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

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

December 5, 2019

Abstract

We investigate how high-frequency traders (HFTs) behave around and during a certain type of flash events known as mini flash crashes (MFCs) on 74 large, medium and small Nasdaq stocks over the period 2008-2010. To do so, we identify MFCs by replicating Nanex MFC detection algorithm and complement our study with a parallel analysis on extreme price movements (EPMs), providing an alternative way to measure them. Our findings reveal that the behavior of HFTs around and during 1.5-second crashes is ambiguous. Based on directional trade imbalance metrics, we find that HFTs do highly significantly exacerbate the crash during 1.5-second extreme price movements, which contradicts the findings by Brogaard et al. (2018), and that NHFTs are the ones who contribute to the resiliency of stock prices after EPMs (at a 1% significance level). However, based on a multivariate regression analysis, we find that, on average, HFTs reduce their liquidity demand during EPMs. When studying MFC and EPM subsamples by market capitalization, we find that HFTs do reduce their liquidity demand during EPMs occurring on large stocks but increase their liquidity demand during EPMs occurring on small stocks. The virtuous behavior of HFTs in large stocks may thus hide a more vicious behavior in small stocks. Moreover, we find that the reduction in HFT liquidity demand is more pronounced during the opening and closing periods. Finally, we find that HFT participation at time *t-1* is by far the main determinant of mini flash crashes at time *t*, whatever the model specification.

JEL classification: G1, G10, G14

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

---

***Preliminary and incomplete.**

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

‡Louvain School of Management, UCLouvain, LFIN-LIDAM, 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, UCLouvain, LFIN-LIDAM, 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 abrupt price changes (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 (MFCs) (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 agent-based 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, 2018), 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 by extension) 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 (2018), 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. These 65 flash crashes have a mean return of 1.58% and a mean duration of 12.25 minutes.

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 Aquilina et al, 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. For a complementary review of the literature on mini flash crashes, see Laly & Petitjean (2019).

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 (EPMs) (Brogaard et al, 2018) even though they are in fact not totally identical in nature. 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 absolute 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 abrupt price changes[1] on a sample of large, medium and small cap Nasdaq equities over a two-year period (2008-2010).

We define mini flash crashes as sudden, extreme and very short-lived abrupt price changes that exhibit at least ten tick movements in the same direction before ticking in the other direction (following Nanex, 2010) and 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 the 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 this time interval before bouncing back to its initial level.

[1]Our analysis focuses on MFCs and EPMs 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.

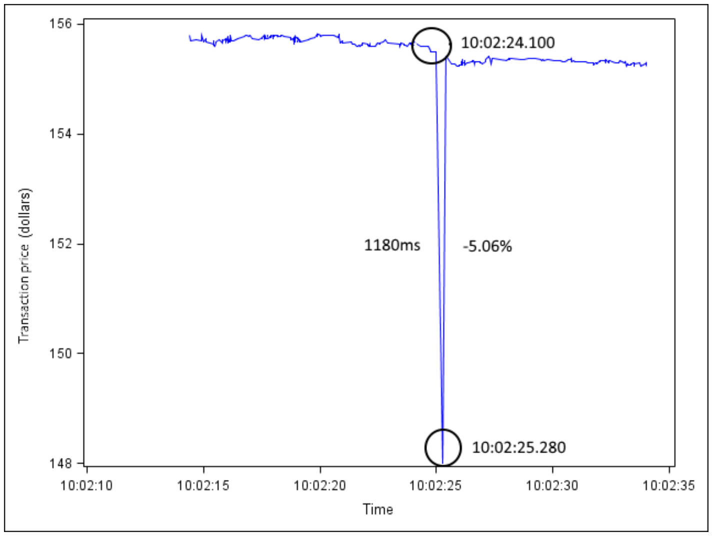
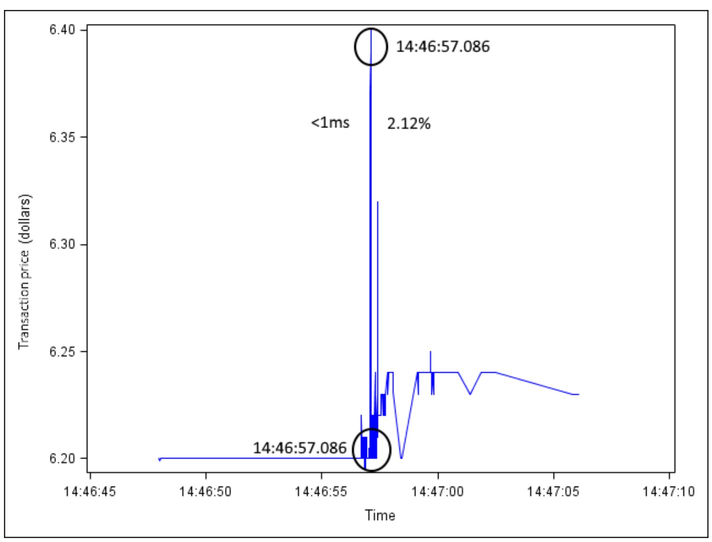


Figure 2 presents an example of an up mini flash crash that impacted the Alcoa (AA) stock on March 16, 2009. The crash from bottom to top occurred in less than 1 millisecond (starting at 10:46:57.086 and ending at 10:46:57.086), the (transaction) price jumping 2.12% within this sub-millisecond interval 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 absolute 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 exact same database in this paper. 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 stocks of the Nasdaq HFT database over the period 2008-2010.

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 within 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.

Figure 3: **Illustration of an EPM on AA - January 22, 2008**

The data are from Tickdata.

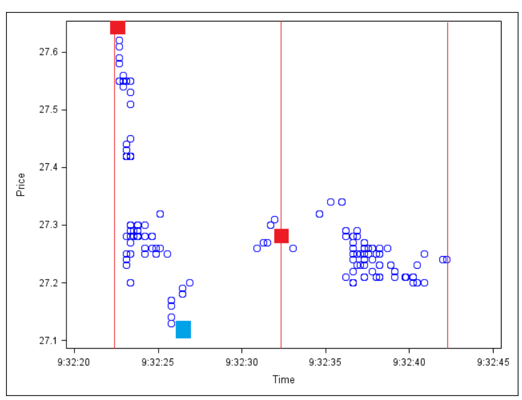
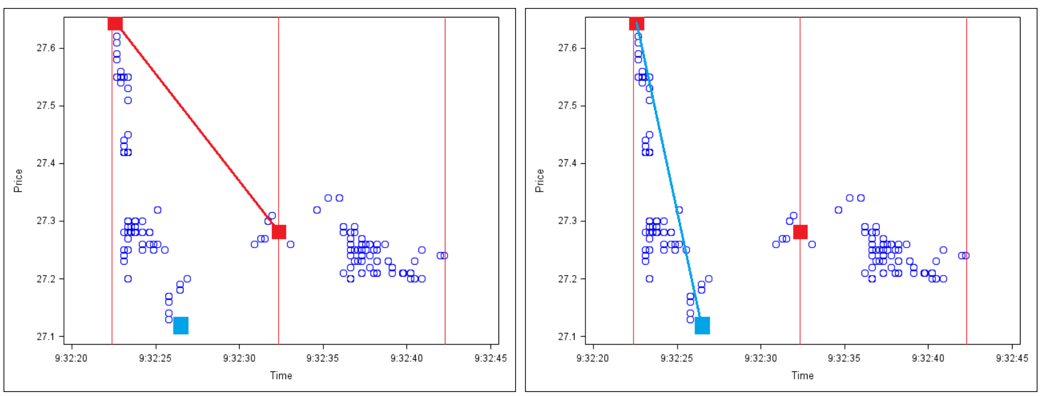


Figure 3 offers a visual illustration of why measuring EPMs from open to close may potentially be misleading. The graph shows an EPM occurring on the Alcoa (AA) stock on January 22, 2008. The red vertical lines represent the 95th, 96th and 97th 10-second intervals respectively for that day. Each dot stands for transactions within the same millisecond.[2] The red square dots represent the opening and closing transaction prices within the 10-second interval (from interval 95 to 96), while the light blue dot represents the true extreme transaction price of the sequence. The measured crash, represented by the red diagonal line (Figure 4), underestimates the true crash, represented by the light blue diagonal line (Figure 4), so that the open-close EPM approach (i.e. measuring the magnitude of the crash based on the close price and opening price of the interval) may not be ideal.

Figure 4: **Measured crash vs true crash on AA - January 22, 2008**
The data are from Tickdata.



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 (i.e. we identify the exact timing of the crash), 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).

Our findings reveal that the behavior of HFTs around and during 1.5-second crashes is ambiguous. We find that (1) HFTs do highly statistically significantly exacerbate the crash during 1.5-second extreme price movements, which contradicts the findings by Brogaard et al. (2018), (2) NHFTs are the ones who contribute to the resiliency of stock prices after the crash (at a 1% significance level) thus driving the price recovery right after EPMs, which corroborates the findings by Bellia et al. (2018), (3) HFTs reduce their liquidity demand during MFCs on an ag-

[2]Note that on Figure 3, each dot is supposed to represent a transaction. However, when two or more transactions occur within the same millisecond, all the transactions appear on the same dot graphically.

gregate basis, which is definitely a virtuous behaviour. However, when studying MFC and EPM subsamples by market capitalization, we find that HFTs do reduce their liquidity demand during EPMs occurring on large stocks but increase their liquidity demand during EPMs occurring on small stocks. The virtuous behavior of HFTs in large stocks may hide a more vicious behavior in less liquid stocks. Finally, we find that HFT participation at time *t-1* is by far the main determinant of mini flash crashes at time t, whatever the model specification.

The remainder of the paper is organized as follows. Section 2 introduces the data, presents the different MFC identification methods we use in the paper as well as the summary statistics, and defines the different variables. Section 3 is dedicated to the empirical part of the paper and presents our results. Section 4 summarizes the different robustness checks performed in the paper. Section 5 concludes.

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 the aggregated data from Tickdata to identify MFCs (as well as EPMs) since MFCs originate from successive transactions occurring on different exchanges (and not on one exchange only, even though that can be the case from time to time). Second, we use the Nasdaq HFT dataset to observe the behavior of HFTs during the identified MFCs of our sample.

In more details, 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 the list in Appendix). The data from Tickdata account for all transactions on U.S. stock exchanges[3] for the 74 stocks at our disposal. Then, 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 exclusively. A flag on Nasdaq trade data enables us to know if the liquidity demander/supplier is a high-frequency trader (HFT) or a non-high frequency trader (NHFT). Indeed, a code (HH, HN, NH, NN) is associated with each transaction on Nasdaq. 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-Exchanges, BATS.

first letter of the code refers to the liquidity demander and the second letter of the code refers to the liquidity supplier. For example, when a transaction occurs between a HFT and an NHFT, the code can either be HN or NH depending on who is on the liquidity demand side and who is on the liquidity supply side. HN indicates that the HFT demands liquidity while the NHFT supplies liquidity. Conversely, NH indicates that the NHFT demands liquidity while the HFT supplies liquidity.

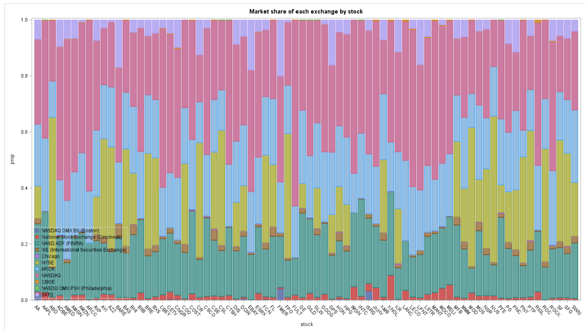
We use a window of 30 minutes (i.e. 15 minutes before and 15 minutes after the crash) around each MFC in our sample, discarding all the other days, so that the final sample period is shorter than the initial reference (Nasdaq) sample period which runs from January 1, 2008 to December 31, 2009. The Tickdata 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 stock 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.

We use transaction prices (from the aggregated data from Tickdata) instead of midquotes in the identification of mini flash crashes in order to take into account the full magnitude of each crash (from top to bottom or from bottom to top) and we use midquotes (from the aggregated data from Tickdata) in the identification of both open-close and high-low extreme price movements as in Brogaard et al (2018). We 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.

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 5.

