

University of Zurich

Department of Economics
Institute for Banking and Finance
Prof. Dr. Thorsten Hens



**University of
Zurich**^{UZH}

Bachelor Thesis

Supervisor: Anastasiia Sokko

The different theories of the 2010 Flash Crash with main focus on high-frequency trading

Author: Laurent Angehrn
Matriculation Number: 13-715-909
Erismannstrasse 33
8004 Zürich

+41 78 952 09 53
Laurent.Angehrn@uzh.ch

Course of Studies: Banking and Finance

Date of Submission: 31.01.2017

Executive Summary

This paper investigates the plausibility of different theories of the “Flash Crash” on May 6, 2010, when the *Dow Jones Industrial Average* experienced its biggest intraday drop, before rebounding within few minutes. It is based on the official report of the *U.S. Securities and Exchange Commission (SEC)* and the *U.S. Commodities Futures Trading Commission (CFTC)*, which concludes, that mainly an erroneous sell algorithm issued by a large fundamental trader is mainly responsible for the market crash. Their claim was proven to be partially flawed and thus, further investigation was needed.

Firstly, early emerged theories are analyzed and inter alia proven untrue. The paper’s primary focus lies on studies conducted by academic and independent researchers, trying to explain the cause of the “Flash Crash” and their criticism towards the official explanation by the SEC and CFTC. Furthermore, the regulators’ claim, whether Navinder Sarao single-handedly was significantly responsible for triggering the “Flash Crash”, is studied. The results show, that he may is guilty of using manipulative trading strategies, but not for causing the “Flash Crash”.

Chapter 4 compares the different theories about the “Flash Crash” and judges its plausibility. In conclusion, the market events of May 6th were a result of the interplay of uncertain market conditions, episodic illiquidity, reporting delays, broken cross-market arbitrage and extraordinary selling pressure. Newer theories are typically more credible, mainly due to the extensive data basis recent studies rely on and previous findings they can build upon.

The next section concentrates on the role of high-frequency traders (HFTs) in the market crash. It provides the reader with a general introduction about high-frequency trading and their most frequently used legal and illegal trading strategies. A comparison of benefits and drawbacks of HFTs for the market stability and overall welfare of market participants is conducted, whereas the negative arguments prevail significantly.

Chapter 6 focuses on the implemented and currently planned regulatory changes in response of the “Flash Crash” to prevent a repetition of this occurrence. Subsequently, a summary of suggested regulations by academics in order to stabilize the markets and inhibit illegal/unethical high-frequency trading techniques is composed.

The last part studies legal actions and detection methods against the manipulative strategy “spoofing”, which is often used by HFTs for the purpose of disrupting market prices in their favor. Further, it discusses the support of third parties for regulators in detecting spoofing activities and provides suggestions for more effective spoofing prevention and detection on an automated basis. This could be achieved by more efficient data collection, improved coordination of market regulators and exchanges and automated pattern recognition.

Table of Contents

Executive Summary	I
List of Figures	IV
List of Abbreviations	V
1. Introduction	1
2. The “Flash Crash” According to CFTC and SEC	3
2.1 The Time Before the “Flash Crash”	3
2.2 Credit Default Swap Spreads	3
2.2.1 Volatility Index	4
2.2.2 Liquidity Replenishment Points.....	5
2.2.3 Liquidity	5
2.3 Events During the “Flash Crash”	6
2.3.1 One Big Bad Trade	6
2.3.2 Crash in the E-Mini Market	6
2.3.3 Crash of Individual Stocks	8
2.4 Aftermath of the “Flash Crash”	9
3. Causes of and Explanations for the “Flash Crash”	10
3.1 Early Theories	10
3.1.1 Financial Terrorism	10
3.1.2 Fat Finger Theory.....	10
3.1.3 Price Reporting Delays	11
3.2 Critics on the SEC/CFTC – Report	12
3.2.1 CME Group	12
3.2.2 Technological Incompetence.....	12
3.2.3 Nanex	13
3.3 Navinder Singh Sarao Case	16
3.3.1 The Charge	16
3.3.2 Sarao Pleads Guilty	17
3.3.3 Criticism of Allegations	17
3.4 Academic Research	18
3.4.1 Kirilenko, Kyle, Samadi and Tuzun (2011, 2014)	19
3.4.2 Easley, Lopez de Prado and O’Hara (2010)	20
3.4.3 Menkveld and Yueshen (2016).....	22
3.4.4 Aldrich, Grundfest and Laughlin (2016)	22
4. Analysis and Comparison	23
4.1 Early Theories	23
4.2 Plausibility of the CFTC/SEC Joint Report	24
4.3 Nanex	25
4.4 Plausibility of Academic Research	26
4.4.1 Kirilenko, Kyle, Samadi and Tuzun (2011, 2014)	26
4.4.2 Easley, Lopez de Prado and O’Hara (2010)	27
4.4.3 Menkveld and Yueshen (2016).....	28
4.4.4 Aldrich, Grundfest and Laughlin (2016)	29

4.5	Conclusion	30
5.	High-Frequency Trading.....	31
5.1	Description	31
5.2	Strategies of HFTs	32
5.2.1	Market Making	32
5.2.2	Arbitrage Trading.....	33
5.2.3	Statistical Arbitrage	33
5.2.4	Key Word Algorithms.....	34
5.2.5	Order Flow Signals	34
5.2.6	Unethical/Illegal Strategies.....	34
5.3	Benefits and Drawbacks for the Markets	36
5.3.1	Benefits of High-Frequency Traders in Financial Markets	36
5.3.2	Drawbacks of High-Frequency Traders in Financial Markets.....	37
6.	Regulator’s Responses to the “Flash Crash”	39
6.1	Implemented Regulations and Measures since May 6, 2010	39
6.1.1	Single Stock “Circuit Breaker”	40
6.1.2	Prohibition of Unfiltered Market Access	40
6.1.3	Prohibition of Market Maker Stub Quotes	41
6.1.4	Consolidated Audit Trail (CAT)	41
6.1.5	Cancellation Fees.....	41
6.1.6	Limit Up-Limit Down Plan	42
6.1.7	Compliance and Integrity Improvement.....	42
6.1.8	CFTC: Disruptive Trade Practices	43
6.2	Suggested Regulations.....	43
6.2.1	VPIN Measure.....	43
6.2.2	Financial Transaction Tax.....	43
6.2.3	Cancellation Fees.....	44
6.2.4	Minimum Resting Time.....	44
6.2.5	Speed Bumps.....	44
7.	Spoofing (Detection).....	45
7.1	Regulators	45
7.2	Third Parties	46
7.3	Suggestions for automated Spoofing Detection	47
8.	Conclusion.....	50
	References	52

List of Figures

Figure 1: Credit Default Spreads on Greek Sovereign Debt	4
Figure 2: CBOE SPX Volatility Index Intraday Levels on May 6, 2010	5
Figure 3: E-Mini Volume and Price.....	7
Figure 4: SPY Volume and Price.....	9
Figure 5: Trading Volume of the W&R Sell Algorithm	14
Figure 6: Cumulative Algorithm Volume Percentage	15
Figure 7: Sarao’s Algorithm Activity	18
Figure 8: “Hot Potato” Volume.....	20
Figure 9: E-Mini S&P 500 Futures’ VPIN metric on May 6, 2010	21
Figure 10: HTG’s Spoofing Illustration	36
Figure 11: Illustration of Spoofing Activity in Gold Futures Market.....	49

List of Abbreviations

%	Percent
ATS	Alternative Trading System
Bp	Basis Points
CAT	Consolidated Audit Trail
CBOE	Chicago Board Options Exchange
CBOT	Board of Trade of the City of Chicago
CDF	Confidence Interval
CFTC	U.S. Commodities Futures Trading Commission
CME	Chicago Mercantile Exchange
CQS	Consolidated Quotation System
CT	Central Time
CTS	Consolidated Tape System
DJIA	Dow Jones Industrial Average
Dodd-Frank Act	Dodd-Frank Wall Street Reform and Consumer Protection Act
DoJ	U.S. Department of Justice
E-Mini	E-Mini S&P 500 Futures
ECB	European Central Bank
EST	Eastern Time
et al.	et alli
FINRA	Financial Industry Regulatory Authority
FTT	Financial Tracking Technologies, LLC
GMT	Greenwich Mean Time
HFT	High-Frequency Trader
ICE	Intercontinental Exchange
IEX	The Investors' Exchange
LRP	Liquidity Replenishment Point
MiFID II	Markets in Financial Instruments Directive II
MPID	Market Participant Identifier
NASDAQ	Nasdaq Composite
NBBO	National Best Bid and Offer
NMS	National Market System
NYSE	New York Stock Exchange
OTC	Over-the-Counter
p.	Page
P&G	Proctor & Gamble
PIN	Probability of Informed Trading
Reg ATS	Regulation Alternative Trading Systems

S&P 500	Standard and Poor's 500
SEC	U.S. Securities Exchange Commission
SPY	S&P 500 SPDR Exchange Traded Fund
TRF	Trade Reporting Facility
TSE	Tokyo Stock Exchange
U.S.	United States
U.S.A.	United States of America
VIX	Volatility Index
VPIN	Volume-Synchronized Probability of Informed Trading
W&R	Waddell and Reed Financial Advisors

1. Introduction

Since the 17th century, stock market crashes and so-called bear markets have had a severe impact on humanity. Starting with the famous “tulip mania bubble” in 1637, caused by speculations on contracts for bulbs of tulips in the Netherlands, these phenomena have occurred on a more and more frequent basis (Dash (2000)). As a legal consequence of the tulip mania bubble, the Dutch parliament implemented a decree that changed the futures contracts of tulip bulbs into options contracts. Hence, purchasers of tulip contracts were not obliged to buy the bulbs anymore, but were able to pay a penalty and disclaim the receipt of the bulbs (Thompson (2007)).

