Flash Crashes: Review of the identification techniques

Octavio Cardona García
University of Konstanz

Introduction

High frequency trading is becoming more popular in the financial markets. The speed of the transactions is fundamental for some companies to make profits in the market and take advantage from arbitrage. However, faster transactions, beyond human reactions, also implies risks, one of those risks is called Flash Crash, abrupt changes in prices with a subsequent reversal in matter of minutes or even seconds.

The literature on Flash Crashes is not abundant. The first observed episode ocurred on 2010 and since then there have been few attempts to establish the causes of these market drawdowns. Yet one kind of Flash Crashes called Mini Flash Crashes (also known as Ultra Fast Extreme Events), a smaller version of the crashes, have called the attention of some researchers. These literature is basically concentrated on identification and description of these events.

With respect to identification, we observe Flash Crashes without any kind of statistical test, while Mini Flash Crashes require the analysis of high frequency data. The most often used method is an ad hoc approach developed by Nanex consisting in a list of characteristics that a trade should fulfill to classify as a Mini Flash Crash. Besides, a relative recent paper by Christensen, Oomen, and Renn (2016) also includes an attempt to test them formally. From the statistical learning theory, there is no particular study dedicated to Flash Crashes. In spite of this, there are attempts to identify rare events and outliers, this approach might be an useful technique as well.

The objective of this document is to describe the existing methodologies and discuss the statistical learning approach as a possible new way of identifying these events. Also, in order to compare the ad hoc and outlier identification approaches, there is an empirical part with a high frequency data set.

The organization of the document is as follows: First, introduction; second, brief definition of Flash and Mini Flash Crashes; third, identification techniques; fourth, empirical analysis; fifth, conclusions.

Flash and Mini Flash Crashes

On May 6, 2010 the financial markets started with an unusual pressure coming from the european debt crisis, specifically the Greek's sovereign debt. The negative feeling about the economic news started with a sharp decline of Euro against the U.S Dollar and the Japanese Yen. Around 1 p.m the number of volatility pauses (known as Liquidity Replacement Points "LRPs") increased in the New York Stock Exchange "NYSE".

One and a half hours later, the market volatility index "VIX" in the S&P 500 had increased up to 22.5% compared to the starting level that day. The most remarkable news was the behavior of the E-Mini S&P 500 futures which had fallen by 20%.

According to the U.S. Commodity Futures Trading Commission / U.S. Securities and Exchange Commission (2010) at 2:32 p.m. a fundamental trader had initiated a sell program of 75.000 E-Mini contracts even when the market showed clear signals of insufficient liquidity and high volatility. This sell program was initiated by a automated execution algorithm or "Sell Algorithm", resulting in the largest net change of any trader in the E-Mini that day. Furthermore, the process was executed extremely fast, approximately 20 minutes. The event would be known as the "Flash Crash".

During this episode, many stocks in the S&P 500 and in the Dow Jones Index lost on average between 5% to 15% of their price to recover the previous levels just few minutes later. The causes and consequences have been studied since then by some authors, nevertheless there is no consensus about it and the literature produced is limited so far.

Although the Flash Crash of the 6 May, 2010 has been the most studied and controversial case, there have been many more flash crashes. Wells and Chemi (2016) of CNBC list some of them: Facebook debut on May 18, 2012; Flash freeze on August 22, 2013; Treasury freeze on October 15, 2014; Lunch-hour halt the NYSE on July 9, 2015; and, Four-digit drop: August 24, 2015. In addition, similar behaviors have been observed in the foreign exchange market as for example the Flash Crash of the British Pound on October 6, 2019 and its Japanese counterpart on January 2, 2019.

Flash Crashes in financial markets are observed more often than bubbles or "normal" crashes, nonetheless they are still relative rare and therefore the specialized literature is commonly interested in a smaller version called Mini Flash Crashes or, to be more technical, Ultra-fast Extreme Event "UEE".

Mini Flash Crashes are characterized for being even faster, some of them appear in seconds or milliseconds, making impossible any kind of human intervention unlike Flash Crashes. On the other hand, UEE do not have the same magnitude and some authors claim they occur quite often. Johnson et al. (2013) have identified 18520 Mini Flash Crashes from January 3, 2006 to February 3, 2011 by using a methodology (that will be explained in the next section) proposed by Nanex, a firm that provides streaming of financial market data.