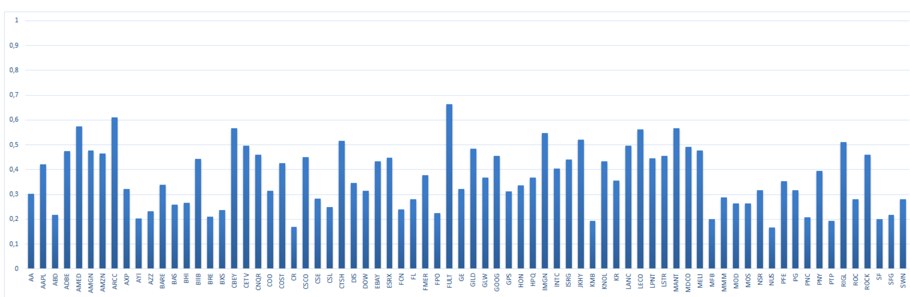
Figure 5: **Market share of each U.S. Stock Exchange over the reference sample period for the 74 stocks of our sample**



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.

In order to zoom on Nasdaq only (exchange on which we perform an analysis of the HFT behavior), we present the market share of Nasdaq for each of the 74 sample stocks (Figure 6). Nasdaq market share based on the number of trades is more than 50% in 10 of the 74 sample stocks.

Figure 6: **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 identified on the U.S. equity market, 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 abrupt price changes 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 log 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 abrupt price changes 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 (see Section 1 for a visual illustration), we provide an alternative third methodology by identifying price movements exceeding the 99.9th percentile of the absolute log 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 abrupt price changes 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.

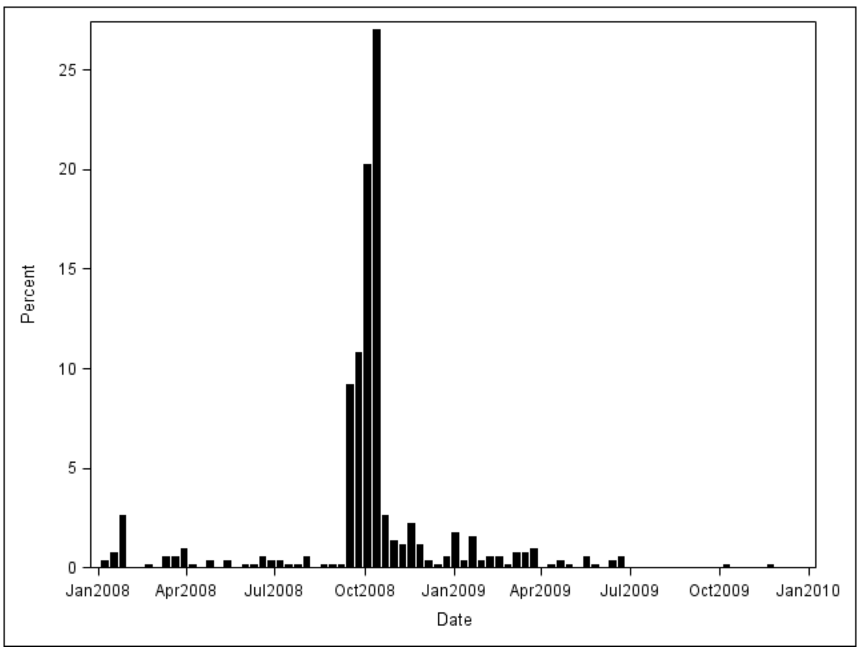
2.6 Descriptive statistics

We present descriptive statistics based on the three different identification methods we use in the paper for MFCs, open-close EPMs and high-low EPMs, keeping in mind MFCs are the focal point of the paper while open-close EPMs and high-low EPMs are comparison points.

Distribution of mini flash crashes

Figure 7 reports the distribution of MFCs over the sample period 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 (September 15, 2008 - September 19, 2008) and 316 MFCs in the month (September 15, 2008 - October 14, 2008) following the news of Lehman Brothers' bankruptcy thus representing respectively about 9% and 62% of all MFCs in our sample.

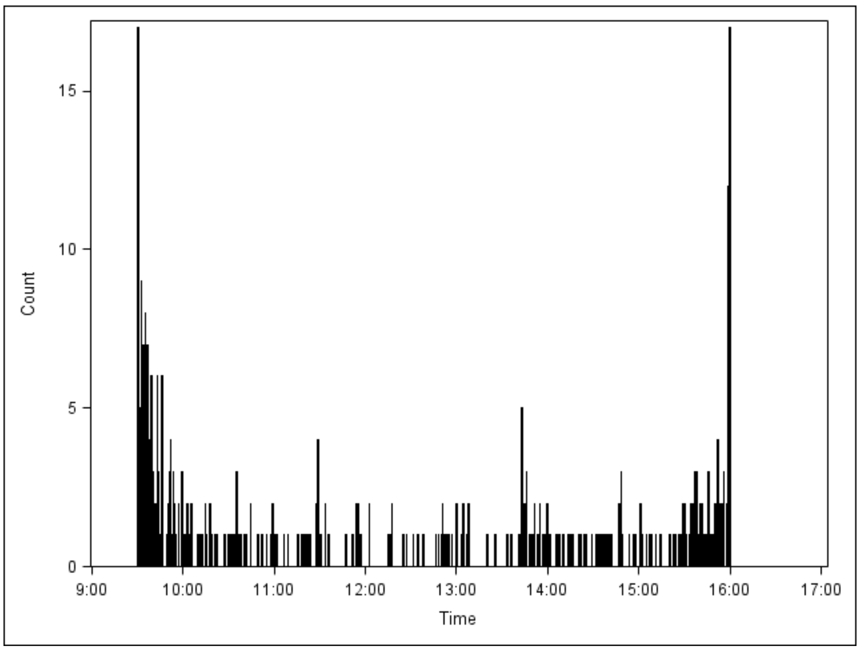
Figure 7: **Distribution of MFCs over the sample period**



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

Figure 8 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 (about 27%) occur in the first and last five minutes of the trading day (about 16% of MFCs occur in the first five minutes while about 12% of MFCs occur in the last five minutes) and more than half of MFCs (about 55%) occur in the first and last half hour of the trading day (about 33% of MFCs occur in the first half hour and about 22% of MFCs occur in the last half hour) so that the overall intraday distribution is U-shaped.

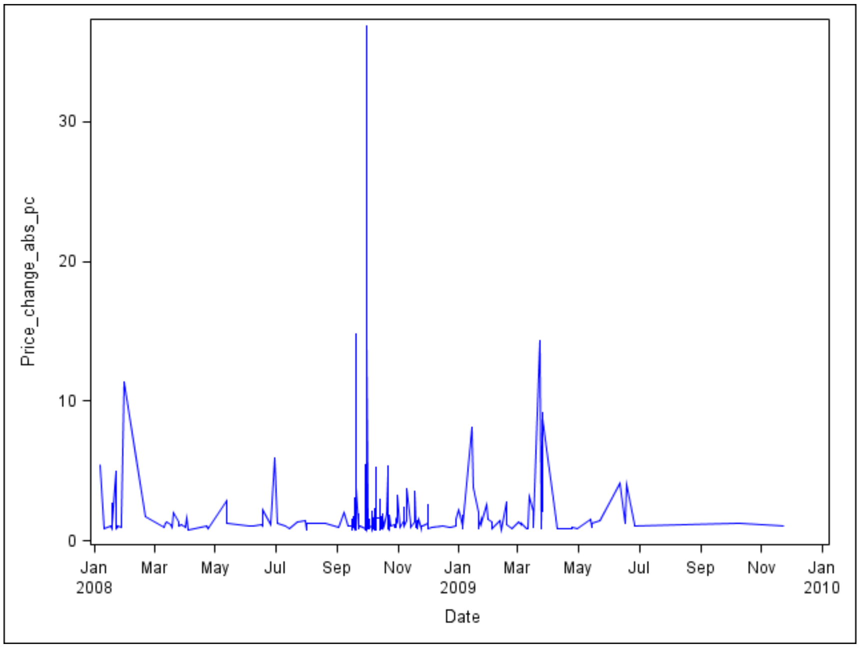
Figure 8: **Intraday distribution of MFCs**



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

Finally, we report the distribution of MFCs over the full sample period (x-axis) taking into account the absolute percentage price change (y-axis) in Figure 9. Not surprisingly, we find that the most extreme crashes ($\geq$ 30%) occur around the bankruptcy of Lehman Brothers. Still, crashes $\geq$ 5% are present all over the sample period. One has to keep in mind that these crashes occur in time intervals lower than 1.5 second so that they are both ultra fast and extreme in terms of magnitude (absolute log return price change).

Figure 9: **Distribution of MFCs with Absolute Price Change in Percent**



The figure plots the distribution of MFCs over the sample period following the Nanex identification method with absolute percentage price change. The data are from Tickdata.

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 (<1.5 second in Panel B and =1.5 second in Panels A, C and D of Table **??**). 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.

HFT participation measures

We measure the participation 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*) following Carrion (2013).

We define *Proportion of HFT Trades* as:

$$ProportionHFTtrades_{i,t} = (HH_{i,t} + HN_{i,t} + NH_{i,t})/(HH_{i,t} + HN_{i,t} + NH_{i,t} + NN_{i,t}), \tag{2.1}$$

based on Nasdaq trades.

We define *Proportion of HFT Shares* as:

$$ProportionHFTshares_{i,t} = (HH_{i,t} + HN_{i,t} + NH_{i,t})/(HH_{i,t} + HN_{i,t} + NH_{i,t} + NN_{i,t}), \tag{2.2}$$

based on Nasdaq shares traded.

We define *Proportion of HFT Volume* as:

$$ProportionHFTvolume_{i,t} = (HH_{i,t} + HN_{i,t} + NH_{i,t})/(HH_{i,t} + HN_{i,t} + NH_{i,t} + NN_{i,t}), \tag{2.3}$$

based on Nasdaq dollar volume.

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:

$$Depth_{i,t} = BestAskSize_{i,t} + BestBidSize_{i,t} \tag{2.4}$$

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:

$$DollarDepth_{i,t} = BestAskPrice_{i,t}BestAskSize_{i,t} + BestBidPrice_{i,t}BestBidSize_{i,t} \tag{2.5}$$

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:

$$QuotedSpread_{i,t} = BestAskPrice_{i,t} - BestBidPrice_{i,t} \tag{2.6}$$

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:

$$RelativeSpread_{i,t} = (BestAskPrice_{i,t} - BestBidPrice_{i,t})/Midquote_{i,t} \tag{2.7}$$

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.

HFT activity measures

We capture HFT and NHFT trading activity around MFCs or EPMs via a measure of trade imbalance (a proxy for their respective behavior) following Brogaard et al. (2018). We compute trade imbalance for both HFTs and NHFTs in the following way:

$$(N)HFT^{D} = (N)HFT^{D^{+}} - (N)HFT^{D^{-}} \tag{2.8}$$

$$(N)HFT^{S} = (N)HFT^{S^{+}} - (N)HFT^{S^{-}} \tag{2.9}$$

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, and where $(N)HFT^{D^{-}}$ and $(N)HFT^{S^{-}}$ represent the liquidity demanded and supplied in the opposite direction of the MFC.

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).

Mini flash crashes

We first investigate the general characteristics of our MFC sample and we carry out a similar investigation on our EPM samples (open-close EPMs and high-low EPMs).

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). We use a variable sub-1.5-second interval for Panel B and a fixed 1.5-second interval for Panels A, C and D.

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

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.

Panel C reports the descriptive statistics for the sample of 29,358 open-close EPMs following the identification method used by Brogaard et al. (2018). Similarly to the MFC sample, the open-close EPM sample presents an absolute return, trading activity (as measured by total trades, share volume and dollar volume), and spread (as measured by quoted spread and relative spread) that are substantially (and highly statistically) larger during open-close EPMs than during the average 1.5-second interval of the full sample (Panel A). The mean absolute open-close EPM return is 0.44% while the full sample mean absolute return is 0.0142%. As such, the mean absolute open-close EPM return is more than 30 times larger than the mean absolute full sample return. Trading activity also appears to be materially higher during open-close EPMs. Indeed, while about 4 trades are executed on average per 1.5 second within the full sample (Panel A) we note that about 32 trades are executed on average per 1.5 second during open-close EPMs, i.e. there are on average 8 times more trades per 1.5-second interval during open-close EPMs. In the same way, share volume and dollar volume are about 12 times higher respectively during open-close EPMs 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, 1,076,800 shares ($436,970) are traded on average per 1.5 second-interval during open-close EPMs. Moreover, the quoted spread is 3.75 times (2.4 times) higher and the relative spread is more than 3.1 times (2.8 times) higher during open-close EPMs based on the mean (median) when compared to the full sample. Finally, the liquidity picture as represented by depth and dollar depth indicates that liquidity dries up considerably during open-close EPMs, with a difference between the open-close EPM sample mean and the full sample mean that is highly statistically significant for both depth and dollar depth.