Apart from the two big stock market crashes in 1929 and 2007/08 that have led to global financial crisis, and the passing of the “Glass-Steagall Act” in 1932 as well as the Dodd-Frank Wall Street Reform and Consumer Protection Act (“Dodd-Frank Act”) and “Basel III” in 2010, many crashes did not have such a severe impact on the global markets, but highlighted fundamental flaws in the corresponding regulatory environments on global, transnational as well as national levels (Acharya (2012)).

One of the most influential stock market crashes of this century occurred on May 6, 2010 and resulted in a temporary loss of the prices of many U.S.-based equities of over 9% in the *Dow Jones Industrial Average* (DJIA), *Standard and Poor’s 500* (S&P 500) and the *NASDAQ Composite* (NASDAQ), before rebounding promptly within a few minutes (SEC (2010a)). This phenomenon is publicly known as the so-called “**Flash Crash of 2:45¹**” (Bates (2015)).

According to the official report of the CFTC and the SEC (2010b), the main cause, additional to the already tense situation in the market due to the investors’ worries about the debt crisis in Greece at that time, was a large fundamental trader’s² sale program which sold 75’000 E-Mini S&P 500 futures contracts³, valued at about \$4.1 billion. The high degree of market uncertainty combined with a thinning of liquidity⁴ provision in addition to the mentioned sale program allegedly resulted in the stated stock market crash.

Throughout this event’s following months and years, several theories and research papers were published by different authors, trying to explain the happenings of May 6, 2010. High-frequency traders (Nanex (2010a)), order imbalances (Easley, Lopez de Prado and O’Hara (2011)), broken cross-market arbitrage (Menkveld and Yueshen (2016)), reporting delays (Aldrich, Grundfest and Laughlin (2016)) and sole individuals (CFTC (2015)) have been blamed to be responsible for the event.

¹ All stated times within this thesis correspond to Eastern Time (EST) = Greenwich Mean Time (GMT) -5.

² The large fundamental trader later turned out to be the American asset management company Waddell & Reed Financial Inc. (Nanex (2010a))

³ E-Mini S&P 500 futures are electronically traded futures contracts and based on the underlying S&P 500 index (CME Group (2017)).

⁴ The term “liquidity” refers to the market depth of buy- and sell-side, which incorporated the resting orders of market participants willing to buy or sell at a certain price (SEC (2010b)).

In this respect, the aim of this paper is to provide the reader with an extended overview on research and different theories that have been conducted in this field of study and regarding the introduced “Flash Crash”, and the resulting impact on market regulations. Hence, this thesis addresses the following questions:

- Which explanation approaches are more plausible, which less?
- What was the response of the regulators to this stock market crash in order to prevent a repetition of this occurrence?
- Is it possible that Navinder Singh Sarao single-handedly was significantly responsible for the “Flash Crash” of 2:45?
- How can regulators detect or even prevent spoofing on an automated basis?

In order to investigate these research questions, this paper analyzes plausibility of the different explanation approaches through an evaluation of the particular argumentations and models used by regulators, academics and individual researchers. Subsequently, the author of this thesis provides a deeper insight into implemented, planned and suggested regulations in response of the “Flash Crash”. Further, it investigates Navinder Sarao’s role in the “Flash Crash” and whether it was caused by his spoofing activity. As academics claim, high-frequency traders (HFTs) played a significant role on that day, the thesis will lastly analyze the impact of high-frequency trading for financial markets and deepen into the spoofing. This thesis will conclude with own suggestions for an automated spoofing detection method to improve regulators surveillance efficiency and effectivity.

This thesis mainly focuses on the developments in U.S. financial markets, developments of other markets are used for comparative purposes only.

2. The “Flash Crash” According to CFTC and SEC

As reported by CFTC and SEC (2010b), a huge selling order of E-Mini futures by a large fundamental trader, additionally to the already volatile market conditions, has caused a lack of liquidity and therefore a severe drop in price of the S&P 500 and other related indices. In the course of about 30 minutes, prices of the most important U.S. indices dropped more than five percent before rebounding quickly. The whole event of the “Flash Crash” as well as the incidents that led to it will be discussed in detail in the following.

2.1 The Time Before the “Flash Crash”

According to the official report of the CFTC and SEC about the preliminary findings regarding the market events which lead to the “Flash Crash” on May 6, 2010 (2010a), the corresponding financial markets have contained growing uncertainty due to alarming news about the European debt crisis, particularly connected to Greece’s looming crisis. As can be seen in Figure 1 below, this uncertainty manifested itself through high volatility, a preference for quality investments among investors⁵, rising prices for credit default swaps⁶ (CDS) used as protection against default by the Greek government, and a decline of the Euro in comparison to the global currency market. Some of these aspects will be discussed further in the following.

2.2 Credit Default Swap Spreads

On May 6, at 8:30 a.m., the *European Central Bank* (ECB) started a press conference during which the possibility of purchasing Greek government bonds through the ECB has not been addressed. This led to a widening of the corresponding spreads on CDS protecting against a default by the Greek government on their sovereign debt by more than 11% on the 6th of May compared to the previous day, from 844.2bp (basis points) up to 937.9bp (see Figure 1) (SEC (2010a)).

⁵ Flight to quality investments, such as for example gold (SEC (2010b)).

⁶ Credit default swaps are an insurance against credit risk. In case of default of the underlying insured asset is the protection seller obliged to pay the insured amount to the protection buyer (Pettinger (2008))

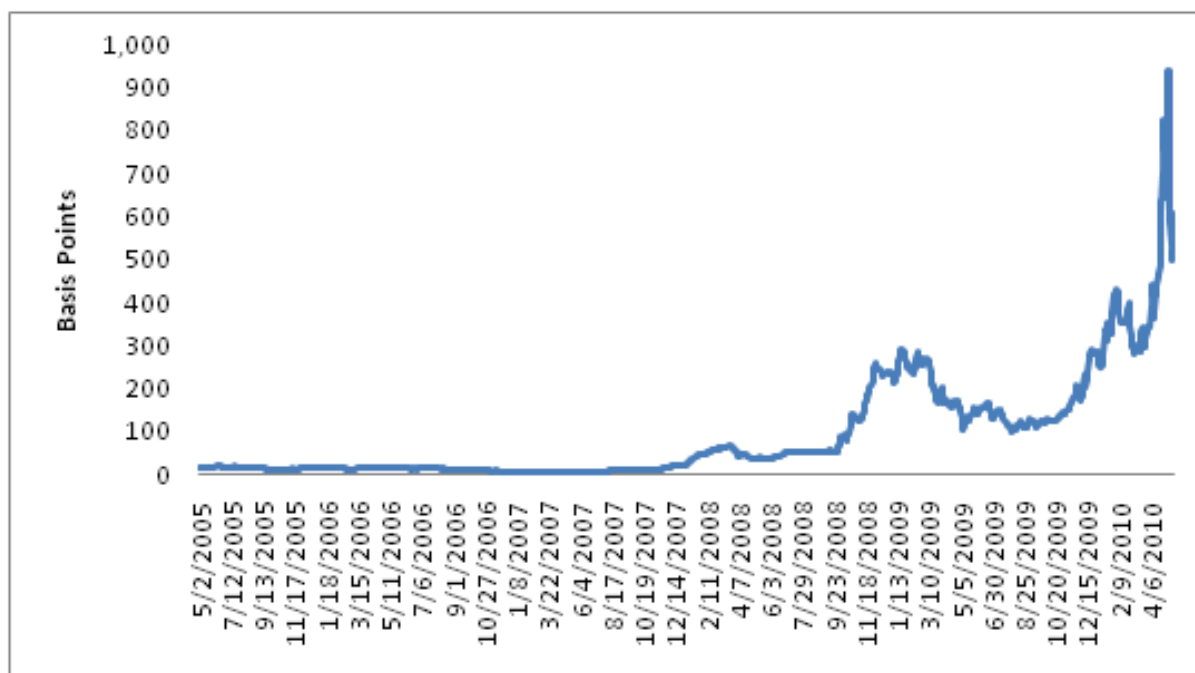


Figure 1: Credit Default Spreads on Greek Sovereign Debt

Source: SEC (2010a), p.15

2.2.1 Volatility Index

Regarding the market volatility, the *Volatility Index (VIX)* by the *Chicago Board Options Exchange SPX* serves as a reliable measurement tool. It is calculated through option prices and is a measure of expected market volatility of the *S&P 500* index. In this respect, the VIX stands for the expected range of returns of the *S&P 500* within the next 30 days, whereas a higher value represents a higher volatility, and thus higher expected range of (positive as well as negative) returns (SEC (2010a)).