There are no formal definitions of Flash and Mini Flash Crashes, they are tacitly identified by the sudden change in prices in a very short period of time. There exists some attempts to model their behavior though, by considering the drift burst hypothesis (Christensen et al., 2016) and the Poisson Hawkes model (Filimonov & Sornette, 2012). These approaches contribute to the debate, but cannot escape from it, any formal model is dependent on the set of assumption we make, and as it was mentioned before, we do not know the causes of these events.

The next section will focus on those documents that try to identify Flash and Mini Flash Crashes such as Johnson et al. (2013) and Christensen et al. (2016) leaving out some other theoretical points of view about this topic. Even though, it is worth to mention some hypothesis regarding the market behavior during these crashes.

One of the hypothesis about the Flash Crash of May, 2010 was proposed by Easley, López de Prado, and O'Hara (2010) who claim that order flow toxicity plays an important role during the Flash Crashes by affecting the liquidity provision and propose a statistical measure of this called Volume-Synchronized Probability of Informed Trading "VPIN". First of all we need to know that "order flow toxicity is the measure of a trader's exposure to the risk that counterparts possess private information or other informational advantages." (Rand Low, Te Li, & Terry Marsh, 2018, p. 1). In our context, flow toxicity happens when High Frequency Trading "HFT" firms face an adverse selection risk scenario in which other participants have different information generating a imbalance in the microstructure of the market. In other words, HFT firms face a problem of liquidity when other actors close their positions causing large losses and forcing them to undo their positions.

Following the U.S. report, some authors came up with the hypothesis that Intermarket Sweep Orders "ISO" triggered the mechanism that makes the algorithms overreact. For instance, McInish, Upson, and Wood (2014) examine the ISOs before and after May 6, finding that there is a significant increase of them during short periods of time the day of the crash. In the same fashion, Golub, Keane, and Poon

(2012) analyzes Mini Flash Crashes in the most volatile months during 2006-2011 attributing the cause to regulatory framework and the inherent problem of fleeting liquidity.

With respect to Mini Flash Crashes, they are supposed to be affected by the same factors: human mistakes, endogenous feedback loops, liquidity problems, value shocks, and market fragmentation. (Bayraktar, 2017). Endogenous feedback loops are treated in the Filimonov and Sornette (2012) work. The model consists in a self excited conditional Poisson-Hawkes model which includes the influence of exogenous shocks. The Hawkes model is a generalization where the frequency of the events are dependent not just on time but also on the history of the process. Additionally Bayraktar (2017) has built a model based on the optimal execution literature, unlike the traditional equilibrium models. The idea is to express mathematically all the mentioned causes for the Mini Flash Crashes by building an individual optimization problem in which all the agents have wrong beliefs.

Lastly, there exist more models based on different hypothesis going from an econophysics perspective (Fry & Serbera, 2017; Mazzeu, Otuki, & Da Silva, 2011) whose working papers do not categorize the Flash Crash of May 6, 2010 as an anomaly, but instead as a power-law distribution that could be modeled by a GARCH(1,1), to a micro-macro agent-based model (Paulin, Calinescu, & Wooldridge, 2019) that describes the flash crash contagion with overlapping portfolio model that predicts the systematic risk of having a diversified or a overcrowded portfolio.

Identification Techniques

Most of the time, identification of Flash Crashes does not require any different technique apart from observing the episode. As mentioned before, these are characterized by an abrupt movement in the price of the asset to recover the previous level just minutes later. In contrast, UEE are more ambiguous. How can we identify them? They are not easily observable since they happen in minuscule time frames with smaller movements in prices. As a result, in the financial econometrics literature exist two ways of identifying them. The first one is proposed by Nanex as a set of rules that a trade should fulfill. The second approach is a formal test similar to a t-student statistic developed by Christensen et al. (2016). Moreover, the idea of "interesting outliers" form the statistical learning literature might help to identify UEE, specifically with the algorithm proposed by Lecué and Lerasle (2017).

The strategy used by Nanex [1] on November 2010 is to capture those stocks with a negative (positive) change in price of more than 0.8% with at least ten continuous ticks going down (up) before going up (down) again in the next tick, all within 1.5 seconds.

Different researches have used this strategy to identify Mini Flash Crashes, modelling them, and describing their consequences. (Braun, Fiegen, Wagner, Krause, & Guhr, 2018; Golub et al., 2012; Johnson et al., 2013; Levine, Hale, & Floridi, 2017). In all of these documents the use of basic statistics is common, there are no econometric models to predict future events or characterize the observed structure.