Panel D reports the descriptive statistics for the sample of 29,362 high-low EPMs following an alternative identification method we propose. Similarly to both the MFC and open-close samples, the high-low EPM sample presents an absolute return, trading activity (as measured by total trades, share volume and dollar volume), and spread (as measured by quoted spread and relative spread) that are substantially larger during high-low EPMs than during the average 1.5-second interval of the full sample (Panel A). The mean absolute high-low EPM return is 0.5481% while the full sample mean absolute return is 0.0142%. As such, the mean absolute high-low EPM return is more than 38 times larger than the mean absolute full sample return. Trading activity also appears to be materially higher during high-low EPMs. Indeed, while about 4 trades are executed on average per 1.5 second within the full sample (Panel A) we note that about 13 trades are executed on average per 1.5 second during high-low EPMs, i.e. there are on average 3 times more trades per 1.5-second interval during high-low EPMs. In the same way, share volume and dollar volume are both about 15 times higher during high-low EPMs 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, 13,667 shares ($532,620) are traded on average per 1.5 second-interval during high-low EPMs. Moreover, the quoted spread is about 4 times (3.2 times) higher and the relative spread is more than 3.5 times (3.5 times) higher during high-low EPMs based on the mean (median) when compared to the full sample. Finally, the liquidity picture as represented by depth and dollar depth indicates that liquidity decreases considerably during high-low EPMs, with a difference between the high-low EPM sample mean and the full sample mean that is highly statistically significant for both depth and dollar depth.

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 respectively representing about 48% and 52% of MFCs in our sample. When comparing the down MFC subsample and the up MFC subsample, we find that the mean (median) of most variables are very close from one another between the two groups so that the difference of means between the variables of both groups are not statistically significant in most cases. An exception is observed for the following four variables: Absolute return (significant at the 5% level), Quoted spread (significant at the 5% level), Relative spread (significant at the 5% level) and Total trades on Nasdaq (significant at the 10% level).

We also 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).

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

Panel A: Full sample					
	Minimum	Median	Mean	Maximum	Std Dev
Absolute return, %	0	0	0.0124	19.64	0.0345
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 shares (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
Depth	2	6	23.67	100,816	99.32
Dollar Depth	6.34	152.3	363.93	2,225,817	1805.47
Quoted spread, $	0.01	0.02	0.0413	113.91	0.1137
Relative spread, %	0.0022	0.0633	0.1224	56.21	0.7707
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	1.1178	1.6676***	35.29	2.3137
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	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, %	0.1098	0.3725	0.4402***	19.64	0.3962
Total trades (all U.S. exchanges)	0	13	31.82***	1,054	55.20
Total trades (Nasdaq)	0	4	14.58***	672	29.25
Proportion HFT trades (Nasdaq)	0	0.6154	0.5272***	1	0.3990
Proportion HFT shares (Nasdaq)	0	0.5222	0.4968***	1	0.4015
Proportion HFT volume (Nasdaq)	0	0.5222	0.4968***	1	0.4015
Share volume	0	2,000	10,768***	10,356,583	77,853
Dollar volume	0	64,739	436,970***	1,458,100,000	8,766,984
Depth	2	5.08	12.80***	1,824	41.69
Dollar depth	6.58	121.41	231.18***	42,745	748
Quoted spread, $	0.01	0.0486	0.1508***	113.91	1.3485
Relative spread, %	0,0025	0,1680	0,3724***	42.89	0.8061
N	29,358				
Panel D: High-Low EPM sample					
	Minimum	Median	Mean	Maximum	Std Dev
Absolute return, %	0.1358	0.4569	0.5480***	30.35	0.5684
Total trades (all U.S. exchanges)	0	13	31.8174***	1,054	55.20
Total trades (Nasdaq)	0	4	14.79***	672	29.83
Proportion HFT trades (Nasdaq)	0	0.6	0.5147***	1	0.3978
Proportion HFT shares (Nasdaq)	0	0.5	0.4834***	1	0.3993
Proportion HFT volume (Nasdaq)	0	0.5002	0.4834***	1	0.3992
Share volume	0	2,100	13,667***	28,368,232	200,287.8
Dollar volume	0	67,990	532,620***	1,458,142,722	9,897,589
Depth	2	4.80	12.31***	1,824	45.80
Dollar depth	6.58	117.65	231.85***	75,122	888.86
Quoted spread, $	0.01	0.0653	0.1724***	113.91	1.36
Relative spread, %	0,0024	0,2223	0,4356***	42.89	0.84
N	29,362				

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.

Mini flash crashes by market capitalization

Among the 120 Nasdaq stocks of the Nasdaq HFT database, the 74 stocks at our disposal all suffer at least one mini flash crash over the sample period based on the Nanex identification method so that at least about 62% of the equities included in the Nasdaq HFT database suffer at least one MFC over the period 2008-2010.[10] The original dataset being made of 40 large stocks, 40 medium stocks and 40 small stocks, the proportion of the large, medium and small subsamples represented in our sample is the following: 80% of the initial large cap subsample, 75% of the initial midcap subsample and 30% of the initial small cap subsample. As a consequence, within our sample of 74 stocks impacted by MFCs, about 43% are large stocks, about 41% are medium stocks and about 16% are small stocks.

We find a total of 510 MFCs over the sample period. Among these MFCs, about 83% impact large stocks, about 15% impact medium stocks and about 2% only impact small stocks within our sample based on the Nanex identification method. However, in order to account for sample size (32 large stocks vs 30 medium stocks vs 12 small stocks), we compute the MFC per stock ratio for each market capitalization subsample. We observe 13.2 MFC per large stock, 2.5 MFC per medium stock and 0.9 MFC per small stock within our sample. As such, we note that the overwhelming majority of MFCs occur on large stocks (sometimes within the same day, the same hour or even within the same minute), while this does not prevent medium and small 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.

We note several interesting characteristics when focusing on market capitalization (Table **??**). First, we note that large, medium and small 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 capitalization 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 capitalization group 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, medium and small stocks respectively. Fourth, based on the proportion of HFT trades on Nasdaq, it appears that HFTs are more active on large stocks

[10]It is possible though that some or all of the other 46 stocks that are not included in our sample also suffer at least one MFC. As such, the descriptive statistics we present in this subsection can be considered as conservative.

(mean of 0.66 and median of 0.73) than on medium stocks (mean of 0.48 and median of 0.49) and small 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 stocks (0.41%) than on medium stocks (0.50%) and small stocks (0.61%), which would tend to indicate that the lower the market capitalization of the stock the higher the impact of MFCs on relative bid-ask spreads.

We also report summary statistics on open-close EPMs (Table 3) and high-low EPMs (Table 4) 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 on Nasdaq cannot be responsible for extreme price movements occurring in medium and small stocks since they do not play any active role in them most of the time (even though HFTs on other exchanges could be responsible for such extreme price movements). On the contrary, we note that HFTs keep playing an active role in medium and small stocks during MFCs, even though their activity is reduced when compared to large cap stocks.

Table 2: Summary statistics of MFCs by market cap

Large					
	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	1.0824	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	0.6576***	1	0.2697
Proportion HFT shares (Nasdaq)	0	0.6129	0.5776***	1	0.2892
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	95.47
Dollar depth	49.56	195.23	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					
	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	1.9159***	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					
	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***	0.9667	0.3162
Proportion HFT shares (Nasdaq)	0	0.25	0.3583***	0.9857	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	51.24	16.03
Dollar depth	23.57	66.69	140.76	505.71	181.32
Quoted spread, $	0.0320	0.0772	0.1142	0.6017	0.1637
Relative spread, %	0.1464	0.3281	0.6127	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. The table also reports univariate tests for means differences. Asterisks ***, ** and * indicate statistical significance at the 1%, 5% and 10% significance levels.

Table 3: Summary statistics of open-close EPMs by market cap

Large					
	Minimum	Median	Mean	Maximum	Std Dev
Absolute return, %	0.1489	0.3647	0.4121***	21.86	0.4005
Total trades (all U.S. exchanges)	0	27	49.10***	1,054	66.06
Total trades (Nasdaq)	0	12	24.59***	672	36.66
Proportion HFT trades (Nasdaq)	0	0.7872	0.6994***	1	0.2998
Proportion HFT shares (Nasdaq)	0	0.7333	0.6516***	1	0.3235
Proportion HFT volume (Nasdaq)	0	0.7335	0.6516***	1	0.3235
Share volume	0	4,668	17,015***	10,356,583	100,781
Dollar volume	0	159,431	709,866***	1,458,142,722	11,429,108
Depth	2	6.03	17.24***	1,824	50.90
Dollar depth	26.65	159.07	313.50	42,745	909.51
Quoted spread, $	0.01	0.03	0.1362***	113.91	1.7430
Relative spread, %	0.0025	0.0905	0.1736***	42.89	0.6492
N	17,242				
Medium					
	Minimum	Median	Mean	Maximum	Std Dev
Absolute return, %	0.1099	0.3751	0.4720***	15.58	0.3944
Total trades (all U.S. exchanges)	0	4	7.97***	225	12.32
Total trades (Nasdaq)	0	1	3.34***	130	6.10
Proportion HFT trades (Nasdaq)	0	0	0.3544***	1	0.4107
Proportion HFT shares (Nasdaq)	0	0	0.3418***	1	0.4126
Proportion HFT volume (Nasdaq)	0	0	0.3418***	1	0.4126
Share volume	0	600	2,033***	452,584	8,873
Dollar volume	0	15,115	54,672***	17,789,123	324,996
Depth	2	4.25	6.75	1,051	23.70
Dollar depth	10.90	93.48	126.10	21,089	439.33
Quoted spread, $	0.01	0.1044	0.1832***	10.81	0.3135
Relative spread, %	0.0285	0.3733	0.5964***	18.40	0.8104
N	9,579				
Small					
	Minimum	Median	Mean	Maximum	Std Dev
Absolute return, %	0.1859	0.4129	0.5125***	5.9822	0.4031
Total trades (all U.S. exchanges)	0	2	4.42***	352	9.97
Total trades (Nasdaq)	0	1	2.04***	349	8.59
Proportion HFT trades (Nasdaq)	0	0	0.2384***	1	0.3792
Proportion HFT shares (Nasdaq)	0	0	0.2360***	1	0.3828
Proportion HFT volume (Nasdaq)	0	0	0.2360***	1	0.3828
Share volume	0	300	1,292***	530,200	13,122
Dollar volume	0	4,695	25,758***	9,277,725	249,238
Depth	2	4	5.53*	181	7.44
Dollar depth	6.58	54.77	68.46	1,152	65.40
Quoted spread, $	0.01	0.0845	0.1276***	2.00	0.1465
Relative spread, %	0.0398	0.5091	0.8777***	90.74	1.21
N	2,537				

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. The table also reports univariate tests for means differences. Asterisks ***, ** and * indicate statistical significance at the 1%, 5% and 10% significance levels.