In the beginning of the 18th week in 2010, on May 3, the VIX level started at 22.41, which was the highest value of the day. Due to the increasing uncertainty in the markets, the VIX level rose to 25.88 as of the beginning of Thursday, May 6. This represents an increment of 15.5% within three trading days. After remaining constant to a greater or lesser extent, the VIX level started to rise at about 1:30 p.m., and half an hour later, it reached an increase of 10.5% from the opening level, to 28.60 points (see figure 2 below). Until 2:30 p.m., the VIX incremented continuously to 31.71 points, an intraday rise of over 22.5%, representing the rising expected volatility of the *S&P 500*. Within the next 16 minutes, this represented “fear” accelerated to 40.26 points, the highest level since August 4, 2009. This 61.6% increase from the previous day’s closing value was the fourth-largest single-day increase of the VIX in its history⁷. At the close of the trading markets on May 6 the VIX level rested on 32.80 points, an increase of 31.7% from the previous day’s closing (SEC (2010a)).

⁷ The three highest single-day increases in the VIX’s history were on 10/19/87 (312.95%), 10/13/89 (68.30%), and 2/27/07 (64.22%) (SEC (2010a)).

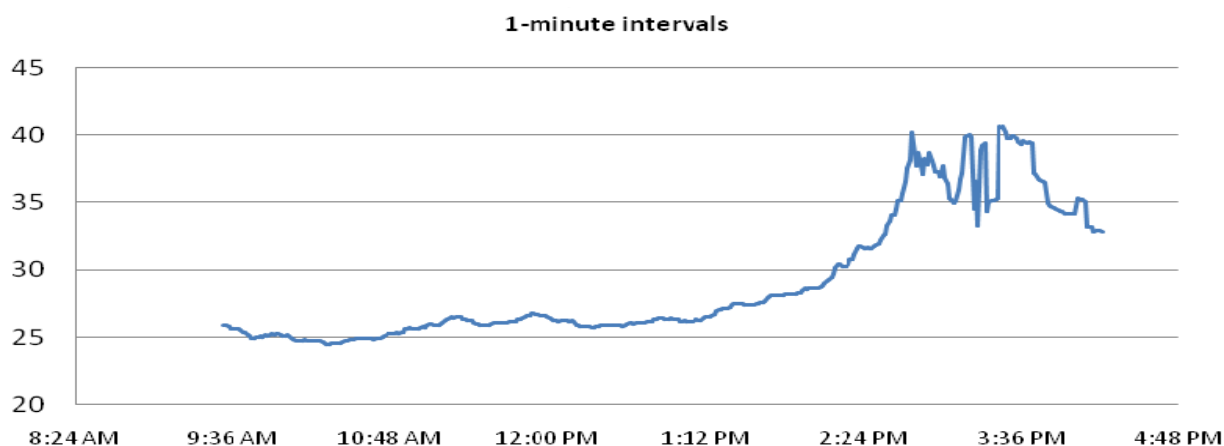


Figure 2: CBOE SPX Volatility Index Intraday Levels on May 6, 2010

Source: SEC (2010a), p.13

2.2.2 Liquidity Replenishment Points

An additional indicator of the overall increased risk in the markets during the time around May 6, 2010, manifested itself in the inclining price volatility of individual equities. So-called “Liquidity Replenishment Points” (LRP’s) which are utilized by the *New York Stock Exchange* (NYSE) and can metaphorically be seen as “speed bumps”, are used to soften a stock’s volatility by temporarily converting the automated market into a manually functioning auction market. This happens if a certain stage of price movement is exceeded. In that case, the trading frequency of the given stock is slowed down and will even be paused in order to allow additional liquidity to enter the market before returning to the automated stock market. Due to the increased volatility, the amount of securities which triggered LRPs were far above average before 2 p.m. on May 6, 2010. Even though more than 75% of all LRP events were resolved within one second or less, between 2:30 p.m. and 3:00 p.m., more than 1’000 securities triggered LRPs for more than one second, whereas a normal day’s average is between 20 to 30 of these events (SEC (2010b)).

2.2.3 Liquidity

According to the CFTC’s and SEC’s report on findings regarding the market events of May 6, 2010 (2010a), another important factor of the market’s insecurity during that time was the thinning buy-side liquidity in the *E-Mini S&P 500 futures contracts* (E-Mini) and the *S&P 500 SPDR exchange traded fund*⁸ (SPY). These two are the most active stock index instruments traded on the electronic futures and equity markets. In the time around May 6, 2010, the E-Mini’s buy-side liquidity had fallen from almost \$6 billion during the early morning to \$2.65 billion, which accounted to a decline of 55%. The SPY fell about 20% from about \$275 million

⁸ The *S&P 500 SPDR exchange traded fund* is an exchange traded fund that tracks the *S&P 500* (Bloomberg (2017)).

to \$220 million intraday. This is representative for the market participant's uncertainty which is correlated to the "flight to quality", such as investments in gold (SEC (2010a)).

2.3 Events During the "Flash Crash"

2.3.1 One Big Bad Trade

This chapter (2.3) is fully based on the official report of the CFTC and SEC (2010b). While the markets were highly volatile and denoted increasing thinning of liquidity during the time of May 6, 2010, at 2.32 p.m., a large fundamental trader started an algorithmic-based selling program of 75,000 E-Mini contracts in order to hedge a portion of the risk in its \$75 billion investment portfolio. The total amount of the selling program was valued at roughly \$4.1 billion dollars and was programmed to feed selling orders into the June 2010 E-Mini market, regardless of the price or time, but targeting an execution rate of 9% of the trading volume calculated over the previous minute. The sell algorithm's execution, while markets were already highly uncertain and volatile, took just 20 minutes, resulting in the largest net change of daily positions of any trader in the E-Mini market since January 1, 2010.

This immense selling order has mainly been absorbed by HFTs, fundamental buyers and other intermediaries in the futures market, as well as "cross-market arbitrageurs". Cross-market arbitrageurs are market participants that benefit from small price differences between related products, by selling the more expensive product and buying the cheaper equivalent, over different markets. Thus, the selling pressure has been transmitted to equity markets through buying E-Mini contracts and selling equities in the *S&P 500* index.

The mentioned algorithm's first selling batch was mainly hit by HFTs, specifically accumulated 3300 contracts. Between 2:41 p.m. and 2:44 p.m. on May 6, 2010, 2'000 of these contracts have been sold again for the sake of reducing the corresponding investors' long positions. During the same time period, over 140'000 E-Mini contracts have been traded by HFTs, equivalent to 33% of the market's total trading volume. Not aggregating a larger inventory than 3'000 to 4'000 contracts while trading a large number of contracts is a common practice among HFTs.

The increased trading volume was anticipated by the sell algorithm of the mentioned, large fundamental trader based on an increase of the selling volume, although the selling orders already sent were not fully absorbed by other market participants. This serves as an example that a high trading volume, especially during times of high market volatility, does not automatically imply high liquidity.

2.3.2 Crash in the E-Mini Market

Due to the increased selling pressure, implemented by the beforementioned selling algorithm, HFTs, and other market participants, the E-Mini decreased by 3% between 2:41 p.m. and 2:44 p.m. on May 6, 2010, as figure 3 shows below. Cross-market arbitrageurs

transferred this impact to the SPY by buying the E-Mini, and synchronously selling equivalent numbers in the equities markets which also resulted in a loss of the price of SPY by 3%.

Since the fundamental buyers and cross-market arbitrageurs did not provide sufficient demand for the overflow of contracts, the HFTs started to aggressively buy and resell the contracts to each other and thereby created a so-called “hot potato”-volume-effect⁹, according to the CFTC and SEC. Within the 14 seconds after 2:43:13, over 27'000 E-Mini contracts have been traded by HFTs while only buying a net of 200 additional contracts. The buy-side market depth of the E-Mini, which is on average approximately 100'000 contracts (\$5.5 billion), dropped by almost 99% to about \$58 million by this time. This resulted in an additional price drop of the E-Mini by 1.7% to its intraday low of 1056.