For example, in Braun et al. (2018) the authors have considered all the stocks that were continuously traded during 2007 and 2008 in the S&P 500, since during this time many UEE were reported in the previous work of Johnson et al. (2013). Among the results, they have found 5529 Mini Flash Crashes, the sector with more occurrences is the financial sector with 33.35 events per company, the energy sector and telecommunications were behind with 14.36 and 12.14 occurrences per company respectively. They also conclude that stocks with lower liquidity are more likely to

[1] Besides of providing this ad hoc method, the company has identified different events from 2006 to 2011: `http://www.nanex.net/FlashCrashEquities/FlashCrashAnalysis_Equities.html`

exhibit abrupt changes in price, along with this, they have observed that Mini Flash Spikes (Flash Crashes with positive changes in price) have 7% more probability to appear than Mini Flash Crashes.

In the same direction Golub et al. (2012) have analyzed the US equity markets during the most volatile months during 2006 to 2011 by using the same strategy. They have found that Mini Flash Crashes are the result of regulatory deficiencies and market fragmentation. Basically they differentiate between ISO-initiated and auto-routine initiated Flash Crashes, concluding that the aggressive behavior of the former make the ISO's a particular generator of UEE with 67.85% events out of the 5140 identified. Another conclusion, in line with the findings of the prior mentioned authors is that UEE are more likely during low liquidity episodes.

The work of Levine et al. (2017) has a slightly different approach. The purpose is to use Mini Flash Crashes as predictors of the Flash Crashes. To do so, they have counted the number of UEE during the first three minutes after the opening of the market (9:30 -9:33) just before the start of the Flash Crash. Although the idea in this document is particularly interesting, most of the analysis consists in descriptive statistics comparing the number of events on October 15, relative to prior and posterior days in the same time frame. The conclusion coincides with their hypothesis of many more UEE during the day of the Flash Crash.

On the contrary, Christensen et al. (2016) have designed a formal test to identify drift burst from noisy high frequency data. They model a continuous- time Itô semimartingale process for the log-price X_t:

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

In which the parameter of interest is the drift μ_t. σ_t is the volatility, W_t represents a Brownian motion, and J_t is a jump process. The idea of the document is to build a theoretical framework and test for locally explosive trends in prices. With respect to the test the authors have build the following non-parametric statistic:

$$T_t^n = \sqrt{\frac{h_n}{K_2}} \frac{\hat{\mu}_t^n}{\hat{\sigma}_t^n} \tag{2}$$

Where h_n is the bandwidth of the mean estimator, K_2 is a kernel-dependent constant, and $\hat{\mu}_t^n$, $\hat{\sigma}_t^n$ are non parametric estimators of the drift and the volatility respectively. This test have power under the alternative hypothesis of fast drift explosion since under the null hypothesis of no drift burst the test is asymptotically normal but it diverges.

As empirical result, the authors have analyzed the most actively traded funds in the Chicago Mercantile Exchange from January 2012 to December 2017, identifying over a thousand significant drift burst episodes of which two thirds were followed by a price reversion as it is expected from the Mini Flash Crashes. Moreover they conclude that the volume during these episodes is highly correlated with price reversals.

One issue with this methodology is that the researcher must select the bandwidths arbitrarily, although the common procedure according to the authors is taking five minutes. This is not a minor issue since the importance of an optimal bandwidth will give us the best variance-bias trade off and therefore it has implication over the results of the test.

From the Statistical Learning literature there is not a direct attempt to measure this phenomena in financial markets. Obviously there exists attempts to predict crashes by using Recurrent Neural Networks and Long Short Term Memory Networks, however, they are not explicitly designed to identify UEE or normal Flash Crashes.

An approach that might be applicable is the interpretation of Flash and Mini

Flash Crashes as "interesting outliers". This notion is given in Lecué and Lerasle (2017), although the original idea of the researches is a robust estimation in a data set contaminated by outliers, their algorithm includes an identification procedure as a by product. As a remark, the use of a similar notion is not new, the work of Johansen and Sornette (2000) is a good example of the market price drawdowns interpreted as outliers.

The basic idea of Lecué and Lerasle (2017) consists in a median of mean estimator "MOM" as a robust alternative to traditional machine learning algorithms. From all their work the $\mathcal{O} \cup \mathcal{I}$ framework is useful in our context due to it gives a formal expression for rare events by making flexible the i.i.d assumption. They assume a data process divided into two groups, that are unknown to the statistician, $1, ..., N = \mathcal{O} \cup \mathcal{I}$ with $\mathcal{O} \cap \mathcal{I} = \emptyset$. We do not make any kind of assumption over data $(X_i, Y_i)_i \in \mathcal{O}$, where $\mathcal{O}$ stands for outliers. In contrast, $\mathcal{O} \cup \mathcal{I}$, where $\mathcal{I}$ stands for informative data, is assumed to be independent.