Table 4: Summary statistics of high-low EPMs by market cap

Large					
	Minimum	Median	Mean	Maximum	Std Dev
Absolute return, %	0.1754	0.4472	0.5132***	26.51	0.6035
Total trades (all U.S. exchanges)	0	27	49.87***	1,054	67.20
Total trades (Nasdaq)	0	12	24.85***	672	37.41
Proportion HFT trades (Nasdaq)	0	0.7692	0.6832***	1	0.3062
Proportion HFT shares (Nasdaq)	0	0.7064	0.6338***	1	0.3277
Proportion HFT volume (Nasdaq)	0	0.7064	0.6338***	1	0.3277
Share volume	0	4,900	21,856***	28,368,232	260,915
Dollar volume	0	172,810	870,702***	1,458,142,722	12,902,407
Depth	2	5.54	16.45***	1,824	56.24
Dollar depth	26.67	151.32	315.16***	75,122	1,098
Quoted spread, $	0.0100	0.0401	0.1572***	113.91	1.7609
Relative spread, %	0.0024	0.1145	0.2119***	42.89	0.6560
N	17,243				
Medium					
	Minimum	Median	Mean	Maximum	Std Dev
Absolute return, %	0.1357	0.4635	0.5858***	16.30	0.4975
Total trades (all U.S. exchanges)	0	4	8.40***	225	13.53
Total trades (Nasdaq)	0	1	3.50***	150	6.64
Proportion HFT trades (Nasdaq)	0	0	0.3457***	1	0.4068
Proportion HFT shares (Nasdaq)	0	0	0.3333***	1	0.4085
Proportion HFT volume (Nasdaq)	0	0	0.3333***	1	0.4085
Share volume	0	600	2,189***	452,584	9,163
Dollar volume	0	14,978	57,926***	17,789,123	326,795
Depth	2	4.13	6.67	1,051	25.43
Dollar depth	10.90	93.02	124.98	21,089	467
Quoted spread, $	0.01	0.126	0.2065***	10.81	0.3131
Relative spread, %	0.0317	0.4470	0.6803***	18.40	0.8344
N	9,582				
Small					
	Minimum	Median	Mean	Maximum	Std Dev
Absolute return, %	0.2272	0.5075	0.6429***	9.07	0.5136
Total trades (all U.S. exchanges)	0	2	4.64***	352	10.27
Total trades (Nasdaq)	0	1	2.08***	349	8.69
Proportion HFT trades (Nasdaq)	0	0	0.2300***	1	0.3737
Proportion HFT shares (Nasdaq)	0	0	0.2265***	1	0.3759
Proportion HFT volume (Nasdaq)	0	0	0.2265***	1	0.3759
Share volume	0	300	1,356***	530,200	13,137
Dollar volume	0	4,626	27,677***	9,277,725	249,741
Depth	2	3.93	5.48**	181	7.16
Dollar depth	6.58	55.60	69.27	1,152	62.51
Quoted spread, $	0.0113	0.1063	0.1476***	2.00	0.1477
Relative spread, %	0.0398	0.6113	1.03***	9.07	1.34
N	2,537				

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. The table also reports univariate tests for means differences. Asterisks ***, ** and * indicate statistical significance at the 1%, 5% and 10% significance levels.

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 a bit more than 58% of all MFCs in our sample (Table 5).

Table 5: 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 6) 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 6: Summary statistics of MFCs by sector

Information Technology					
	Minimum	Median	Mean	Maximum	Std Dev
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 (Table 5) to open-close EPMs and high-low EPMs by sector (Table 7 and Table 8 respectively), we observe that the ranking is the same so that the sectors that suffer the highest number of abrupt price changes, 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). That being said, we note that the combination MFC/Sector can lead to very small samples in some cases. In addition, it is possible for HFTs not to be present on Nasdaq during MFCs but on other exchanges.

Table 7: 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 8: 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.

Mini flash crashes by U.S. Exchange

We finally investigate the distribution of trades among the different U.S. stock exchanges during MFCs (Table Table 9). 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 9: 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.

2.7 Methodology

We initially capture HFT (respectively NHFT) trading activity around MFCs via a measure of directional trade imbalance, which enables us to determine the role played by both market participants during the phase preceding the crash (pre-crash), during the crash (crash), as well as during the recovery phase (post-crash). We then proceed in two steps.

First, we run several multivariate regressions to study the net trading contribution of HFTs (respectively NHFTs) 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.

Second, we run logistic regressions to measure the probability for a stock to undergo a mini flash crash as a function of lagged values of HFT^{NET}, absolute log return, share volume and relative spread. Results are presented in the following section.

The methodology is replicated on open-close EPMs and high-low EPMs so as to compare the behavior of HFTs (respectively NHFTs) during both types of crashes (MFCs vs EPMs).

3 Empirical results

3.1 HFT trading activity around mini flash crashes

We capture HFT and NHFT trading activity around mini flash crashes via a measure of net directional trade imbalance following Brogaard et al. (2018) so as to see who among HFTs and NHFTs trigger the crash, exacerbate the crash and participate to resiliency right after the crash.[11]

The measure of directional trade imbalance is computed in three steps. First, we compute *(N)*HFT^{D^+}, *(N)*HFT^{D^-}, *(N)*HFT^{S^+} and *(N)*HFT^{S^-}, which respectively represent the liquidity demanded in the direction of the MFC (vicious), the liquidity demanded in the opposite direction of the MFC (virtuous), the liquidity supplied in the direction of the MFC (vicious) and the liquidity supplied in the opposite direction of the MFC (virtuous). Second, we compute *(N)*HFT^{D} and *(N)*HFT^{S}. *(N)*HFT^{D} represents liquidity demanded by (N)HFTs and is computed as the substraction of (N)HFT^{D^-} from (N)HFT^{D^+}, that is to say liquidity demanded

[11] Bellia et al. (2018) use a similar measure of net directional trade imbalance except they use monetary volume instead of traded shares.

in the direction of the MFC (vicious) minus liquidity demanded in the opposite direction of the MFC (virtuous). *(N)*HFT^S represents liquidity supplied by (N)HFTs and is computed as the substraction of (N)HFT^{S^-} from (N)HFT^{S^+}, that is to say liquidity supplied in the direction of the MFC (vicious) minus liquidity supplied in the opposite direction of the MFC (virtuous).

Net imbalance (HFT^{NET} and $NHFT^{NET}$) informs us on the direction of net trading activity vis-à-vis the MFC. A positive net imbalance implies trading activity in the direction of the MFC on an aggregated basis, which can be considered a vicious behavior whether this positive imbalance is observed during the pre-crash phase (crash triggering), the crash phase (crash exacerbation) or during the post-crash phase (resiliency prevention). On the contrary, a negative net imbalance implies trading activity in the opposite direction of the MFC on an aggregated basis, which can be considered a virtuous behavior in the pre-crash phase (crash prevention), the crash phase (crash absorption) or during the post-crash phase (resiliency promotion).

We compute net imbalance of both types of traders at times $t-2$, $t-1$, t, $t+1$ and $t+2$ for MFCs as well as for open-close EPMs and high-low EPMs using fixed 1.5-second intervals. In particular, we focus on time intervals $t-1$, t and $t+1$, 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 when compared to the prior or successive interval.

A filter is applied on the MFC, open-close EPM and high-low EPM samples so as to prevent consecutive crashes to pollute the different windows from time $t-2$ to time $t+2$. For example, if a crash occurs at time t and another crash occurs at time $t+1$ then both crashes are removed from our sample so as to prevent any bias in the computation of the directional trade imbalance metrics. Our three initial samples (MFCs, open-close EPMs, high-low EPMs) are screened for crashes that occur in at least two of five consecutive time windows. We end up with final samples of 405 MFCs, 22,021 open-close EPMs and 17,840 high-low EPMs after the cleaning.

3.2 Trade imbalance around MFCs

Table 10: Trade imbalance around MFCs at 1.5-second intervals

	t-2	t-1	t	t+1	t+2
		pre-crash	crash	post-crash	
HFT^{NET}	51.8222	-4.9383	-119.7	29.1827	-2.9852
HFT^{D}	-17.1012	16.4346	196.2*	4.1235	-2.8173
HFT^{S}	68.9235*	-21.3728	-315.9***	25.0593	-0.1679
$NHFT^{NET}$	-51.8222	4.9383	119.7	-29.1827	2.9852
$NHFT^{D}$	-206.9	61.4815	1,262.3***	-41.7235	86.3654
$NHFT^{S}$	155.1	-56.5432	-1,142.6***	12.540	-83.3802*

The table reports directional 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 3 seconds prior to the crash ($t-2$, $t-1$) and 3 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.

Before the crash

Just before the crash, at time interval $t-1$, we note that HFT^{S} is negative ($HFT^{S} = -21.3728$) while HFT^{D} is positive ($HFT^{D} = 16.4346$), which would imply HFTs both supplying liquidity in the opposite direction of the MFC to come while also demanding liquidity in the direction of the MFC to come. We note that HFTs supply 1.3 times more liquidity in the opposite direction of the crash to come than they demand (21.3728/16.4346). That being said, results are not significant so that one cannot reject the null hypothesis according to which: (1) liquidity demanded by HFTs in the direction of the crash to come at time $t-1$ is zero in the population, (2) liquidity supplied by HFTs in the opposite direction of the crash at time $t-1$ is zero in the population, (3) net liquidity from HFTs at time $t-1$ is zero in the population.

NHFTs appear to be the ones that are active during this phase. While NHFTs supply about 2.6 times (=56.5432/21.3728) more liquidity in the opposite direction of the crash than HFTs during this time interval, they also seem to consume more liquidity in the direction of the MFC to come ($NHFT^{D}$=61.4815). Again, results are not significant so that one cannot reject the null hypothesis according to which: (1) liquidity demanded by NHFTs in the direction of the crash to come at time $t-1$ is zero in the population, (2) liquidity supplied by NHFTs in the opposite direction of the crash at time $t-1$ is zero in the population, (3) net liquidity from NHFTs at time $t-1$ is zero in the population. As such one cannot conclude about who triggers MFCs.

During the crash

At time interval t, we note that HFT^{NET} ($NHFT^{NET}$) is negative (positive) but not statistically significant, which implies that one cannot reject the null hypothesis according to which net liquidity from both HFTs and NHFTs is zero in the population. However, interesting stylized facts are worth mentioning. First, we observe that NHFTs are the ones who become particularly active during the crash. Indeed, NHFTs consume 20.5 times more liquidity in the direction of the crash at time interval t than within the preceding time interval (1,262.3/61.4815) while they also supply 20.2 times more liquidity in the opposite direction of the crash at time interval t than within the preceding time interval (1,142.6/56.5432). As for HFTs, they consume 11.9 times more liquidity in the direction of the crash at time interval t than within the preceding time interval (196.2/16.4346) and supply 14.8 times more liquidity in the opposite direction of the crash at time interval t than within the preceding time interval (315.9/21.3728). Overall, NHFTs supply 3.6 times (=1,142.6/315.9) more liquidity in the opposite direction of the crash than HFTs during the crash phase but also consume 6.4 times (=1,262.3/196.2) more liquidity in the direction of the MFC than HFTs. In net values however, no conclusion can be reached.

After the crash

At time interval $t + 1$, results are not significant so that one cannot reject the null hypothesis according to which: (1) liquidity demanded by both HFTs and NHFTs in the direction of the crash is zero in the population, (2) liquidity supplied by both HFTs and NHFTs in the opposite direction of the crash is zero in the population, (3) net liquidity from both HFTs and NHFTs is zero in the population. As such one cannot conclude about who enables stock prices to recover after a mini flash crash.

To sum it up, due to a lack of statistical significance, one cannot conclude about who triggers MFCs, who exacerbates the crash and who enables stock prices to recover. We thus 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.

3.3 Trade imbalance around open-close EPMs

Table 11: Trade imbalance around open-close EPMs at 1.5-second intervals

	t-2	t-1	t	t+1	t+2
		pre-crash	crash	post-crash	
HFT^{NET}	2.9854**	0.8153	14.2781***	9.5093***	2.0845
HFT^{D}	0.7710	11.0562***	116.5***	5.6254***	-3.9869***
HFT^{S}	2.2144**	-10.2409***	-102.2***	3.8839***	6.0714***
$NHFT^{NET}$	-2.9854**	-0.8153	-14.2781***	-9.5093***	-2.0845
$NHFT^{D}$	1.3620	35.1578***	260.6***	28.7946***	8.2640***
$NHFT^{S}$	-4.3474	-35.9731***	-274.9***	-38.3039***	-10.3485***

The table reports directional trade imbalance (based on share volume) around open-close extreme price movements (open-close EPMs) computed from Nasdaq. Time interval t is the fixed-1.5-second interval corresponding to the crash. We also report the trade imbalance figures 3 seconds prior to the crash ($t-2$, $t-1$) and 3 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.

Before the crash

Just before the crash, at time interval $t-1$, we note that HFT^{S} is highly significantly negative ($HFT^{S} = -10.2409^{***}$) while HFT^{D} is highly significantly positive ($HFT^{D} = 11.0562^{***}$), which would imply HFTs both supplying liquidity in the opposite direction of the crash to come while also demanding liquidity in the direction of the crash to come. We note that HFTs demand 1.08 times more liquidity in the direction of the crash to come than they supply (11.0562/10.2409). In net values however, no statistical conclusion can be reached.