Between 2:41 p.m. and 2:45:27 p.m. on May 6, 2010, the prices of the E-Mini had decreased by more than 5%, and the ones of the SPY by more than 6%. At 2:45:28 p.m., the E-Mini's buy-side depth was less than 1'050 contracts (>1% of the beginning of the day) and the buy-side resting orders of the SPY had dropped about 75% of the beginning of the day's level, to just about 600'000 shares which is equivalent to 1'200 E-Mini contracts.

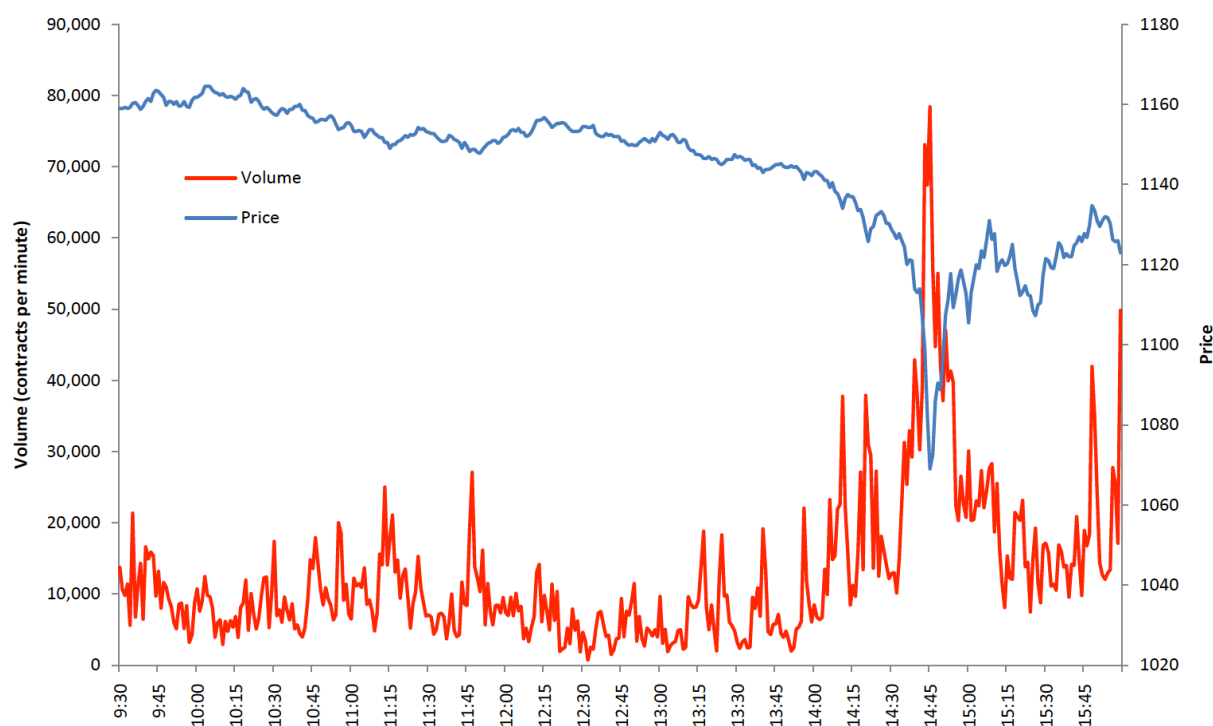


Figure 3: E-Mini Volume and Price
 Source: SEC (2010b), p.19

In the following 13 minutes after 2:32 p.m., the above mentioned sell algorithm managed to sell about 35'000 E-Mini contracts at a combined value of almost \$2 billion. All fundamental sellers together sold more than 80'000 contracts net in that time period, which is about 15 times more than during the same time interval in the previous three days.

⁹ The „hot potato“-volume-effect occurs, if securities are quickly bought and sold again between a group of market participants. Fast back and forth trading of securities can lead to immense increase of trading volume (SEC (2010b)).

In order to dampen the sell pressure in the E-Mini, the *Chicago Mercantile Exchange's* (CME) "Stop Logic" functionality¹⁰ was triggered at 2:45:28 which caused a trading pause of five seconds. After trading did resume, prices started to stabilize and recover, in the E-Mini as well as in the SPY case. The sell algorithm continued its operations at about 2:51 p.m.

2.3.3 Crash of Individual Stocks

The impact of the mentioned large fundamental trader's sell algorithm in the E-Mini market was transferred as well to the equity markets, mainly by market-arbitrageurs and, thus, influenced the prices of individual stocks. Many over-the-counter (OTC) market makers who usually trade internally with orders from retail customers started directing their trades into the public exchanges where they competed with other market participants, while there was a lack of liquidity. At about 2:45 p.m. on May 6, 2010, many market participants stopped trading due to the declined prices observed in the minutes before. This resulted in a strong decrease of provided liquidity in the markets. Figure 4 below shows the trading volume and price of the SPY during May 6th. Various market makers and liquidity providers started to widen their spreads, decrement liquidity or even stop trading at all.

Even though prices started to increase again after 2:45 p.m. in the E-Mini and SPY market segments, selling orders still found decreased buying interest, thus, led to further price declines in certain securities. As liquidity had completely vanished for some securities and ETFs, buying and selling interest could not be immediately fulfilled, which led to trades being executed at irrational prices, such as one penny or \$100'000. These trades were the result of so-called "stub quotes"¹¹.

Between 2:40 p.m. and 3:00 p.m. on May 6, 2010, the total trading volume exceeded \$58 billion, while over 98% of all shares were traded within 10% of their initial value from 2:40 p.m. By 3:00 p.m., most securities had been traded again at realistic prices, but during the previous 20 minutes, over 20'000 trades representing 5.5 million shares had been executed at prices of more than 60% away from the initial prices from 2:40 p.m.

¹⁰ CME's Stop Logic functionality is a detection tool for potential market movements triggered by Stop orders which would lead to a price move beyond pre-specified thresholds. In this case, trading halts for usually five to ten seconds (CME Group (2017)).

¹¹ "Stub quotes" are generated by market makers (and exchanges, on their behalf) in order to fulfill their two-sided quoting requirements without the intention of having them filled. In times of high volatility, market makers place orders (like bid prices of \$0.01 and ask prices of \$100'000) to maintain their requirement. Their intention is not to fill those orders, but if computerized systems cannot find any other offers, these stub quotes may eventually get hit (SEC (2010b)).

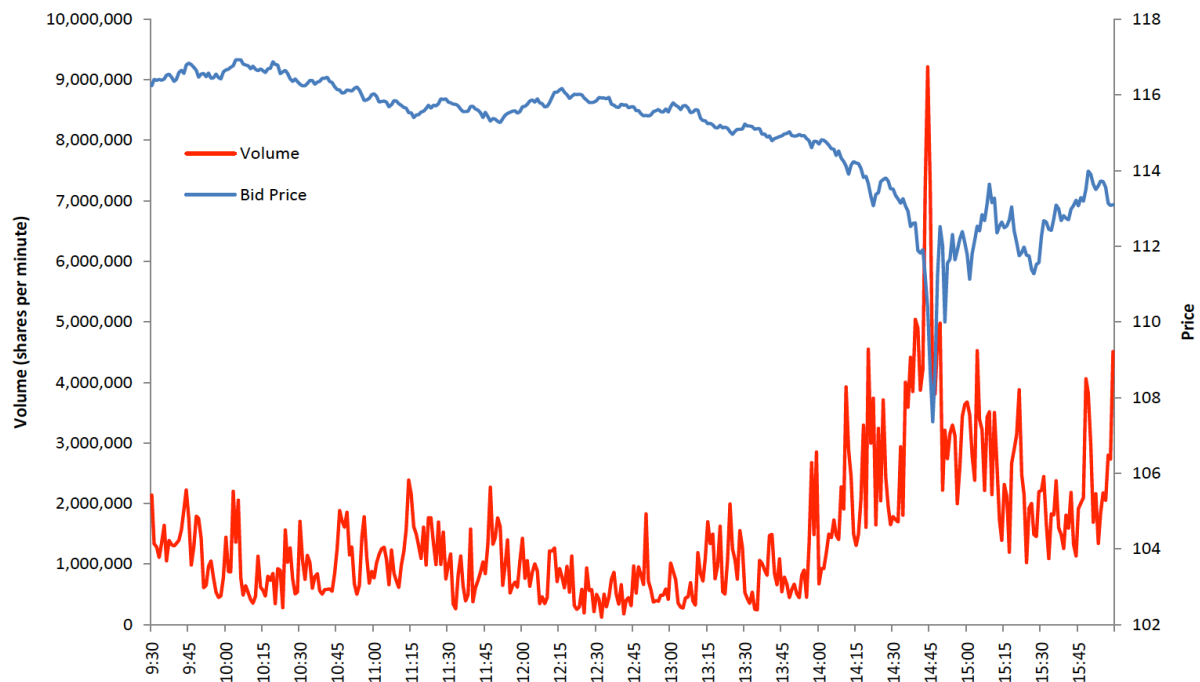


Figure 4: SPY Volume and Price
 Source: SEC (2010b), p.20

2.4 Aftermath of the “Flash Crash”

At 3:42 p.m. on May 6, 2010, stock prices were more or less back on their initial price level before the large fundamental trader’s sell algorithm started executing on that day. But the immense shock had left its impact on the corresponding markets. After their closing on May 6, 2010, the affected exchanges and the *Financial Industry Regulatory Authority* (FINRA) agreed on breaking trades that had been executed at unrealistic prices under severe market conditions. More than 60% of the cancelled trades were executed at prices under \$1.00 and about 5% above \$100. The number of broken trades concentrated on only a few market participants which were still providing liquidity, while most of the other market participants stopped trading. As an example, at approximately 3:29 p.m., 895 shares of Apple Inc. had been traded at stub quote prices of \$100’000, but at the end of the day, they were broken by the exchanges and the FINRA (SEC (2010b)).