Now the challenge is, how can we identify those outliers? If we assume that after a cleaning process the data consists just in informative data and interesting outliers, the simple procedure proposed in the document is to identify a group of data free of outliers, the so called "median block": Given (f_t, g_t) and $l_f(x, y) = (y - f(x))^2$, one can find a median block B_{med} such that

$$R_{B_{med}}[l_{f_t} - l_{g_t}] = median(R_{B_k}[l_{f_t} - l_{g_t}], k \in \{1, ...K\}) \tag{3}$$

Where, K is the total number blocks in which the data is partitioned and R_B are risk functions.

This estimator is intended to be used in a multivariable settings, but in our case a simpler univariate version would be enough:

$$\hat{\mu}_n = median(Z_1, ..., Z_K) \tag{4}$$

Where, $Z_j = \frac{1}{|B_j|} \sum_{i \in j} X_i$

With this estimator the only thing we know is that an outlier is less likely to be included into a median block, but it does not mean that those non-median blocks contain an outlier within them. Therefore, the second step suggested by Lecué and Lerasle (2017) is to iterate the estimator over randomly shuffled blocks and score how many times each observation was into a median block, this score will give us an empirical measure of the centrality of the data. In other words, one observation with null score means that is an outlier.

One advantage of this estimators (under i.i.d assumption) is that they have a sub-gaussian performance for all distributions with a finite variance as one can see in the survey document by Lugosi and Mendelson (2019) :

Theorem 1. *Let $X_i, ..., X_n$ be an i.i.d random variables with mean μ and variance σ^2. Let m, k be positive integers and assume that $N = mK$. Then the MOM estimator $\hat{\mu}_n$ with K blocks satisfies*

$$\mathbb{P}\left\{ |\hat{\mu}_n - \mu| > \sigma\sqrt{4/m} \right\} \leq e^{-K/8} \tag{5}$$

Particularly, for any $\delta \in (0, 1)$ if $K = [8log(1/\delta)]$, then, with probability at least $1-\delta$,

$$|\hat{\mu}_n - \mu| \leq \sigma\sqrt{\frac{32log(1/\delta)}{N}} \tag{6}$$

Proof. One can prove the Theorem 1 with the help the Chebyshev's inequality that claims that for each block $j = 1, ..., K$ with probability 3/4 ,

$$|Z_j - \mu| \leq \sigma\sqrt{\frac{4}{m}}$$

As a result, $|\hat{\mu}_n - \mu| > \sigma\sqrt{4/m}$, implying that at least $K/2$ of the means Z_j are such that $|Z_j - \mu| > \sigma\sqrt{4/m}$. Hence,

$$\begin{aligned}
\mathbb{P}\Big\{|\hat{\mu}_n - \mu| > \sigma\sqrt{4/m}\Big\} &\leq \mathbb{P}\Big\{Bin(K, 1/4) \geq \frac{K}{2}\Big\} \\
&= \mathbb{P}\Big\{Bin(K, 1/4) - \mathbb{E}Bin(K, 1/4) \geq \frac{K}{4}\Big\} \\
&\leq e^{K/8}(by\ Hoeffding's\ inequality)
\end{aligned}$$

□

To conclude this section, I want to mention a couple of observations about the previous methodology. First, the optimal number of iterations is not described by the authors. In their example, they iterate five thousand times their simulated data, that is made of two hundred observations, but they do not provide a reason for this number; second, how can we choose an optimal number of blocks? For the multivariate case, the researchers have provided an adaptive way of selecting K, which empirically can be done with V-Fold Cross Validation. On the contrary, in the univariate case one might notice in the Theorem 1 that the optimal number of blocks K are dependent on the confidence level δ.

Empirical Analysis

The data set used in this analysis corresponds to trade data from Apple (AAPL) during the whole month of May, 2010. This set was obtained from the TAQ database at Wharton Research Data Services (WRDS) and it contains tick data with a precision of milliseconds. As a general rule, just those trades between 9:30 - 16:00 are included. The data set was cleaned following the routine proposed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009) and the algorithms to run the analysis are based in the Lecué and Lerasle (2017) and Levine et al. (2017) Github repositories.