NHFTs appear to be the ones that are (highly significantly) active during this phase. While NHFTs supply about 3.5 times (=35.9731/10.2409) more liquidity in the opposite direction of the crash than HFTs during this time interval, they are the ones who consume (highly significantly) more liquidity in the direction of the EPM to come ($NHFT^{D}$=35.1578***). In net values however, no statistical conclusion can be reached and one cannot conclude that HFTs trigger EPMs.

We note that the behavior of both types of traders, HFTs and NHFTs, is more or less similar from $t-2$ to $t+2$ when one focuses on net values only, Their behavior is more heterogeneous when one focuses on demand and supply only.

During the crash

During the crash, at time t, we note that HFT^{NET} ($NHFT^{NET}$) is positive (negative) and highly statistically significant, which implies that HFTs (as a group) exacerbate the liquidity imbalance during the crash phase while NHFTs (as a group) counteract on this liquidity imbalance. Indeed, it appears that HFTs trade in the direction of the EPM on an aggregated basis (vicious behavior) while NHFTs trade in the opposite direction of the crash on an aggregated basis (virtuous behavior).

In more details, we note that HFTs and NHFTs become very active during the crash, NHFTs supplying 2.7 times (=274.9/102.2) more liquidity in the opposite direction of the crash than HFTs while also consuming 2.2 times (=260.6/116.5) more liquidity in the direction of the EPM than HFTs. Moreover, we note that the change in liquidity from the pre-crash phase to the crash phase follows a proportional relationship for both HFTs and NHFTs. On the one hand, HFTs supply about 10 times more liquidity in the opposite direction of the crash at time t than during the prior time interval (102.2/10.2409), to be compared to the fact that NHFTs supply about 7.6 times more liquidity in the opposite direction of the crash at time t than during the prior time interval (274.9/35.9731). However, HFTs also consume about 10.5 times more liquidity in the direction of the crash at time t than during the prior time interval (116.5/11.0562), to be compared to the fact that NHFTs consume about 7.4 times more liquidity in the direction of the crash at time t than during the prior time interval (260.6/35.1578).

After the crash

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) ($NHFT^{NET}$=-9.5093***) while HFTs keep demanding liquidity in the direction of the crash both at time $t+1$ (HFT^{NET}=9.5093***).

To sum it up, HFTs do not appear to trigger open-close EPMs, but they appear to exacerbate the crash during open-close EPMs and do not appear to contribute to the price recovery. On the contrary, NHFTs appear to be the ones who counteract on the directional trade imbalance during the crash and who contribute to the price recovery.

3.4 Trade imbalance around high-low EPMs

Table 12: Trade imbalance around high-low EPMs at 1.5-second intervals

	t-2	t-1	t	t+1	t+2
		pre-crash	crash	post-crash	
HFT^{NET}	1.8178	3.0115	7.4711*	11.2685***	2.2533
HFT^{D}	-0.1455	10.2201***	98.8703***	3.7589**	-3.6143***
HFT^{S}	1.9633	-7.2086***	-91.3993***	7.5096***	5.8677***
$NHFT^{NET}$	-1.8178	-3.0115	-7.4711*	-11.2685***	-2.2533
$NHFT^{D}$	-1.3218	24.8295***	242.2***	19.1936***	6.5059**
$NHFT^{S}$	-0.4960	-27.8410***	-249.7***	-30.4621***	-8.7592***

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 3 seconds prior to the crash ($t-2$, $t-1$) and 3 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.

3.4.1 Before the crash

Just before the crash, at time interval $t-1$, we note that HFT^{S} is highly significantly negative ($HFT^{S} = -7.2086^{***}$) while HFT^{D} is highly significantly positive ($HFT^{D} = 10.2201^{***}$), which would imply HFTs both supplying liquidity in the opposite direction of the crash to come while also demanding liquidity in the direction of the crash to come. We note that HFTs demand 1.4 times more liquidity in the direction of the crash to come than they supply (10.2201/7.2086). In net values however, no statistical conclusion can be reached.

NHFTs appear to be the ones that are (highly significantly) active during this phase. While NHFTs supply about 3.9 times (=27.8410/7.2086) more liquidity in the opposite direction of the crash than HFTs during this time interval, they are the ones who consume (highly significantly) more liquidity in the direction of the EPM to come ($NHFT^{D}$=24.8295***). In net values however, no statistical conclusion can be reached and one cannot conclude that HFTs trigger EPMs.

We note that the behavior of both types of traders, HFTs and NHFTs, is more or less similar from $t-2$ to $t+2$ when one focuses on net values only, Their behavior is more heterogeneous when one focuses on demand and supply only.

During the crash

During the crash, at time t, we note that HFT^{NET} ($NHFT^{NET}$) is positive (negative) and statistically significant (at the 10% level), which implies that HFTs (as a group) exacerbate the liquidity imbalance during the crash phase while NHFTs (as a group) counteract on this liquidity imbalance. Indeed, it appears that HFTs trade in the direction of the EPM on an aggregated basis (vicious behavior) while NHFTs trade in the opposite direction of the crash on an aggregated basis (virtuous behavior).

In more details, we note that HFTs and NHFTs become very active during the crash, NHFTs supplying 2.7 times (=249.7/91.3993) more liquidity in the opposite direction of the crash than HFTs while also consuming 2.4 times (=242.2/98.8703) more liquidity in the direction of the EPM than HFTs. Moreover, we note that the change in liquidity from the pre-crash phase to the crash phase follows a proportional relationship for both HFTs and NHFTs. On the one hand, HFTs supply about 12.7 times more liquidity in the opposite direction of the crash at time t than during the prior time interval (91.3993/7.2086), to be compared to the fact that NHFTs supply about 9 times more liquidity in the opposite direction of the crash at time t than during the prior time interval (249.7/27.8410). However, HFTs also consume about 9.7 times more liquidity in the direction of the crash at time t than during the prior time interval (98.8703/10.2201), to be compared to the fact that NHFTs consume about9.8 times more liquidity in the direction of the crash at time t than during the prior time interval (242.2/24.8295).

After the crash

NHFTs appear to be the ones that contribute to resiliency (at a 1% level of significance) during high-low 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) ($NHFT^{NET}$=-11.2685***) while HFTs keep demanding liquidity in the direction of the crash at time $t+1$ (HFT^{NET}=11.2685***).

To sum it up, our findings are similar whether we use open-close EPMs and high-low EPMs. Again, HFTs do not appear to trigger high-low EPMs, but they appear to exacerbate the crash during high-low EPMs and do not appear to contribute to the price recovery that follows the crash. On the contrary, NHFTs appear to be the ones who counteract on the directional trade imbalance during the crash and who contribute to the price recovery, helping the stock price to recover (in line with the findings of Bellia et al. (2018).

3.5 HFT activity during crashes

We now focus on the trading activity of HFTs during mini flash crashes and perform a similar analysis on the trading activity of HFTs during both open-close and high-low extreme price movements. To do so, we run several multivariate regressions in which HFT^{NET} is a function of MFC (respectively open-close EPM and high-low EPM), absolute log return, share volume and relative spread. The different regressions account for the type of crash (standalone vs simultaneous) as well as the time of occurrence of the crash (regular hours vs extreme hours).[12]

An important aspect of our methodological approach must be emphasized at this point. Instead of considering all MFCs (respectively EPMs) occurring on all U.S. exchanges, we consider MFCs (respectively EPMs) for which the proportion of transactions occurring on Nasdaq during the interval represents at least 50% of all transactions on U.S. stock exchanges. Indeed, transactions may occur on different exchanges during mini flash crashes (respectively extreme price movements) and they do sometimes occur outside of the Nasdaq exchange. By doing so, we filter out MFCs (respectively EPMs) 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 (respectively EPM samples), i.e. MFCs (EPMs) for which the proportion of transactions on Nasdaq is comprised between 0% and 100%, thus taking into account all MFCs (EPMs), including MFCs (EPMs) where no transaction is observed on Nasdaq, and on (2) an MFC subsample (respectively EPM subsamples) in which the proportion of transactions on Nasdaq is equal to 100%, thus taking into account MFCs (EPMs) where all transactions during the crash occur on Nasdaq exclusively.[13] Moreover, we use an event window of 30 minutes (i.e. 15 minutes before the crash and 15 minutes after the crash) in our base case methodology. We also test the robustness of our results by running our regressions on 15-minute and 60-minute windows. We find similar results whatever the window size we use.

Mini flash crashes

Following Brogaard et al. (2018), we run multivariate regressions in which HFT^{NET} is a function of mini flash crashes, absolute log return, share volume and relative spread. Each specification of the model either takes into account the type of MFC (all vs standalone vs simul-

[12]In the paper, extreme hours refer to opening and closing hours, i.e. from 9:00 a.m. to 9:05 a.m. and from 15:55 p.m. to 16:00 p.m. The term 'extreme' in "extreme hours" is thus not related whatsoever to the term 'extreme' in "extreme price movements".

[13]Results are available upon request.

taneous) or the time of occurrence of the MFC (regular hours vs extreme hours). Results for the different specifications are presented in Table 13.

First, we focus on all MFCs and do not discriminate MFCs depending on their type (standalone or simultaneous) or the time of occurence of the crash (extreme hours). As such, our central multivariate regression is the following:

$$HFT_{it}^{NET} = \alpha_i + \beta_1 MFC_{it} + \beta_2 AbsRet_{it} + \beta_3 SV_{it} + \beta_4 RS_{it} + (Lags_{kit-\sigma}\gamma_{k\sigma}) + \epsilon_{it} \tag{3.1}$$

where HFT^{NET} is the difference between HFT^D and HFT^S, MFC_{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 log return, SV is the share volume and RS is the relative spread. $Lags_{kit-\sigma}\gamma_{k\sigma}$ is a vector of 10 lags for both the dependent variable and all of the independent variables of the regression, with $\sigma \in \{1,2,...,10\}$ 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.

Second, we 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[14] :

$$HFT_{it}^{NET} = \alpha_i + \beta_1 MFC_{it}^{STA} + \beta_2 MFC_{it}^{SIM} + \beta_3 AbsRet_{it} + \beta_4 SV_{it} + \beta_5 RS_{it} + (Lags_{kit-\sigma}\gamma_{k\sigma}) + \epsilon_{it} \tag{3.2}$$

where all the variables are as previously defined, MFC_{it}^{STA} is a dummy variable equal to one if the 1.5-second interval t in stock i is identified as a standalone MFC and is equal to zero otherwise, and where MFC_{it}^{SIM} is a dummy variable equal to one if the 1.5-second interval t in stock i is identified as a simultaneous MFC and is equal to zero otherwise.

[14]Note that simultaneous MFCs are not the same as co-EPMs in Brogaard et al. (2018). While Brogaard et al. define co-EPMs as EPMs occurring on two or more stocks within the same 10-second interval, which would correspond to MFCs occurring on two or more stocks within the same 1.5-second interval in this paper, we instead measure simultaneous MFCs as MFCs occurring on two or more stocks within the same minute due to a limited number of MFCs in our sample.

Third, we focus on MFCs accounting for the time of occurence of the crash. We use an interaction variable ($DUMMY\_MFC$ times $DUMMY\_EXTREME\_HOURS$) to capture the additional effect of the MFCs that occur during the first five and last five minutes of the trading day on HFT^{NET} :

$$\begin{aligned} HFT_{it}^{NET} &= \alpha_i + \beta_1 MFC_{it} + \beta_2 EH_{it} + \beta_3 MFC_{it} * EH_{it} \\ &+ \beta_4 AbsRet_{it} + \beta_5 SV_{it} + \beta_6 RS_{it} + (Lags_{kit-\sigma}\gamma_{k\sigma}) + \epsilon_{it} \end{aligned} \tag{3.3}$$

where all the variables are as previously defined, EH_{it} is a dummy variable equal to one if the 1.5-second interval t in stock i is identified as occurring during extreme hours (EH), i.e. during the first five and last five minutes of the trading day and $MFC_{it} * EH_{it}$ is an interaction variable that captures the additional effect of MFCs that occur during extreme hours (EH).