The “Flash Crash” was the first financial market crash caused by algorithms that gained a large amount of media attention. From that day on, the mysteries around HFTs began to disassemble and the topic “High-Frequency Trading” first perceived public interest. One of the “Flash Crash’s” consequences was the widespread concern about the reliability of electronic markets and their regulators, in combination with untrustworthy trade algorithms (Aldrich, Grundfest and Laughlin (2016)). Agreeably, the greatest fear of exchanges and regulators was the loss of trust in their capabilities by investors. SEC chairperson Mary Schapiro confirmed this by stating at a round-table regarding the “Flash Crash”, “[...] our concern is not whether a single firm might fail, but whether it causes collateral damage to

investors and their confidence in the integrity and stability of our markets.” (SEC (2012a), p.6).

3. Causes of and Explanations for the “Flash Crash”

3.1 Early Theories

During the occurrence of the “Flash Crash” and shortly after, many different speculations and theories about its causes were spread around the markets. Often, these were disproven shortly after. In the following, the author will provide an overview on those theories and explanations.

3.1.1 Financial Terrorism

One of the first theories to explain the happenings of the “Flash Crash” of May 6, 2010, was that it was an act of financial terrorism on financial markets. It has been speculated that enemies of the western society may have caused the above discussed turbulences on the corresponding financial markets. In particular, China, Russia, and even the terrorist group Al-Qaeda have been blamed for issuing a cyber-attack in the futures markets, and thus disrupting the prices. This should have led to the 998.5-point drop of the *Dow Jones Industrial Average* (DJIA) within a few minutes (Simpson (2010)). But the regulators and investigators have found no evidence to prove any severe interaction of these groups within the American financial markets which could have caused such an impact (SEC (2010a)).

Nevertheless, since the CFTC and SEC came up with their explanation of what caused the “Flash Crash”, namely that one single trade had caused such havoc, a discussion started if well-funded groups could provoke a deliberately engineered “Flash Crash” (Ghosh (2010)). On September 11, 2010, members of the *House Committee on Homeland Security* held an informal meeting regarding this topic. According to their testimony, “foreigners from overseas” could use the sponsorship of an American brokerage firm in order to gain “naked access” to the exchanges. Additionally, their identity could be hidden by a sub-custodian chain. On August 22, 2010, the FINRA announced that it will conduct special examinations on broker firms in order to make sure that their customers had risk control in place to prevent the servers from disrupting the financial markets (McTague (2010)).

3.1.2 Fat Finger Theory

According to other market observers, the “Flash Crash” could have been triggered by a so-called “Fat Finger”. The corresponding “Fat Finger Theory” describes an erroneous trade by a human market participant in a financial market. Adding too many zeros to an order or by (e.g.) falsely typing billion instead of million may cause huge turbulences on exchanges and great losses for the issuer of the order (Phillips (2010)). Rumors were spread that a trader of

Citibank accidentally entered a trade to sell shares worth \$16 billion instead of \$16 millions of *Proctor & Gamble* (P&G). Since P&G represents 4.5% of the *DJIA*, this trade had triggered the downward trend of the markets (Chevreau (2010)).

However, this theory has been disproved quickly by the fact that the share price of P&G dropped after a significant decline in the E-Mini and that safe guards of the CME Group would make it impossible to conduct such an erroneous trade (CME Group (2010)).

3.1.3 Price Reporting Delays

Price reporting delays of exchanges and so-called “Alternative Trading Systems¹²” (ATs) were accused of being significantly responsible for triggering the “Flash Crash”. According to the analysis of a “bank that shall not be named¹³”, during May 6, 2010, many exchanges were overwhelmed by the tremendous trading volume, and thus could not coordinate their “Circuit Breakers” in time (Flood (2010)). “Circuit breakers” are measures approved by the SEC and used by ATs and exchanges to temporarily stop trading activity if the volatility rises to a critical level in order to curb panic-selling. “Circuit breakers” activate if e.g. the *S&P 500* has an intraday drop/rise of more than 13% or if an individual stock drops or rises more than 15% within five minutes. When exchanges halt the trading in such an event, liquidity shall be restored and the volatility should decrease (SEC (2010b)).

As reported by the above mentioned “bank that shall not be named”, ATs had mechanical problems with the reporting and pricing of stocks being traded on May 6, 2010, with delays of up to five minutes. Also, the NYSE, the world’s largest stock exchange, confirmed that delays were occurring on that day, mainly caused by an internal system update of the NYSE (Flood (2010)). After 20 minutes, the delays of the ATs were caught up, but the NYSE still had delays of up to 24 seconds according to market research firm *Nanex*. As a result, brokers had great difficulties to serve the “National Best Bid and Offer” (NBBO) system¹⁴ and best prices were often outdated, so that they could not be executed. This led to a drying out of liquidity, thus, also profitable opportunities for HFTs which had access to the direct quote services (such as NYSE’s “Open Book”, a service costing thousands of dollars every month) (Nanex (2010a)).

Jeff Donovan from *Nanex* stated that the market crash of May 6, 2010, began when the 24 seconds lag in the NYSE appeared. Such delays are not uncommon, further research of *Nanex* shows that a double in quote rates is enough to provoke these delays. HFTs would be able to trigger such delays and generate profits out of the exchanges’ incapability (Nanex (2010a)). NYSE spokesman Ray Pellecchia claimed that these delays were caused by the issues on May 6, 2010, and definitely not a contributor to it (Flood (2010)).

¹² Alternative Trading Systems work similar to exchanges, as a market place for buyers and sellers of securities, but comply with the regulations of broker dealers and do not perform self-regulation (Tuttle (2013)).

¹³ The bank prefers to stay anonymous (Flood (2010)).

¹⁴ The NBBO is a regulatory system, issued by the SEC, that obliges brokers to execute customer trades at the best available prices across all exchanges (SEC (2010a)).

3.2 Critics on the SEC/CFTC – Report

The official report of the SEC and CFTC on the happenings around May 6, 2010 (2010b), blames one large institutional trader for issuing an erroneous sell algorithm for the E-Mini market for being responsible for triggering the “Flash Crash”. After the report had been released, critics scrutinized its reliability by proving the analysis and conclusions as wrong. In the following, the author will highlight the main arguments of said critics.

3.2.1 CME Group

The most prominent critic of the joint SEC/CFTC report on the “Flash Crash” was released by the *Chicago Mercantile Exchange Group* (CME Group) within 24 hours after the report has been published. CME Group’s main argument is the outspoken doubt that a single trader could have caused such havoc in the stock markets. First of all, CME Group affirmed that its markets did indeed function the way they were supposed to, and that their automated credit controls, order quantity limitations, stop and market price protection points, price banding procedures¹⁵, and stop logic functionality were effective in the challenging market environment around said events. CME Group stated that the 75’000 contracts traded by the blamed institutional asset manager only represented 1.3% of the total traded E-Mini volume of 5.7 million contracts. During the time period the above described sell algorithm had executed its trades, the corresponding activity was representing less than 9% of the total trading volume (CME Group (2010)).

The algorithm was designed to place orders in relatively small quantities and which dynamically adapted to the market liquidity. At 13:45:28 on May 6, 2010, when the markets were at their lowest prices, the institutional trader’s algorithm represented less than 5% of the total sales in the market. More than half of the corresponding trader’s order volume was executed as the markets were recovering. Thus, the CME Group deems the SEC’s and CFTC’s allegation that the sell algorithm of the institutional trader was the main influencer of the “Flash Crash”, as implausible (CME Group (2010)). Nanex (2010a) backs the statement of the CME Group with own research and concludes as well, that the algorithm did not have such a severe impact as claimed in the official report of the SEC and CFTC (2010b).