The idea is to check the results produced by the Nanex ad-hoc procedure and compare it with the "interesting outlier" search under the mentioned $\mathcal{O} \cup \mathcal{I}$ framework. In addition, to check the robustness of the ad-hoc routine I report results with slightly different parameters in the table 1. The appendix contains all the events identified by both techniques.

	Original	Ticks = 5	Price change = 0.4%
Number of crashes	2	30	4
Avg. Return (Crash)	-1.21%	-1.68%	-0.86%
Avg. Return (Spikes)	-	1.95%	-
Min. Return	-1.57%	-4.29%	-1.57%
Max. Return	-0.84%	3.87%	-0.45%
Avg. Duration in sec.	0.13	0.18	0.26
Avg. Number of ticks	12	7	13
Min. Number of ticks	11	6	11
Max. Number of ticks	14	14	15

Table 1
Summary statistics: Nanex ad-hoc procedure

Table 1 shows the summary of the identified events following the Nanex routine. The first column presents the original version of the routine, with this specification I have found two crashes, none of them spikes. Both have occurred on May 6 2010, specifically at 14:45:51 and 14:46:22. The return, which is calculated as usual, $(P_{t_f} -$

$P_{t_i})/P_{t_i}$ where i and f stand for the beginning and the end of the crash respectively, was in average -1.21%. Both episodes last in average 0.13 seconds.

In the second and third column, I have changed one of the three parameters to contrast the results and see the sensibility of the routine. In the second column I have modified the minimum number of ticks from 10 to 5 keeping the rest constant. In the same fashion, in the third column the minimum price change was lowered from 0.8% to 0.4%. It is worth to mention, that a shift in the maximum time span from 1.5 to 5 seconds does not change the results of the original version, therefore it was omitted. I did not consider time spans beyond 5 seconds to avoid distortions in prices due to human interaction.

As we can see, in the second column the change in the number of ticks heavily affects the number of identified events. From these thirty Mini Flash Crashes, twenty nine occurred on May 6, one on May 21 at 9:30, thirteen were positive (spikes) and seventeen negative. The average return of the spikes exceeds that of the crashes in absolute value and the average duration in seconds is higher than the original case. The minimum return is a crash of -4.29% that started exactly on May 6 at 14:46:32.990 until 14:46:33.110 and 6 ticks. On the opposite side the maximum return is a spike of 3.87% starting the same day at 14:46:44.847 to 14:46:44.930 with 6 ticks as well.

In the last column just two more events were identified. Both of them lie in between of those founded by the original routine. The average duration in seconds again shows a larger increase even when the number of ticks basically remains the same.

On the other hand, table 2 contains the results of our "interesting outlier" identification. The whole data set was divided into no overlapping time spans of 30 seconds, the number of blocks K is 3 and 100 iterations per span. I use (4) to estimate the median block in each iteration. This specification came basically after an iterative process starting from taking all the observation during the May 6 Flash Crash to the inclusion of the whole sample, and in all cases with use of the optimal number of blocks suggested by Theorem 1 at $\delta = 0.05$. Unfortunately, none of these procedures worked for different reasons: First, when we take the whole day (or the whole month) as our X_i, the huge number of observations implies that we need to iterate many more times to give each data point the chance of being included in a median block; second, the number of outliers are low compared to the number of informative data, increasing the probability of incorrectly classify the outlier into a median block if the outlier is not large enough to deviate the mean of the block from its central tendency; third, if we go to the other extreme, dividing the data into time windows of less than 10 or 20 seconds (i.e. X_i contains all the ticks in a window of 10 or 20 seconds) some of the blocks could be empty if $N < K$, making impossible the estimation.

In summary, just the selected specification allowed to identify the outliers without problems or by making unnecessary assumptions over the distribution of the data. The time span and the number of blocks were selected taking into account the third problem I have just mentioned, because sometimes spans just include two or even one observation. These cases also happened a couple of times in the fifth and sixth day of the month in our data set, for them the algorithm automatically takes each data point as a median block. In addition, the number of iterations should respond to the cardinality of X_i , which is sometimes less than the average number of ticks per second in this data set (four), therefore 100 iterations should be enough for most of the time spans. Nevertheless, the optimal number of iterations is another weak point with this technique.

