials	16
AAPL	Apple Inc	Large	Information Technology	27
ABD	Acco Brands Corp	Small	Industrials	2
ADBE	Adobe Systems Inc	Large	Information Technology	6
AMED	Amedisys Inc	Medium	Health Care	1
AMGN	Amgen Inc	Large	Health Care	12
AMZN	Amazon.com Inc	Large	Consumer Discretionary	13
ARCC	Ares Capital Corp	Medium	Financials	1
AXP	American Express Co	Large	Financials	51
AYI	Acuity Brands Inc	Medium	Industrials	6
AZZ	AZZ Inc	Small	Industrials	2
BARE	Bare Escentuals Inc	Medium	Consumer Staples	2
BAS	Basic Energy Services Inc	Small	Energy	0
BHI	Baker Hughes a GE Co LLC	Large	Energy	7
BIIB	Biogen Inc	Large	Health Care	15
BRE	BEX Portfolio LLC	Medium	Real Estate	3
BXS	Bancorp South Inc	Medium	Financials	5
CBEY	Cbeyond Inc	Small	Telecommunications Services	0
CETV	Central European Media Enterprises Ltd	Medium	Consumer Discretionary	12
CNQR	Concur Technologies Inc	Medium	Information Technology	4
COO	Cooper Cos IncThe	Medium	Health Care	0
COST	Costco Wholesale Corp	Large	Consumer Staples	2
CR	Crane Co	Medium	Industrials	1
CSCO	Cisco Systems Inc	Large	Information Technology	23
CSE	Capital Source Inc	Medium	Financials	4
CSL	Carlisle Cos. Inc	Large	Industrials	2
CTSH	Cognizant Technology Solutions Corp	Large	Information Technology	2
DIS	Walt Disney CoThe	Large	Consumer Discretionary	11
DOW	Dow Chemical CoThe	Large	Materials	4
EBAY	eBay Inc	Large	Information Technology	3
ESRX	Express Scripts Holding Co	Large	Health Care	1
FCN	FTI Consulting Inc	Medium	Industrials	3
FL	Foot Locker Inc	Medium	Consumer Discretionary	3
FMER	FirstMerit Corp	Medium	Financials	0
FPO	First Potomac Realty Trust	Small	Real Estate	2
FULT	Fulton Financial Corp	Medium	Financials	0
GE	General Electric Co	Large	Industrials	66
GILD	Gilead Sciences Inc	Large	Health Care	12
GLW	Corning Inc	Large	Information Technology	10
GOOG	Alphabet Inc	Large	Information Technology	16
GPS	Gap IncThe	Large	Consumer Discretionary	1
HON	Honeywell International Inc	Large	Industrials	5
HPQ	HP Inc	Large	Information Technology	9
IMGN	ImmunoGen Inc	Small	Health Care	0
INTC	Intel Corp	Large	Information Technology	3
ISRG	Intuitive Surgical Inc	Large	Health Care	1
JKHY	Jack Henry & Associates Inc	Medium	Information Technology	1
KMB	Kimberly-Clark Corp	Large	Consumer Staples	2
KNOL	Knology Inc	Small	Telecommunications Services	0
KR	Kroger CoThe	Large	Consumer Staples	1
LANC	Lancaster Colony Corp	Medium	Consumer Staples	0
LECO	Lincoln Electric Holdings Inc	Medium	Industrials	0
LPNT	LifePoint Health Inc	Medium	Health Care	2
LSTR	Landstar System Inc	Medium	Industrials	1
MANT	ManTech International CorpThe	Medium	Information Technology	1
MDCO	Medicines CoThe	Small	Health Care	4
MELI	Mercadorlibre Inc	Medium	Information Technology	1
MFB	Maidenform Brands LLC	Small	Consumer Discretionary	0
MMM	3M Co	Large	Industrials	5
MOD	Modine Manufacturing Co	Small	Consumer Discretionary	1
MOS	Mosaic CoThe	Large	Materials	39
NSR	NewStar Inc	Medium	Information Technology	4
NUS	Nu Skin Enterprises Inc	Medium	Consumer Staples	1
PFE	Pfizer Inc	Large	Health Care	7
PG	Procter & Gamble CoThe	Large	Consumer Staples	11
PNC	PNC Financial Services Group IncThe	Large	Financials	18
PNY	Piedmont National Gas Co Inc	Medium	Utilities	2
PTP	Platinum Underwriters Holdings Ltd	Medium	Financials	3
RIGL	Rigel Pharmaceuticals Inc	Small	Health Care	0
ROC	Rockwood Holdings Inc	Medium	Materials	1
ROCK	Gibraltar Industries Inc	Medium	Industrials	0
SF	Stifel Financial Corp	Medium	Financials	5
SFG	StanCorp Financial Group Inc	Medium	Financials	1
SWN	Southwestern Energy Co	Large	Energy	9

The table reports the ticker, company name, market cap, sector (GICS classification) as well as the number of MFCs for the 74 stocks of our sample. Market capitalization data are provided by Bloomberg. The group classification is the same whether we use Bloomberg's historical market capitalization as of February 26, 2010, Bloomberg's 2008, 2009, 2010 historical market capitalizations or Bloomberg's 2008, 2009, 2010 average historical market capitalization.