3.2.2 Technological Incompetence

In their paper “Federal Market Information Technology in the Post Flash Crash Era: Roles for Supercomputing”, the authors Bethel, Leinweber, Ruebel and Wu (2011) accused the SEC and CFTC of being incapable to do their job efficiently with their actual infrastructure. The authors argue with the fact that it took more than four months to issue an official report about

¹⁵ Price banding is a mechanism used by the CME Group to ensure all incoming orders are within a realistic price spectrum. Obvious erroneously priced orders are being rejected by the price banding system (CME Group (2016)).

a few minutes of market activity, which demonstrates that improvements on the SEC's and CFTC's side are overdue. Most of the corresponding data had to be collected and analyzed manually which shows that the regulators' infrastructure is not up-to-date. According to Dan Hubscher, head of capital markets strategy for the *Progress Software Corp.*, the technology used by HFTs is way more advanced than the regulators' IT infrastructure (Vicci (2012)).

In order to effectively regulate the corresponding markets, the SEC and CFTC have to use state of the art technology (Vicci (2012)). SEC Chair Mary Jo White on the other hand stated in a press release on November 15, 2016, that the SEC has approved the "national market system" (NMS) plan to create a "consolidated audit trail" (CAT). This system will enable the regulators to more efficiently oversee market participants and assess a comprehensive data set of trading activities more timely, enabling them to conduct research and identify or investigate misconduct (SEC (2016a)). The implementation process of the CAT started on May 26, 2010, as a consequence of the "Flash Crash", with its first proposal (Sifma (2016)). However, it took more than six years to develop and approve the NMS plan, which, according to the author of this thesis, again represents the difficulties of bureaucrats trying to adapt to state-of-the-art technology used in the market.

3.2.3 Nanex

Nanex LLC is a market transaction data stream service that also provides clients with real-time data. Apart from its business, *Nanex* became famous for its own independent in-house research about market events and market micro-structure analysis. Four days before the release of the official CFTC/SEC Report on "Findings Regarding the Market Events on May 6, 2010", *Nanex* published its own "Flash Crash Summary Report". This report mainly blamed the quote delays of the NYSE to the "Consolidated Quoting System¹⁶" (CQS) of up to 24 seconds additional to the negative news of the Greek Parliament, as well as quote saturation. About 400 milliseconds before the above described heavy selling of E-Mini futures began, the quote traffic of all stocks in the NYSE, *NYSE Arca* and *Nasdaq* surged to saturation levels. This surging led to delays in many feed processing systems, networks, and, consequentially, in addition to the huge selling program caused the crash in the stock markets. *Nanex* also named the beforementioned large institutional investor which initiated the described sell algorithm as *Waddell and Reed Financial Advisors (W&R)*, according to the actual trading data of that day (Nanex (2010a)).

After the release of the CFTC/SEC's report, *Nanex* started to contradict the findings by its own research and interviews of people involved. As it is illustrated in Figure 5 below which was published by *Nanex*, the selling rate of the before discussed selling algorithm dropped sharply while the E-Mini futures market declined precipitously – and contrary to the CFTC/SEC's argumentation in their report (Nanex (2014)). The SEC's and CFTC's argumentation was that the algorithm "responded to the increased volume by increasing the

¹⁶ The Consolidated Quoting System collects and publishes bid, ask, volume and price information of all underlying securities. Exchanges are obliged to transmit timely and accurate data to the CQS in order to inform all market participants about actual quoting activities (SEC (2010b)).

rate at which it was feeding the orders into the market [...]” (SEC (2010b)). As illustrated in Figure 5, the trading volume of the algorithm reached its peak during the time of recovery in the markets, not causing the severe drop (Nanex (2014)).

In collaboration with *ZeroHedge*¹⁷, *Nanex* also discovered that the sell algorithm was not executed by *W&R*, but by their broker *Barclays Capital*. They have used their time-tested algorithm called “Participation” which allegedly cannot crash a market. This algorithm is programmed in such a manner that trades will never execute at bid price, but only at the offer price. After an own matching of the 6’438 *W&R* executions during the time of the “Flash Crash”, *Nanex* concluded that the algorithm in question had always posted sell orders above the market price and waited for a buyer, but had never crossed the bid/ask-spread. Thus, the sell algorithm used by *Barclays Capital* never took nor required liquidity. It did not drive the prices down with its increased selling volume after the intraday-low was reached, because it acted passively, not actively (Nanex (2010d)).

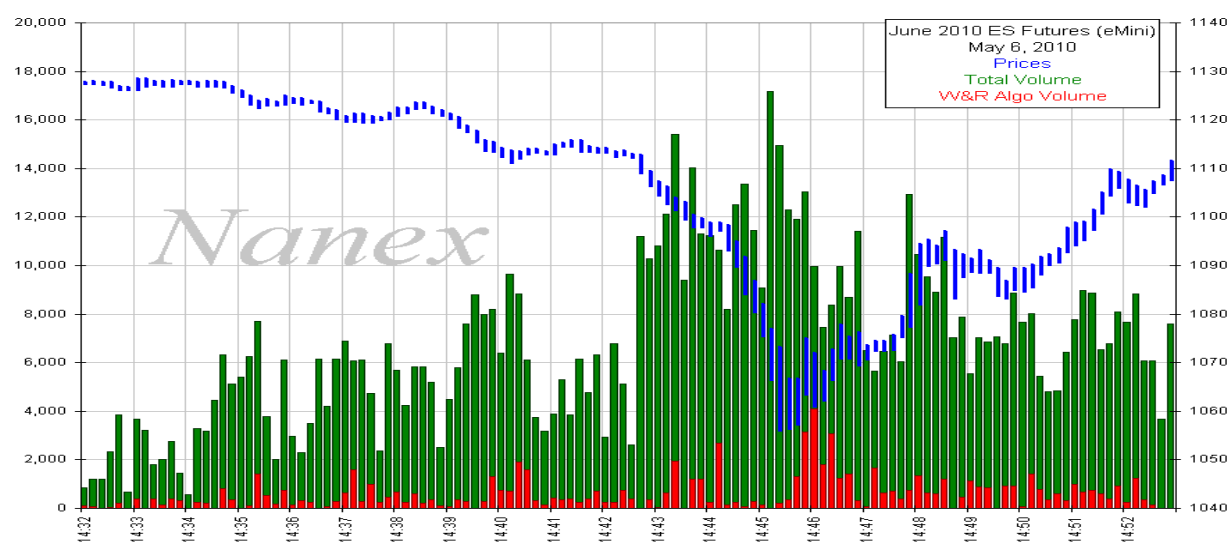


Figure 5: Trading Volume of the W&R Sell Algorithm
 Source: Nanex (2014)

Another flaw in the SEC’s and CFTC’s explanation of the “Flash Crash” found by *Nanex* did cover the programming target of the algorithm in question. As stated in the official report by the SEC and CFTC, the algorithm was “programmed to feed orders into the June 2010 E-Mini market to target an execution rate set to 9% of the trading volume calculated over the previous minute, but without regard to price or time.” (SEC (2010b), p.2) After reconstructing the execution data of the algorithm between 14:32 and 14:52, and comparing the cumulative algorithm volume as percentage of the total volume, it is clearly visible that on the one hand, the algorithm never exceeded 9% of the total volume (see Figure 6). The program indeed prevented trading more than 9%, thus, the upside limitation code worked properly. On the other hand, however, are several drops below 8% and even 7%. The drops in the beginning between 14:32 and 14:34 are explainable by the calibration of the algorithm, but after these

¹⁷ Zerohedge.com is a leading news site for global finance, economics, market, and political analysis.

two minutes, the cumulative algorithm volume percentage should have been stable at 9% (Nanex (2014)).