With the described procedure I got thirty nine events, twenty were crashes and nineteen spikes. The crashes(spikes) were defined as those trades with a price above(bellow) the average price in their thirty seconds windows. Conversely, their

Number of Flash Crashes	20
Number of Flash Spikes	19
Number of Reversions	31
Avg. Return (Crash)	-0.008%
Avg. Return (Spike)	0.008%
Min. Return	-0.031%
Max. Return	0.035%
Number of ISO initiated	5

Table 2
Summary statistics: Outliers

returns are calculated based in the immediately previous trade due to we do not a have a notion of when the crash has started. Unlike the Nanex routine, there are no outliers during May 6, they are mainly located on the 3rd and 13th days, but there are also observations on the 14th, 17th, 18th, 24th, 26th, and 28th.

Apart from the previous facts, thirty one outliers fulfill in the next trade the reversal condition that characterize the crashes. Just five of them were marked as a ISO initiated and the average returns are not particularly large as we might expect from a crash, but this is just because is the return with respect to the last trade.

The problem with the outlier identification is that we probably get just the peak of the flash crash, assuming that we have identified a real one, but we do not have any notion about the start or duration of the event and therefore it is difficult to estimate the returns. In this case I used the mean price of the last 30 seconds before the crash, just because the identification includes this time span, but this is an arbitrary number (although it is conditioned by the data as well since we do not observe trades every second). This theory could work better if we had an identification technique for a group of outliers instead of punctual points. So far, the development of the theory allows us to obtain robust estimators or even clean the data, but as an algorithm for rare events it is not completely accurate.

Comparing the two techniques, it seems that the Nanex routine is the best option, it provides more information about the beginning and the end of the Flash Crashes but it lacks a theory to back it up, as a result, one can arbitrarily induce small changes in the parameters to obtain better outputs. On the contrary, the $\mathcal{O} \cup \mathcal{I}$ framework is quite interesting if we want to relax the assumption about the distribution of the variables and probably it is a good first attempt to identify rare events, but in this specific case, the theory also depends on arbitrary modifications that surely bias the results.

Conclusion

The aim of this paper was to review and contrast the ad hoc identification procedure proposed by Nanex, against the algorithm for outlier identification proposed in Lecué and Lerasle (2017). Economic modelling and theoretical discussion about causes of these events were left aside as they go beyond the scope of the document.

Since Flash and Mini Flash Crashes are a relative new phenomena in financial markets, they represent a risk for investor and market makers when they do not have the change to react to these abrupt changes in prices, it is important to develop a reliable way of identify them.

The two methodologies compared here have two different approaches: While the ad hoc procedure bases its prediction in the previous expertise of their creators, the outlier identification establishes a theoretical framework. By contrasts, just looking at the different results, I would say that the ad hoc routine delivers better output that makes more sense, even with its particular sensitivity to the minimum number of ticks. An interesting observation about it, is the non coincidence of any identified event between the two procedures. In other words, one might expect both methods to find at least one event during the day of the Flash Crash as it was suggested in the previous literature, but the "interesting outlier" algorithm fails to do so.

Even though the ad hoc procedure makes more sense in the context of this data set, its biggest flaw comes from the fact that it is not supported by a theory. Also, if we look at the change in the results once we change the number of ticks, it is clear that the identification should not be dependent on that parameter, we do not have any reason to take ten consecutive trades because the distance between them vary along the whole set, unlike lower frequency that are equally spaced points in time.

By comparison, Mini Flash Crashes were hard to identify as outliers. Some problems I faced were the following: The optimal number of blocks depends on the confidence level δ and sometimes it makes the number of blocks higher than the number of observations in many of the defined time spans; the theory is too general and does not provide a tool for further analysis of the outliers; the algorithm fails to identify outliers when the number of observations increases; the results are not consistent with the outputs presented in previous researches with larger data sets.

Finally, I want to highlight that both techniques are in a developing process yet. The ad hoc procedure is a good start to find Mini Flash Crashes and probably the fleeting liquidity or the market manipulation hypotheses will provide a theoretical background in future researches. At the same time, the $\mathcal{O} \cup \mathcal{I}$ framework is not intended to provide precise information about the outliers, but it is an interesting start for rare events classification.