We start by commenting the coefficients of absolute log return, share volume and relative spread and then move to the coefficient associated with the MFC variable (for the different specifications). First, the positive (highly statistically significant) coefficient on the absolute log return variable (whatever the specification) indicates that HFTs tend to demand liquidity in the direction of the return (which is in line with the literature). Second, the (highly statistically significant) coefficient on the relative spread indicates that HFTs provide more liquidity when spreads widen (which is also in line with the literature). Third, we note that share volume is always non statistically significant. Fourth, and more importantly, we note that the coefficient of the MFC dummy variable (whatever the specification) is not statistically significant in the full sample. However, when focusing on different subsamples (large, medium, small) instead of the full sample, we find (in unreported results) that the coefficient of the MFC dummy variable is sometimes statistically significant. Each time the coefficient is statistically significant (whether at the 1%, 5% or 10% confidence level), the sign of the MFC coefficient is always negative, implying that HFTs reduce their liquidity demand during MFCs occurring on large, medium and small stocks and trade in the opposite direction of the crash (in line with the findings of Brogaard et al. (2018) regarding EPMs). Finally, regarding the extreme hour specification, we find (in unreported results) that the extreme hours dummy variable is statistically significant at the 1% confidence level with a negative coefficient (-23.72***), implying that the liquidity demand reduction of HFTs is even more important during the first five and last five minutes of the trading day. As such, the decline in HFT liquidity demand is potentially more pronounced during periods of known market stress (opening and closing hours).

Table 13: Multivariate regressions for Net HFT during MFCs

Proportion of transactions on Nasdaq $\geq$ 50%			
	(3.1)	(3.2)	(3.3)
MFC	-124.36		
$MFC^{STANDALONE}$		-270.64	
$MFC^{SIMULTANEOUS}$		57.95	
$MFC^{EXTREME\_HOURS}$			-119.19
$AbsRet$	4.3076***	4.2780***	4.3167***
SV	-0.0922	-0.1317	-0.0455
RS	-3.5688***	-3.5773***	-3.5273***
$Adj.R^2$	0.0078	0.0060	0.0060
N	5,137,465	5,137,465	5,137,465

The table reports the estimated coefficients of equations 3.1, 3.2 and 3.3, where HFT^{NET} is the difference between HFT^D, the liquidity demanded by HFTs, and HFT^S, the liquidity supplied by HFTs. A negative HFT^{NET} implies an aggregate trading activity in the opposite direction of the crash (virtuous behavior) while a positive HFT^{NET} implies an aggregate trading activity in the direction of the crash (vicious behavior). MFC_{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. $MFC^{STANDALONE}$ is a dummy that captures MFCs that occur on their own, $MFC^{SIMULTANEOUS}$ is a dummy that captures MFCs that occur on several sample stocks within the same minute, $MFC^{EXTREME\_HOURS}$ is an interaction variable that captures the additional effect of MFCs that occur in the first five and last five minutes of the trading day, $AbsRet$ is the absolute log 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 at least 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.

Open-close extreme price movements

In a similar process, we then focus on the trading activity of HFTs during open-close extreme price movements. Our central multivariate regression is the following:

$$HFT_{it}^{NET} = \alpha_i + \beta_1 EPM_{it}^{OC} + \beta_2 AbsRet_{it} + \beta_3 SV_{it} + \beta_4 RS_{it} + (Lags_{kit-\sigma}\gamma_{k\sigma}) + \epsilon_{it} \qquad (3.4)$$

where HFT^{NET} is the difference between HFT^D and HFT^S, EPM_{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, $AbsRet$ is the absolute log 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.

We also focus on both standalone and simultaneous open-close EPMs as well as on extreme-hour open-close EPMs:

$$HFT_{it}^{NET} = \alpha_i + \beta_1 EPM_{it}^{OC-STA} + \beta_2 EPM_{it}^{OC-SIM} + \beta_3 AbsRet_{it} + \beta_4 SV_{it} + \beta_5 RS_{it} + (Lags_{kit-\sigma}\gamma_{k\sigma}) + \epsilon_{it} \tag{3.5}$$

$$\begin{aligned} HFT_{it}^{NET} = \alpha_i + \beta_1 EPM_{it}^{OC} + \beta_2 EH_{it} + \beta_3 EPM_{it}^{OC} * EH_{it} \\ + \beta_4 AbsRet_{it} + \beta_5 SV_{it} + \beta_6 RS_{it} + (Lags_{kit-\sigma}\gamma_{k\sigma}) + \epsilon_{it} \end{aligned} \tag{3.6}$$

Results of the different regressions are presented in Table 14.

We start by commenting the coefficients of absolute log return, share volume and relative spread and then move to the coefficient associated with the open-close EPM variable (for the different specifications). First, the positive (highly statistically significant) coefficient on the absolute log return variable (whatever the specification) indicates that HFTs tend to demand liquidity in the direction of the return (which is in line with the literature). Second, the (highly statistically significant) coefficient on the relative spread indicates that HFTs provide more liquidity when spreads widen (which is also in line with the literature). Third, we note that share volume is always non statistically significant. Fourth, and more importantly, we note that the coefficient of the open-close EPM dummy variable is highly statistically significant in the full sample as well as in the case of standalone EPMs and simultaneous EPMs. However, it is not statistically significant in the case of extreme-hour EPMs.

When focusing on different subsamples (large, medium, small) instead of the full sample, we find (in unreported results) that the coefficient of the open-close EPM dummy variable is significantly negative for large stocks (-36.6851**) but highly significantly positive for small stocks (27.8919***), implying that while HFTs reduce their liquidity demand during open-close EPMs occurring on large stocks, they increase their liquidity demand during open-close EPMs occurring on small stocks and trade in the direction of the crash. Moreover, we find that the coefficient of the simultaneous open-close EPM variable is always negative when statistically significant. As a consequence, while HFTs increase their liquidity demand during co-EPMs (Brogaard et al., 2018), i.e. when multiple EPMs occur within the same interval, due to the need for HFTs to reduce their cumulative exposure to stocks undergoing an EPM, they do not show a similar behavior regarding multiple EPMs occurring within a longer time frame (EPMs occurring within 40 intervals in our case). We conjecture that one minute is enough for HFTs to reset their inventory risk constraints and get their level of risk exposure in order.

Finally, regarding the extreme hour specification, we find (in unreported results) that the extreme hours dummy variable is statistically significant at the 1% confidence level with a negative coefficient (-23.41***), implying that the liquidity demand reduction of HFTs is even more important during the first five and last five minutes of the trading day. As such, the decline in HFT liquidity demand is potentially more pronounced during periods of known market stress (opening and closing hours), similarly to MFCs.

Table 14: Multivariate regressions for Net HFT activity during open-close EPMs

Proportion of transactions on Nasdaq $\geq$ 50%			
	(3.4)	(3.5)	(3.6)
EPM^{OC}	-18.20**		
$EPM^{OC-STANDALONE}$		-25.10**	
$EPM^{OC-SIMULTANEOUS}$		-25.48***	
$EPM^{OC-EXTREME-HOURS}$			-4.36
$AbsRet$	4.5160***	5.2053***	4.2560***
SV	-0.1584	-0.1774	-0.1014
RS	-3.5765***	-3.5642***	-3.4771***
$Adj.R^2$	0.0078	0.0079	0.0080
N	5,137,465	5,137,465	5,137,465

The table reports the estimated coefficients of equations 3.4, 3.5 and 3.6, where HFT^{NET} is the difference between HFT^D, the liquidity demanded by HFTs, and HFT^S, the liquidity supplied by HFTs. A negative HFT^{NET} implies an aggregate trading activity in the opposite direction of the crash (virtuous behavior) while a positive HFT^{NET} implies an aggregate trading activity in the direction of the crash (vicious behavior). EPM^{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. $EPM^{OC-STANDALONE}$ is a dummy that captures open-close EPMs that occur on their own, $EPM^{OC-SIMULTANEOUS}$ is a dummy that captures open-close EPMs that occur on several sample stocks within the same minute, $EPM^{OC-EXTREME-HOURS}$ is an interaction variable that captures the additional effect of open-close EPMs that occur in the first five and last five minutes of the trading day, $AbsRet$ is the absolute log return, SV is the share volume and RS is the relative spread. The regressions are run on an open-close EPM subsample whose proportion of transactions on Nasdaq is set to at least 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.

High-low extreme price movements

We finally focus on the trading activity of HFTs during high-low extreme price movements. Our central multivariate regression in this case is the following:

$$HFT_{it}^{NET} = \alpha_i + \beta_1 EPM_{it}^{HL} + \beta_2 AbsRet_{it} + \beta_3 SV_{it} + \beta_4 RS_{it} + (Lags_{kit-\sigma}\gamma_{k\sigma}) + \epsilon_{it} \qquad (3.7)$$

where HFT^{NET} is the difference between HFT^D and HFT^S, EPM_{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, $AbsRet$ is the absolute log 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 non-dummy variables are standardized at the stock level and we include stock fixed effects. Results of the different regressions are presented in Table 15.

We also focus on both standalone and simultaneous high-low EPMs as well as on extreme-hour high-low EPMs:

$$HFT_{it}^{NET} = \alpha_i + \beta_1 EPM_{it}^{HL-STA} + \beta_2 EPM_{it}^{HL-SIM} + \beta_3 AbsRet_{it} + \beta_4 SV_{it} + \beta_5 RS_{it} + (Lags_{kit-\sigma}\gamma_{k\sigma}) + \epsilon_{it} \tag{3.8}$$

$$\begin{aligned} HFT_{it}^{NET} = \alpha_i + \beta_1 EPM_{it}^{HL} + \beta_2 EH_{it} + \beta_3 EPM_{it}^{HL} * EH_{it} \\ + \beta_4 AbsRet_{it} + \beta_5 SV_{it} + \beta_6 RS_{it} + (Lags_{kit-\sigma}\gamma_{k\sigma}) + \epsilon_{it} \end{aligned} \tag{3.9}$$

We find similar results for open-close EPMs and high-low EPMs. When focusing on different subsamples (large, medium, small) instead of the full sample, we find (in unreported results) that the coefficient of the high-low EPM dummy variable is significantly negative for large stocks (-55.5087**) but highly significantly positive for small stocks (32.9600***), implying that while HFTs reduce their liquidity demand during high-low EPMs occurring on large stocks, they increase their liquidity demand during open-close EPMs occurring on small stocks and trade in the direction of the crash. Moreover, we find that the coefficient of the simultaneous open-close EPM variable is always negative when statistically significant. Again, while HFTs increase their liquidity demand during co-EPMs (Brogaard et al., 2018), i.e. when multiple EPMs occur within the same interval, due to the need for HFTs to reduce their cumulative exposure to stocks undergoing an EPM, they do not show a similar behavior regarding multiple EPMs occurring within a longer time frame (EPMs occurring within 40 intervals in our case). In the same way as for open-close EPMs, we conjecture that one minute is enough for HFTs to reset their inventory risk constraints and get their level of risk exposure in order.

Finally, regarding the extreme hour specification, we find (in unreported results) that the extreme hours dummy variable is statistically significant at the 1% confidence level with a negative coefficient (-42.02***), implying that the liquidity demand reduction of HFTs is even more important during the first five and last five minutes of the trading day. As such, the decline in

HFT liquidity demand is potentially more pronounced during periods of known market stress (opening and closing hours), similarly to MFCs and open-close EPMs.

Table 15: Multivariate regressions for Net HFT activity during high-low EPMs

Proportion of transactions on Nasdaq $\geq$ 50%			
	(3.7)	(3.8)	(3.9)
EPM^{HL}	-31.12***		
$EPM^{HL-STANDALONE}$		-36.62***	
$EPM^{HL-SIMULTANEOUS}$		-27.65***	
$EPM^{HL-EXTREME-HOURS}$			-20.79*
$AbsRet$	4.6021***	5.2831***	4.6338***
SV	-0.1726	-0.1783	0.1250
RS	-3.8133***	-3.8082***	3.7380***
$Adj.R^2$	0.0077	0.0078	0.0078
N	5,137,465	5,137,465	5,137,465

The table reports the estimated coefficients of equations 3.7, 3.8 and 3.9, where HFT^{NET} is the difference between HFT^D, the liquidity demanded by HFTs, and HFT^S, the liquidity supplied by HFTs. A negative HFT^{NET} implies an aggregate trading activity in the opposite direction of the crash (virtuous behavior) while a positive HFT^{NET} implies an aggregate trading activity in the direction of the crash (vicious behavior). EPM^{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. $EPM^{HL-STANDALONE}$ is a dummy that captures high-low EPMs that occur on their own, $EPM^{HL-SIMULTANEOUS}$ is a dummy that captures high-low EPMs that occur on several sample stocks within the same minute, $EPM^{HL-EXTREME-HOURS}$ is an interaction variable that captures the additional effect of high-low EPMs that occur in the first five and last five minutes of the trading day, $AbsRet$ is the absolute log return, SV is the share volume and RS is the relative spread. The regressions are run on a high-low EPM subsample whose proportion of transactions on Nasdaq is set to at least 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 this section is that based on multivariate regressions in which HFT^{NET} is a function of mini flash crashes, absolute log return, share volume and relative spread, HFTs present a rather virtuous behavior during mini flash crashes. However, it is important to put the emphasize on the fact that HFTs also present an ambiguous behavior during extreme price movements, reducing their liquidity demand during crashes occurring on large stocks but increasing their liquidity demand during crashes occurring on small stocks. As such, the virtuous conduct of HFTs in large stocks may hide a more vicious conduct in small stocks during EPMs. Another takeaway is that HFTs do not behave in the same way during co-crashes (within the same interval) or during repeated crashes (within a set of intervals). While Brogaard et al. (2018) highlight the fact that HFT liquidity supply is sensitive to inventory risk, we find that one minute is enough time for HFTs to reset their inventory constrains. A final important takeaway is that the reduction in HFT liquidity demand is more pronounced during extreme hours, i.e. during the first five and last five minutes of the trading day. We conjecture that the decline in HFT liquidity demand is more pronounced during periods of known market stress (opening and closing hours) as opposed to periods of unknown market stress.