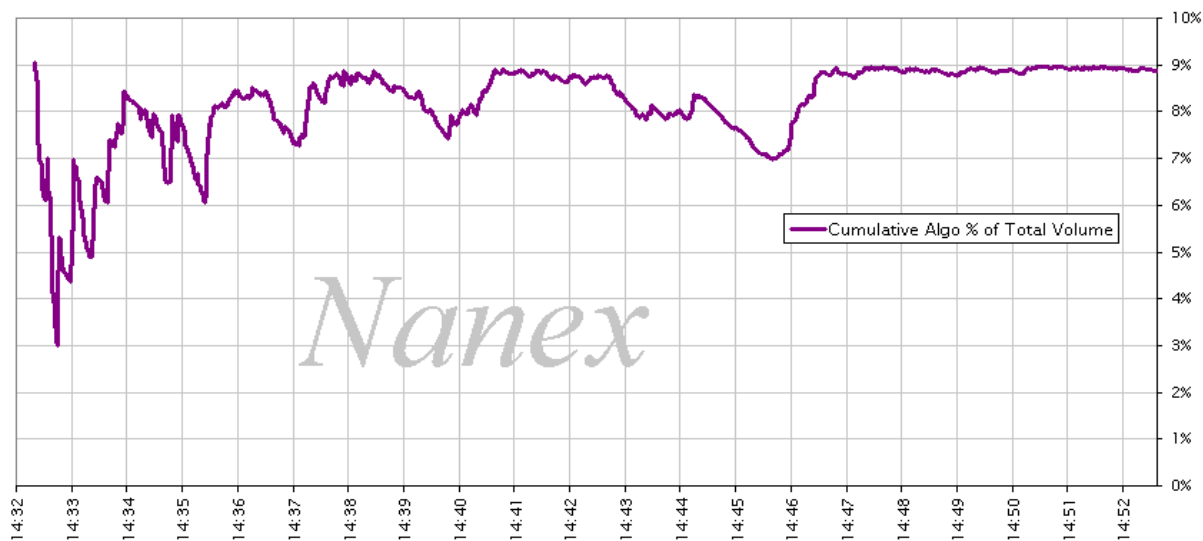


Figure 6: Cumulative Algorithm Volume Percentage

Source: Nanex (2014)

According to the official report of the CFTC and SEC on the “Flash Crash”, the large fundamental trader in question was significantly responsible for triggering the crash, and the HFTs did not have a severe negative impact on the markets on May 6, 2010 (SEC (2010b)). Instead of blaming the sell algorithm of *W&R*, *Nanex* accused the HFTs for being responsible for causing the “Flash Crash”. According to their conclusion, released on October 14, 2010, the algorithm behaved well and careful, and did not impact the market by, for example, selling at the bid price. But from *Nanex*’ point of view, the buyers of those E-Mini contracts were less careful: after the prices declined, they aggressively sold¹⁸ the contracts in batches of more than 2’000 future contracts each, which led to a further price drop. These buyers were mostly HFTs that tried to clear out their positions and transmitted the sudden price drop to other related markets, such as the *S&P 500*, in order to still benefit of arbitrage. Hence, *Nanex* pleads the HFTs of triggering the “Flash Crash” (Nanex (2010b)).

Furthermore, *Nanex* accuses the SEC and CFTC of whitewashing the HFTs in their report, in order to put the main focus on the accused fundamental large trader. In their report on the “Flash Crash”, on page 13, the SEC and CFTC divide the market makers into intermediaries and HFTs, without giving any explicit explanation for it. They designated the top 3% intermediaries (sorted by the number of trades) as HFTs. Andrei Kirilenko, CFTC chief economist and co-author of the CFTC/SEC report, published a research paper on the “Flash Crash” the day after the official “Flash Crash” report was released (Kirilenko et al. (2011)). In his paper, he classified, out of 195 market makers, 16 as HFTs which represents the top 8.2% without clearly explaining his thoughts behind it (Nanex (2014)).

¹⁸ Selling aggressively means to place sell orders at or even below the bid price which leads to price decline (SEC (2010b)).

3.3 Navinder Singh Sarao Case

In 2015, a UK based trader called Navinder Singh Sarao has been indicted on 22 counts of fraud to manipulate the markets and was thereby allegedly significantly responsible for triggering the “Flash Crash” of 2010. On November 9th, 2016, the said trader pleaded guilty to his charge (Whipp and Scannell (2016)). In the following, the author will provide an overview on the Navinder Singh Sarao case.

3.3.1 The Charge

On February 11th, 2015, Navinder Singh Sarao (“Sarao” in the following) was accused of violating the law on 22 counts, ranging from wire fraud to market manipulation between June 2009 and April 2014, including May 4 to May 6, 2010, by the *U.S. Department of Justice* (DoJ) (Whipp and Scannell (2016)). By the time of his arrest, Sarao denied all charges. He mainly traded through his own company, *Nav Sarao Futures Limited Plc.* in West-London, and allegedly used the techniques of “layering” and “spoofing” algorithms¹⁹ in order to trade thousands of E-Mini contracts (Brush, Schoenberg and Ring (2015)). These trading techniques became illegal under 2010’s “Dodd Frank Act”. In this respect, the government accused Sarao of causing artificial prices to exist, due to his aggressive spoofing tactics. Sarao’s trading activity artificially depressed the prices for the E-Mini futures, and that the price rebounded as the algorithm was turned off. His orders represented between \$170 million and \$200 million of downward pressure on the E-Mini price (CFTC vs. Sarao (2016)). The case dragged enormous attention from media all over the world. If Sarao would have been found guilty in all allegations, a maximum sentence of 380 years, as well as civil charges, would have applied (Dakers (2016)).

Five years after the “Flash Crash”, Navinder Sarao was the first individual held responsible for contributing to the market events of May 6, 2010. Allegedly, he gained over \$40 million with illegal trades between 2010 and 2014 (Whipp and Scannell (2016)). Towards all the allegations, Sarao kept claiming that he did nothing wrong, and that he was an “old school point and click prop trader” (U.S.A. vs. Sarao (2015), p.13), using his mouse to trade, having great reflexes, and that he changes his mind very quickly. This should explain why he often cancelled orders before they could be executed (Hope (2015)). Sarao later admitted in the official Criminal Complaint (U.S.A. vs. Sarao (2015)) that he paid a trading company to build a program in addition to his basic trading software in order to conceal his orders more efficiently and protect them for other manipulative market participants.

¹⁹ See section 5.2.6 of this thesis for explanations of these techniques.

3.3.2 Sarao Pleads Guilty

On November 9th, 2016, after months of hearings, Sarao admitted on one count of spoofing, and one count of wire fraud in the *Chicago Federal District Court*. His spoofing activity lasted over a five-year span, including May 6, 2010. On that day, he entered more than 85 so called “spoof-orders”²⁰ in order to manipulate the markets. His orders represented, at certain times, more than 20% of all outstanding sell orders (Whipp and Scannell (2016)). The DoJ dropped the remaining 20 counts as part of an agreement, in exchange of Sarao co-operating with the *Fraud Section* of the DoJ. He agreed to forfeit almost \$13 million of gains of his illegal actions. His sentencing is not yet officially released, but he is expected to be jailed for 78 to 97 months. Additionally, the CFTC announced to declare \$38 million in monetary sanctions and permanent trading bans. Allegedly, Sarao gained \$789’000 in profits during May 6, 2010 (Telegraph (2016)).

3.3.3 Criticism of Allegations

With the accusations of Navinder Sarao, many raised their voice against the fact that one single individual could have single-handedly triggered the market to that extent. *Nanex*’ founder, Eric S. Hunsader, who analyzed the trades of that day, argued that Sarao probably is guilty of using illegal market practices, such as “spoofing” and “layering”, but that his behavior did not trigger the “Flash Crash” itself. He used the following metaphor to explain this statement: “If the “Flash Crash” was a cake recipe, then Sarao was a pinch of salt” (Gandel (2016), 7th paragraph). The following graph in Figure 7 illustrates why, in Hunsader’s opinion, Navinder Sarao is to be held innocent. Roughly two-and-a-half minutes before the “Flash Crash” began due to massive selling orders, Sarao’s algorithm stopped its activity, probably due to reaching a historic extreme level in the E-Mini price.

²⁰ Placing orders without the intention of fulfilling them to manipulate the market (see section 5.2.6 of this thesis).