References

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2009). Realized kernels in practice: trades and quotes. *The Econometrics Journal*, *12*(3), C1-C32. doi: 10.1111/j.1368-423X.2008.00275.x

Bayraktar, E. (2017). Mini-flash crashes, model risk, and optimal execution. *SSRN Electronic Journal*. doi: 10.2139/ssrn.2975769

Braun, T., Fiegen, J. A., Wagner, D. C., Krause, S. M., & Guhr, T. (2018). Impact and recovery process of mini flash crashes: An empirical study. *PloS one*, *13*(5), e0196920. doi: 10.1371/journal.pone.0196920

Christensen, K., Oomen, R. C. A., & Renn, R. (2016). The drift burst hypothesis. *SSRN Electronic Journal*. doi: 10.2139/ssrn.2842535

Easley, D., López de Prado, M., & O'Hara, M. (2010). The microstructure of the 'flash crash': Flow toxicity, liquidity crashes and the probability of informed trading. *SSRN Electronic Journal*. doi: 10.2139/ssrn.1695041

Filimonov, V., & Sornette, D. (2012). Quantifying reflexivity in financial markets: Towards a prediction of flash crashes. *SSRN Electronic Journal*. doi: 10.2139/ssrn.1998832

Fry, J., & Serbera, J.-P. (2017). *Modelling and mitigation of flash crashes.* Germany.

Golub, A., Keane, J., & Poon, S.-H. (2012). High frequency trading and mini flash crashes. *SSRN Electronic Journal*. doi: 10.2139/ssrn.2182097

Johansen, A., & Sornette, D. (2000). Large stock market price drawdowns are outliers. *SSRN Electronic Journal*. doi: 10.2139/ssrn.244563

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. doi: 10.1038/srep02627

Lecué, G., & Lerasle, M. (2017). Robust machine learning by median-of-means : theory and practice. *Annals of Statistics*. Retrieved from `http://arxiv.org/pdf/1711.10306v2`

Levine, Z. S., Hale, S. A., & Floridi, L. (2017). The october 2014 united states treasury bond flash crash and the contributory effect of mini flash crashes. *PloS one*, *12*(11), e0186688. doi: 10.1371/journal.pone.0186688

Lugosi, G., & Mendelson, S. (2019). Mean estimation and regression under heavy-tailed distributions: A survey. *Foundations of Computational Mathematics*, *19*(5), 1145–1190. doi: 10.1007/s10208-019-09427-x

Mazzeu, J., Otuki, T., & Da Silva, S. (2011). *The canonical econophysics approach to the flash crash of may 6, 2010.* Germany.

McInish, T., Upson, J., & Wood, R. A. (2014). The flash crash: Trading aggressiveness, liquidity supply, and the impact of intermarket sweep orders. *Financial Review*, *49*(3), 481–509. doi: 10.1111/fire.12047

Paulin, J., Calinescu, A., & Wooldridge, M. (2019). Understanding flash crash contagion and systemic risk: A micro–macro agent-based approach. *Journal of Economic Dynamics and Control*, *100*, 200–229. doi: 10.1016/j.jedc.2018.12.008

Rand Low, Te Li, & Terry Marsh. (2018). Bv–vpin: Measuring the impact of order flow toxicity and liquidity on international equity markets. *Journal of Risk*. Retrieved from https://www.risk.net//media/download/

U.S. Commodity Futures Trading Commission / U.S. Securities and Exchange Commission. (2010). *Findings regarding the market events of may 6, 2010: Report of the staffs of the cftc and sec to the joint advisory committee on emerging regulatory issues.* Retrieved from http://www.sec.gov/news/studies/2010/marketevents-report.pdf

Wells, N., & Chemi, E. (2016, 08 24). A short history of stock market crashes. *CNBC*. Retrieved 14/03/2020, from https://www.cnbc.com/2016/08/24/a-short-history-of-stock-market-crashes.html