3.6 Determinants of future crashes

We model the probability for a stock to undergo a mini flash crash as a function of lagged values of HFT^{NET}, absolute log return, share volume, relative spread and HFT participation based on trades and look at the determinants of future MFCs. We use the same explanatory variables as in Brogaard et al. (2018) except we use HFT participation based on trades as a supplementary explanatory variable. To do so, we use a logistic (logit) regression model (dependent variable = MFC) so as to test the relationship between the dependent variable and related potential factors and rank them by relative importance.

We then perform a similar logistic regression analysis on open-close extreme price movements (dependent variable = open-close EPM) and on high-low extreme price movements (dependent variable = high-low EPM) so as to put the different results into perspective. Again, we consider MFCs (respectively EPMs) for which the proportion of transactions occurring on Nasdaq during the interval represents at least 50% of all transactions on U.S. stock exchanges.

Determinants of future mini flash crashes

We first model the probability for a stock to undergo a mini flash crash. Specifically, we model the probability for a stock to experience a mini flash crash as a function of lagged values of HFT^{NET}, absolute log return, share volume, relative spread as well as HFT Participation based on trades:

$$Prob(MFC=1)_{it} = \alpha_i + \beta_1 {HFT^{NET}}_{it-1} + \beta_2 AbsRet_{it-1} + \beta_3 SV_{it-1} + \beta_4 RS_{it-1} + \beta_5 HFTP_{it-1} + \epsilon_{it} \tag{3.10}$$

where the dependent variable is a binary variable equal to one if the 1.5-second interval t contains a mini flash crash on stock i 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 log return, SV is the share volume, RS is the relative spread and $HFTP$ is the level of HFT participation based on trades. We estimate four specifications of the model so that we run the logit regression focusing in turn on all MFCs (All), standalone MFCs (Standalone), simultaneous MFCs (Simultaneous) as well as MFCs occurring during extreme hours exclusively, i.e. extreme-hour MFCs (Extreme hours). Results are presented in Table 16.

Due to the potential bias engendered by the imbalance between events and non events within our samples (number of MFCs/EPMs vs number of non-MFCs/non-EPMs), we implement a penalized maximum likelihood estimation, as advocated by Firth (1993).

First, we take into account all MFCs (equation 3.10) and observe that four out of the five parameters in our model are most of the time highly statistically significant. Lagged HFT^{NET}, lagged relative spread and lagged HFT participation are always statistically significant at the 1% confidence level and lagged absolute log return is either statistically significant at the 1% level or at the 10% level (when dependent variable = Extreme hour MFC). It implies that the probability for a stock to undergo a mini flash crash at time t depends on the level of HFT^{NET}, absolute log return, relative spread and HFT participation at time *t-1*. We note that results are similar whatever the model specification (MFC, standalone MFC, simultaneous MFC or extreme-hour MFC as the dependent variable) at the exception of HFT^{NET} whose coefficient, even though always very close to 0.0000 is sometimes positive and sometimes negative.

Second, we observe that absolute log return, share volume, relative spread and HFT participation all present positive coefficients. In other words, an increase in each of these variables increases the probability for a stock to undergo a mini flash crash in the next interval. Based on coefficients, HFT participation has by far the strongest impact (whatever the specification) on the probability for a stock to undergo an MFC in the next interval. We note that absolute log return and relative spread also have a relatively strong impact on this probability. The impact of HFT participation on futures MFCs is substantial, an increase by one standard deviation in HFT participation increasing the odds for a stock to suffer an MFC in the next interval by 26.9% (in the base specification), while the impact of absolute log return and relative spread on future MFCs is relatively substantial, an increase by one standard deviation in absolute log return and relative spread increasing the odds for a stock to suffer an MFC in the next interval by 4.9% and 9.6% respectively (in the base specification).

Third, the impact of lagged HFT^{NET} (when the coeeficient is positive) remains negligible (whatever the dependent variable we use). Indeed, the odds ratio (comprised between 1.000 and 1.001) associated with lagged HFT^{NET} implies that an increase by one standard deviation of HFT^{NET} increases the odds of a future MFC from 0.00% to 0.10% depending on the specification. In other words, MFCs appear to be closely related to the level of HFT participation as well as to relative spread and absolute return (even though to a lesser extent). On the contrary, MFCs do not seem to that related to the level HFT^{NET}. We thus conclude that the most important determinants of future MFCs are (in order of importance) HFT participation, relative spread and absolute return. The impact of HFT^{NET} is negligible.

Table 16: Logistic regressions - MFCs

Proportion of transactions on Nasdaq $\geq$ 50%						
		Coefficient	Odds ratio	P-value	C-stat	N
	All (Nb of MFCs=274)				0.73	4,787,047
HFT^{NET}_{it-1}		0.0004	1.000	***		
$AbsRet_{it-1}$		0.0476	1.049	***		
SV_{it-1}		0.0005	1.001			
RS_{it-1}		0.0921	1.096	***		
$HFTP_{it-1}$		0.2385	1.269	***		
	Standalone (Nb of MFCs=87)				0.70	4,787,047
HFT^{NET}_{it-1}		-0.0002	1.000	***		
$AbsRet_{it-1}$		0.0460	1.047	***		
SV_{it-1}		0.0054	1.005			
RS_{it-1}		0.0546	1.056	***		
$HFTP_{it-1}$		0.2734	1.314	***		
	Simultaneous (Nb of MFCs=93)				0.77	4,787,047
HFT^{NET}_{it-1}		0.0005	1.001	***		
$AbsRet_{it-1}$		0.0393	1.040	***		
SV_{it-1}		0.0032	1.003			
RS_{it-1}		0.1016	1.107	***		
$HFTP_{it-1}$		0.1987	1.220	***		
	Extreme hours (Nb of MFCs=79)				0.78	4,787,047
HFT^{NET}_{it-1}		0.0003	1.000	***		
$AbsRet_{it-1}$		0.0225	1.023	*		
SV_{it-1}		0.0099	1.010			
RS_{it-1}		0.1205	1.128	***		
$HFTP_{it-1}$		0.2217	1.248	***		

The table reports results of the different logistic regressions (see equation 3.10) regarding the probability for a stock to undergo a mini flash crash at time t as a function of HFT^{NET}, absolute log return (AbsRet), share volume (SV), relative spread (RS) and HFT participation based on trades (HFTP) at time *t-1.* The event window is 30 minutes, i.e. 15 minutes before the crash and 15 minutes after the crash. All non-dummy variables are standardized at the stock level. Asterisks ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels. The c-statistic is a measure of goodness of fit for binary outcomes in a logistic regression model. A c-stat over 0.7 indicates a good model. As an illustration, if we randomly choose two stocks from our sample, one that has the outcome characteristic (experienced a mini flash crash) and one that does not have the outcome characteristic (did not experience a mini flash crash), each stock has a predicted probability of experiencing a crash from the logistic regression model. The c-statistic is the probability that the stock that truly has the outcome characteristic will have a higher predicted probability from the logistic regression equation than the stock that truly does not have the characteristic.

Determinants of future open-close extreme price movements

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

$$Prob(EPM^{OC} = 1)_{it} = \alpha_i + \beta_1 HFT^{NET}{}_{it-1} + \beta_2 AbsRet_{it-1} + \beta_3 SV_{it-1} + \beta_4 RS_{it-1} + \beta_5 HFTP_{it-1} + \epsilon_{it} \tag{3.11}$$

where the dependent variable is equal to one if the 1.5-second interval t contains an open-close extreme price movement on stock i and zero otherwise. All the independent variables are lagged by one interval and all the variables are standardized at the stock level. All the independent variables are as previously described. Results are presented in Table 17.

First, we take into account all open-close EPMs (equation 3.11) and observe that five out the five parameters in our model are most of the time highly statistically significant. Lagged absolute return, lagged share volume, lagged relative spread and lagged HFT participation are always statistically significant at the 1% confidence level and lagged HFT^{NET} is highly statistically significant in three out of the four specifications. It implies that the probability for a stock to undergo an open-close extreme price movement at time t depends on the level of HFT^{NET}, absolute log return, share volume, relative spread and HFT participation at time *t-1*. We note that results are similar whatever the model specification (MFC, standalone MFC, simultaneous MFC or extreme-hour MFC as the dependent variable) at the exception of HFT^{NET} whose coefficient, even though always very close to 0.0000 is sometimes positive and sometimes negative.

Second, we observe that absolute log return, share volume, relative spread and HFT participation all present positive coefficients. In other words, an increase in each of these variables increases the probability for a stock to undergo an open-close EPM in the next interval. Based on coefficients, HFT participation, absolute return and relative spread have the strongest impact (whatever the specification) on the probability for a stock to undergo an open-close EPM in the next interval. For example, an increase by one standard deviation in HFT participation, relative spread and absolute return increases the odds for a stock to suffer an open-close EPM in the next interval by 18.35%, 14.39% and 13.10% respectively (in the base specification).

Third, the impact of lagged HFT^{NET} remains negligible (whatever the dependent variable we use). Indeed, when the coefficient associated with HFT^{NET} is positive (which is not always the case), then the odds ratio is comprised between 1.000 and 1.001, which implies that an increase by one standard deviation of HFT^{NET} increases the odds of a future open-close EPM from 0.00% to 0.10% depending on the specification. In other words, open-close EPMs appear to be closely related to the level of HFT participation, relative spread and absolute return. However the ranking of these factors varies from one specification to the other, which is not the case for MFCs where HFT participation always ranks first. Similarly to MFCs, open-close EPMs do not seem that related to the level of HFT^{NET}. We thus conclude that the most important determinants of future MFCs are (excluding any ranking) absolute return, relative spread and HFT participation. As in the case of MFCs, the impact of HFT^{NET} is negligible.

Table 17: Logistic regressions - open-close EPMs

Proportion of transactions on Nasdaq ≥ 50%						
		Coefficient	Odds ratio	P-value	C-stat	N
	All (Nb of EPMs=7,185)				0.72	4,787,047
HFT^{NET}_{it-1}		-0.0001	1.000	***		
$AbsRet_{it-1}$		0.1310	1.140	***		
SV_{it-1}		0.0125	1.013	***		
RS_{it-1}		0.1439	1.155	***		
$HFTP_{it-1}$		0.1835	1.201	***		
	Standalone (Nb of EPMs=5,603)				0.70	4,787,047
HFT^{NET}_{it-1}		0.0001	1.000	***		
$AbsRet_{it-1}$		0.1206	1.128	***		
SV_{it-1}		0.0138	1.014	***		
RS_{it-1}		0.0756	1.078	***		
$HFTP_{it-1}$		0.2350	1.265	***		
	Simultaneous (Nb of EPMs=13,652)				0.77	4,787,047
HFT^{NET}_{it-1}		-0.0001	1.000			
$AbsRet_{it-1}$		0.1897	1.209	***		
SV_{it-1}		0.0187	1.019	***		
RS_{it-1}		0.2098	1.233	***		
$HFTP_{it-1}$		0.1522	1.164	***		
	Extreme hours (Nb of EPMs=1,695)				0.78	4,787,047
HFT^{NET}_{it-1}		-0.0002	1.000	***		
$AbsRet_{it-1}$		0.0898	1.094	***		
SV_{it-1}		0.0120	1.012	***		
RS_{it-1}		0.1958	1.216	***		
$HFTP_{it-1}$		0.1764	1.193	***		

The table reports results of the different logistic regressions (see equation 3.10) regarding the probability for a stock to undergo a mini flash crash at time t as a function of HFT^{NET}, absolute log return (AbsRet), share volume (SV), relative spread (RS) and HFT participation based on trades (HFTP) at time *t-1*. The event window is 30 minutes, i.e. 15 minutes before the crash and 15 minutes after the crash. All non-dummy variables are standardized at the stock level. Asterisks ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels. The c-statistic is a measure of goodness of fit for binary outcomes in a logistic regression model. A c-stat over 0.7 indicate a good model. As an illustration, if we randomly choose two stocks from our sample, one that has the outcome characteristic (experienced an open-close extreme price movement) and one that does not have the outcome characteristic (did not experience an open-close extreme price movement), each stock has a predicted probability of experiencing a crash from the logistic regression model. The c-statistic is the probability that the stock that truly has the outcome characteristic will have a higher predicted probability from the logistic regression equation than the stock that truly does not have the characteristic.

Determinants of future high-low extreme price movements

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

$$Prob(EPM^{HL} = 1)_{it} = \alpha_i + \beta_1 HFT^{NET}{}_{it-1} + \beta_2 AbsRet_{it-1} + \beta_3 SV_{it-1} + \beta_4 RS_{it-1} + \beta_5 HFTP_{it-1} + \epsilon_{it} \tag{3.12}$$

where the dependent variable is equal to one if the 1.5-second interval t contains a high-low extreme price movement on stock i and zero otherwise. All the independent variables are lagged by one interval and all the variables are standardized at the stock level. All the independent variables are as previously described. Results are presented in Table 18.

Our findings regarding high-low EPMs are very similar to our findings vis-à-vis open-close EPMs and thus do not necessitate additional comments.

Overall, we find that HFT participation is the main determinant of mini flash crashes while the impact played by HFT participation is more ambiguous in the case of extreme price movements as it does not always rank first depending on the specification. Moreover, we find that absolute log return and relative spread also have a strong impact (whatever the specification) on the probability for a stock to undergo an MFC (respectively an EPM) in the next interval. Finally, we find that the impact of HFT^{NET} is negligible (whatever the dependent variable we use).

Table 18: Logistic regressions - high-low EPMs

Proportion of transactions on Nasdaq $\geq$ 50%

	Coefficient	Odds ratio	P-value	C-stat	N
All (Nb of EPMs=6,639)				0.78	4,787,047
HFT_{it-1}^{NET}	-0.0001	1.000	*		
$AbsRet_{it-1}$	0.1803	1.198	**		
SV_{it-1}	0.0106	1.011	***		
RS_{it-1}	0.1549	1.168	***		
$HFTP_{it-1}$	0.1778	1.195	***		
Standalone (Nb of EPMs=5,534)				0.75	4,787,047
HFT_{it-1}^{NET}	0.0000	1.000			
$AbsRet_{it-1}$	0.1653	1.180	***		
SV_{it-1}	0.0137	1.014	***		
RS_{it-1}	0.0647	1.067	***		
$HFTP_{it-1}$	0.2419	1.274	***		
Simultaneous (Nb of EPMs=13,978)				0.834	4,787,047
HFT_{it-1}^{NET}	0.0000	1.000			
$AbsRet_{it-1}$	0.3008	1.351	***		
SV_{it-1}	0.0135	1.014	***		
RS_{it-1}	0.2244	1.252	***		
$HFTP_{it-1}$	0.1000	1.105	***		
Extreme hours (Nb of EPMs=1,949)				0.88	4,787,047
HFT_{it-1}^{NET}	-0.0001	1.000	***		
$AbsRet_{it-1}$	0.1172	1.124	***		
SV_{it-1}	0.0140	1.014	***		
RS_{it-1}	0.2111	1.235	***		
$HFTP_{it-1}$	0.1593	1.173	***		

The table reports results of the different logistic regressions (see equation 3.10) regarding the probability for a stock to undergo a mini flash crash at time t as a function of HFT^{NET}, absolute log return (AbsRet), share volume (SV), relative spread (RS) and HFT participation based on trades (HFTP) at time *t-1*. The event window is 30 minutes, i.e. 15 minutes before the crash and 15 minutes after the crash. All non-dummy variables are standardized at the stock level. Asterisks ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels. The c-statistic is a measure of goodness of fit for binary outcomes in a logistic regression model. A c-stat over 0.7 indicates a good model. As an illustration, if we randomly choose two stocks from our sample, one that has the outcome characteristic (experienced a high-low extreme price movement) and one that does not have the outcome characteristic (did not experience a high-low extreme price movement), each stock has a predicted probability of experiencing a crash from the logistic regression model. The c-statistic is the probability that the stock that truly has the outcome characteristic will have a higher predicted probability from the logistic regression equation than the stock that truly does not have the characteristic.

4 Robustness checks

We perform robustness checks at several 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 regarding (up to) 1.5-second crashes. We note that even though mini flash crashes are somewhat different from extreme price movements, our findings on mini flash crashes are most of the time corroborated by similar findings related to extreme price movements and vice versa.

Second, we consider MFCs (respectively EPMs) for which the proportion of transactions occurring on Nasdaq during the interval represents at least 50% of all transactions on U.S. stock exchanges. As a robustness check however, we also run the regressions on (1) the full MFC sample (respectively EPM samples), i.e. MFCs (EPMs) for which the proportion of transactions on Nasdaq is comprised between 0% and 100%, thus taking into account all MFCs (EPMs), including MFCs (EPMs) where no transaction is observed on Nasdaq, and on (2) an MFC subsample (respectively EPM subsamples) in which the proportion of transactions on Nasdaq is equal to 100%, thus taking into account MFCs (EPMs) where all transactions during the crash occur on Nasdaq exclusively.

Third, we use an event window of 30 minutes (i.e. 15 minutes before the crash and 15 minutes after the crash) in our base case methodology. We also test the robustness of our results by running our regressions on 15-minute and 60-minute windows. We find similar results whatever the window size we use.

Fourth, 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 for mini flash crashes and extreme price movements, we also perform a similar analysis using alternative time intervals of 1 second and 2 seconds respectively. Whatever the time interval we use, our results remain very similar.[15]

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

5 Conclusion

In this paper, we identify mini flash crashes by replicating Nanex MFC detection algorithm (2010) and we complement our study with a parallel analysis on extreme price movements, providing an alternative way to measure them. To the best of our knowledge, we are the first to run a parallel analysis on both mini flash crashes and extreme price movements and look at mini flash crashes on large, medium and small stocks (previous studies focus exclusively on large stocks).

We study how high-frequency traders behave around and during mini flash crashes (respectively extreme price movements), which we define as crashes that are sudden, extreme, characterized by very short-lived abrupt price changes that exhibit at least ten tick movements in the same direction before ticking in the other direction and that partially or totally self-correct within a few seconds (respectively crashes that belong to the 99.9th percentile of the absolute log return distribution). We identify 510 MFCs, 29,424 open-close EPMs and 29,427 high-low EPMs over a two-year period (2008-2010), representing about one mini flash crash and fifty-eight extreme price movements per day on average based on 74 large, medium and small U.S. equities traded on the Nasdaq stock exchange.

Overall, we find that computing extreme price movements from open to close or from high to low (respectively low to high) does not fundamentally change the nature of our results. We thus conclude that the methodology used by Brogaard et al. (2018) may not be that misleading. Still, we offer an alternative methodology, which we believe, might be useful in the identification of EPMs at higher frequencies.

Following Brogaard et al. (2018), we compute directional trade imbalance metrics for the pre-crash, crash and post-crash phases so as to see who (HFTs or NHFTs) triggers the crash, exacerbates the crash or leads the price recovery after the crash. To the question 'Do HFTs trigger mini flash crashes ?' and based on directional trade imbalance metrics computed right before the crash, the answer is uncertain. We do not find any evidence that HFTs trigger mini flash crashes or extreme price movements. To the question 'Do HFTs exacerbate the crash ?' the answer is mixed. While no conclusion can be reached about MFCs, due to a lack of statistical significance, it appears that HFTs do (highly statistically significantly) exacerbate the crash during 1.5-second extreme price movements (in line with the findings by Bellia et al., 2018). This contradicts the findings by Brogaard et al. (2018) and may be explained by the fact we apply a filter to remove consecutive crashes that pollute the different windows around the crash. Finally, to the question 'Do HFTs lead the price recovery right after the crash?' and based on the directional trade imbalance metrics computed right after the crash, we do not find any evidence regarding MFCs, agin due to a lack of statistical significance, but find that NHFTs are the ones that contribute to the resiliency of stock prices after the crash (at a 1% significance level) in the case of extreme price movements thus driving the price recovery right after crash. On the contrary, HFTs keep demanding liquidity in the direction of the crash during the post-crash phase. Based on directional trade imbalance metrics, some results indicate that HFTs may have a vicious conduct during 1.5-second crashes.

Moreover, we find via a multivariate regression analysis similar to Brogaard et al. (2018) that HFTs present a rather ambiguous behavior during 1.5-second crashes. Indeed, when studying the full MFC and EPM samples, we do no find any evidence regarding MFCs due to a lack of statistical significance but we find that on average HFTs reduce their liquidity demand during extreme price movements on an aggregate basis, which appears to be a virtuous behaviour. However, when studying EPM subsamples by market capitalization, we find that HFTs do reduce their liquidity demand during EPMs occurring on large stocks but increase their liquidity demand during EPMs occurring on small stocks and trade in the direction of the crash. In other words, the virtuous conduct of HFTs in large stocks may hide a more vicious conduct in small stocks during EPMs.

We also find that HFTs do not behave in the same way during co-crashes (within the same interval) as during repeated crashes (within a set of intervals).[16] While Brogaard et al. (2018) highlight the fact that HFT liquidity supply is sensitive to inventory risk, we find that one minute is long enough for HFTs to reset their inventory constrains. In addition, we find that the reduction in HFT liquidity demand is more pronounced during extreme hours, i.e. during the first five and last five minutes of the trading day, that during the rest of the day. We conjecture that the decline in HFT liquidity demand is more pronounced during periods of known (anticipated) market stress (opening and closing hours) as opposed to periods of unknown (unanticipated) market stress.

Finally, we model the probability for a stock to undergo a mini flash crash (respectively an extreme price movement) as a function of lagged values of HFT^{NET}, absolute log return, share volume, relative spread and HFT participation based on trades and perform a logistic regression analysis so as to test the relationship between the dependent (binary) variable and related potential factors and rank them by relative importance. We find that HFT participation at time *t-1* is by far the main determinant of mini flash crashes at time *t*, whatever the model specification. Contrary to MFCs, we do not find any dominant determinant of the crash to come in the case of extreme price movements. HFT participation at time *t-1* still appears as one of

[16] While Brogaard et al. (2018) focus on crashes occurring within the same 10-second interval (co-crashes), we focus on crashes occurring within the same 60-second interval (simultaneous crashes).

the main determinants of crashes at time t along with absolute return and relative spread but it does not always rank first depending on the specification.

In terms of limits, our findings regarding the behavior of HFTs around and during mini flash crashes and extreme price movements only relate to the Nasdaq stock exchange. As such, it is possible that the behavior of HFTs during MFCs and EPMs may differ from one exchange to another. One avenue for further research might be to study the behavior of HFTs on all the U.S. stock exchanges at the same time. Moreover, it is also possible that the behavior of some market participants might be hidden within the aggregated data at our disposal as pointed out by Bellia et al. (2018). Another avenue for further research might thus be to perform a similar study on a more refined dataset where the trading activity of individual HFTs would be flagged. Finally, a final avenue for further research might be to focus on higher frequencies so as to observe changes in both trading and liquidity dynamics more closely.

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.

Firth, D. (1993), 'Bias reduction of maximum likelihood estimates', *Biometrika* **80**(1), 27–38.

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.

Laly, F. & Petitjean, M. (2019), 'Mini flash crashes: Review, taxonomy and policy responses', *Bulletin of Economic Research* (forthcoming).

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. Z. (2018), 'The flash crash: A cautionary tale about highly fragmented markets', *Management Science* .

Nanex (2010), 'Flash equity failures in 2006, 2007, 2008, 2009, 2010, and 2011, technical report'.

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 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 Studies* **43**(6), 767–797.