Appendix

Start Time	Start Price	End Time	End Price	Price Change	Ticks
2010-05-06 14:45:36.873	228.5200	14:45:37.043	226.0200	-0.010940	6
2010-05-06 14:45:37.857	227.7500	14:45:37.897	225.1000	-0.011636	6
2010-05-06 14:45:46.060	222.0800	14:45:46.250	220.2600	-0.008195	6
2010-05-06 14:45:51.090	224.5200	14:45:51.163	220.9800	-0.015767	11
2010-05-06 14:45:58.960	217.0450	14:45:59.163	220.0000	0.013615	6
2010-05-06 14:46:03.403	215.6900	14:46:03.453	213.4400	-0.010432	6
2010-05-06 14:46:03.960	212.6000	14:46:04.320	218.3600	0.027093	6
2010-05-06 14:46:05.480	215.0100	14:46:05.667	220.0000	0.023208	6
2010-05-06 14:46:06.643	217.1650	14:46:06.733	214.9900	-0.010015	7
2010-05-06 14:46:09.397	214.4900	14:46:09.897	210.8600	-0.016924	7
2010-05-06 14:46:14.847	210.0225	14:46:15.017	214.9999	0.023699	6
2010-05-06 14:46:15.260	213.9465	14:46:15.480	215.8600	0.008944	6
2010-05-06 14:46:21.783	209.0200	14:46:21.830	206.1900	-0.013539	9
2010-05-06 14:46:22.280	212.3900	14:46:22.697	205.0000	-0.034794	7
2010-05-06 14:46:22.733	206.0000	14:46:22.920	204.2500	-0.008495	14
2010-05-06 14:46:23.523	205.0000	14:46:23.647	203.0000	-0.009756	8
2010-05-06 14:46:27.630	206.0700	14:46:27.860	201.0000	-0.024603	7
2010-05-06 14:46:32.990	210.2400	14:46:33.110	201.2100	-0.042951	6
2010-05-06 14:46:33.353	202.5800	14:46:33.637	200.1000	-0.012242	9
2010-05-06 14:46:38.340	207.7700	14:46:38.380	210.4900	0.013091	6
2010-05-06 14:46:43.507	199.8500	14:46:43.633	205.3800	0.027671	6
2010-05-06 14:46:44.847	200.1400	14:46:44.930	207.8900	0.038723	6
2010-05-06 14:46:51.537	201.7100	14:46:51.557	205.7100	0.019830	6
2010-05-06 14:47:07.473	210.6800	14:47:08.063	215.8400	0.024492	6
2010-05-06 14:47:08.650	216.7800	14:47:08.773	212.0400	-0.021865	6
2010-05-06 14:47:36.380	220.0000	14:47:36.457	223.0000	0.013636	7
2010-05-06 14:47:36.633	222.0400	14:47:36.717	224.0000	0.008827	6
2010-05-06 14:49:16.897	229.5200	14:49:17.057	231.9800	0.010718	6

2010-05-06 14:49:18.993	231.6900	14:49:19.120	229.5300	-0.009323	7
2010-05-21 09:30:00.080	237.5700	09:30:00.473	232.0000	-0.023446	6

List of all crashes found by the Nanex routine

Start Time	Price	Reversion
2010-05-03 11:35:46.300	265.9000	True
2010-05-03 11:35:56.843	265.9700	True
2010-05-03 11:54:57.140	266.4300	True
2010-05-03 12:10:32.670	266.2592	True
2010-05-03 12:10:48.420	266.2001	True
2010-05-03 12:36:25.567	267.3500	True
2010-05-03 12:36:29.780	267.3505	False
2010-05-03 12:53:54.870	267.3200	True
2010-05-03 13:05:17.817	267.3900	True
2010-05-03 13:06:32.200	267.3100	True
2010-05-03 13:06:46.187	267.3501	True
2010-05-03 13:18:34.883	267.2900	True
2010-05-03 13:18:54.330	267.3001	False
2010-05-03 13:25:09.843	267.2709	True
2010-05-03 13:57:05.740	267.6337	False
2010-05-03 14:31:40.213	267.2999	False
2010-05-13 11:46:36.180	263.1230	True
2010-05-13 13:27:58.030	262.9899	True
2010-05-13 13:30:18.547	263.0980	False
2010-05-13 13:30:22.943	263.1400	True

2010-05-13 13:47:38.780	262.5012	True
2010-05-13 13:47:58.823	262.5699	True
2010-05-13 14:13:03.800	262.8501	True
2010-05-13 14:13:26.490	262.8960	True
2010-05-13 14:14:18.363	262.8801	True
2010-05-13 14:15:07.273	262.8250	False
2010-05-13 14:24:01.707	262.4800	True
2010-05-14 13:29:28.700	252.2911	True
2010-05-17 13:53:07.213	249.3200	True
2010-05-17 13:53:08.920	249.3600	True
2010-05-18 12:15:06.707	253.1250	True
2010-05-18 12:15:22.820	253.2685	True
2010-05-24 13:48:37.510	249.9086	True
2010-05-26 13:14:43.167	250.0420	True
2010-05-27 14:04:00.397	250.9750	True
2010-05-27 14:04:24.653	251.0400	True
2010-05-28 13:42:10.277	255.1100	False
2010-05-28 13:44:01.610	255.1000	False
2010-05-28 13:44:05.000	255.1800	True

List of all crashes found as outliers