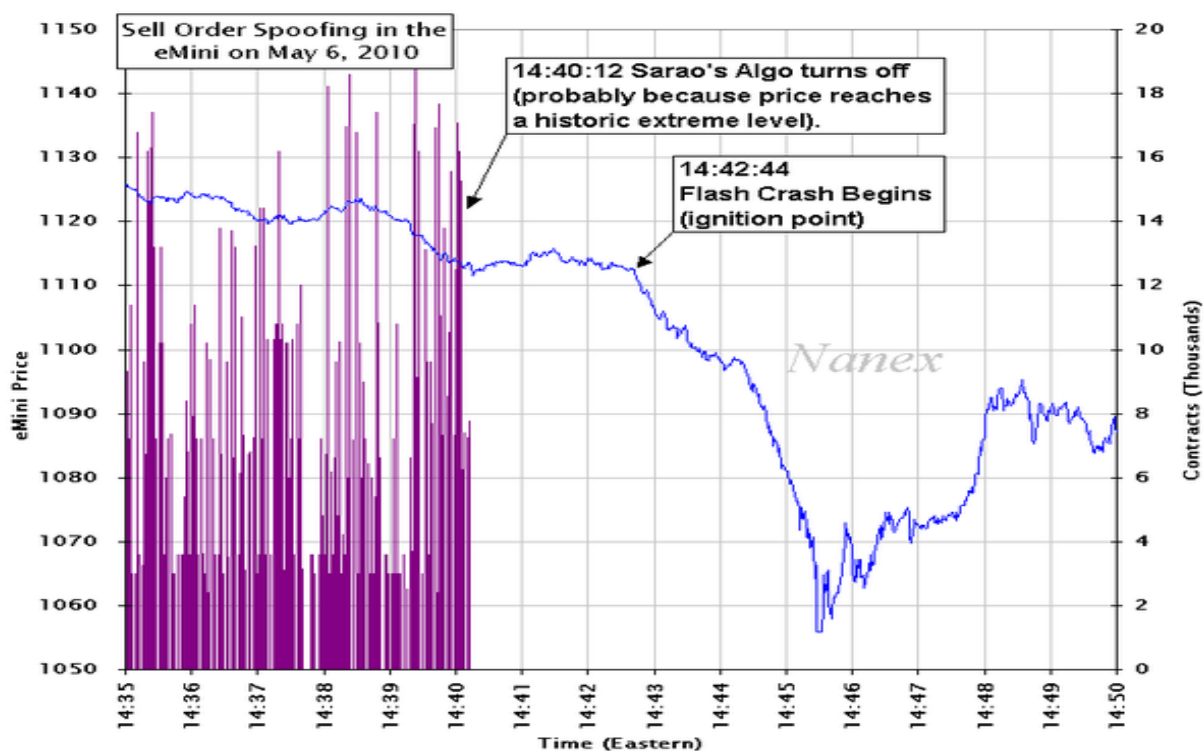


Figure 7: Sarao's Algorithm Activity
 Source: Hunsader (2015)

Joseph Grundfest, former SEC commissioner, published an academic paper along with two professors from the *University of California* and *Yale University* on September 22, 2016. The authors found that Sarao probably did not, and could not, have caused the “Flash Crash”. The study called “The Flash Crash: A New Deconstruction” is allegedly the first analysis of the events from May 6, 2010, at millisecond granularity. It concluded that even though Sarao’s activities might have been illegal, the “Flash Crash” most probably would still have occurred without his participation in the markets. His “layering” should have had none or at max a minuscule impact on the market price, since his orders were sufficiently far away from the marketable prices (Aldrich, Grundfest and Laughlin (2016)).

3.4 Academic Research

Since May 6, 2010, academics tried to simulate and explain the causes and happenings on the markets during the “Flash Crash”. Their findings often contradict each other and the official report of the SEC and CFTC. Therefore, it has to be stated that a comprehensive and inclusive explanation has yet not been found. In the following, the author will present some of these mentioned studies.

3.4.1 Kirilenko, Kyle, Samadi and Tuzun (2011, 2014)

One day after the official report of the CFTC/SEC regarding the events of May 6, 2010, was released, Kirilenko et al. (2011) published a study about the intraday intermediation between May 3rd and May 6th of 2010. Since Andrei Kirilenko was the co-author of the official report, his findings obviously were mostly similar to the ones of the CFTC/SEC report. In their study, Kirilenko et al. (2011) used audit-trail data²¹ and analyzed the trades in the *E-Mini S&P 500* futures market on May 6, 2010. They concluded that HFTs did not cause the “Flash Crash”, but intensified the price movement triggered by the sell algorithm of a single, large individual trader.

In this respect, it is important to understand that a typical practice of HFTs is the so-called “immediacy-absorption”, meaning the removal of all contracts at the bid or ask price, and replacing new best bids and asks orders at adjacent price levels. Under normal market conditions, this practice accelerates the price movement and trading volume, but does not lead to a directional price movement. But in times of high volatility in which prices are moving directionally, this trading activity can further increase the price movement and amplify an already existing market volatility. This increased volatility in turn leads to HFTs acting even faster, replacing the outstanding bid or ask orders with new ones at adjacent price levels (Kirilenko et al. (2014)).

Especially in the last minute of the “Flash Crash’s” down phase, the above described tactic resulted in a “hot potato” effect, meaning that HFTs reduced their inventories against other HFTs. Figure 8 below visualizes the scaled trading volume²² of HFTs and market makers in second intervals during the “Flash Crash” of May 6, 2010. Between 13:45:13 and 13:45:27²³, when prices were plunging with enormous velocity, as it can be seen in figure 8 below, HFTs were responsible for 49% (over 27’000 E-Mini contracts) of the total trading volume, while their net position only changed by 200 contracts (Kirilenko et al. (2014)).

²¹ Audit trail data is collected by the exchanges and consists of time, size and price of each security traded. Further, it involves information on quotes and usually an identification number of the actors involved in each trade (Hasbrouck, Sofianos and Sosebee (1993)).

²² The scaled trading volume is the five seconds moving average of the traded contracts in relation to absolute value of their net holdings (Kirilenko et al. (2011)).

²³ Note, that times is given in Central Time (CT), which is equal to EST -1.

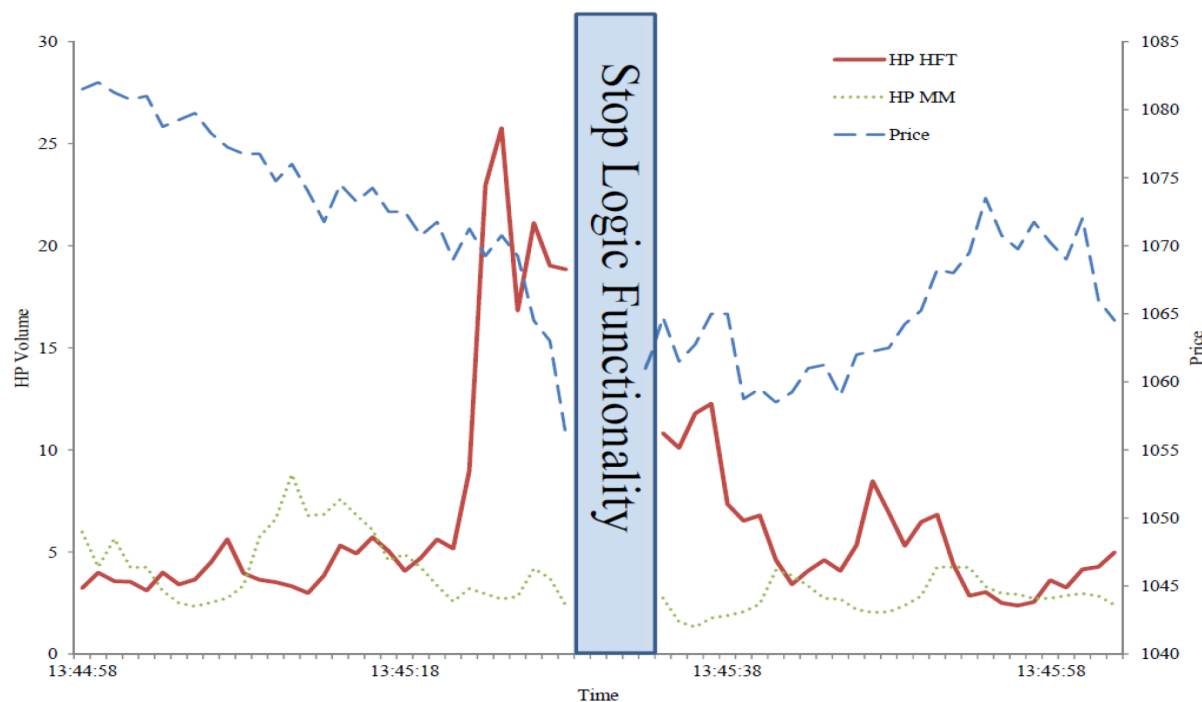


Figure 8: "Hot Potato" Volume
Source: Kirilenko et al. (2011), p.57

According to Kirilenko et al. (2014), regulators should implement actions to encourage HFTs to continuously provide immediacy for other market participants without demanding it. They recommend to desist from additional taxes, limitations, or restrictions, and just change the market design, as for example by a more diligent use of short market wide-trading pauses. This would benefit slower algorithms to decide under which conditions they are willing to replenish liquidity from the markets.

3.4.2 Easley, Lopez de Prado and O'Hara (2010)

In their academic paper "The Microstructure of the 'Flash Crash' – Flow toxicity, liquidity crashes and the Probability of Informed Trading", Easley, Lopez de Prado and O'Hara (2010) acknowledged the findings of the official report by the CFTC/SEC as "factors [that] may have played a role" (Easley, Lopez de Prado and O'Hara (2010), p.13). But their analysis presents a possible explanation, using their measure of the "Volume-Synchronized Probability of Informed Trading" (VPIN). VPIN is a time-varying renewal and a high-frequency estimate of the "Probability of Informed Trading" (PIN) which computes a measure of asymmetric information in financial markets. VPIN allows to capture the order toxicity which affects liquidity provision in stressful market situations, such as the "Flash Crash".

The authors of the paper observed a rising order toxicity in the previous days and hours of the "Flash Crash", and minutes before the before stated and discussed sell algorithm was initiated, the trading volume was high and unbalanced, but liquidity was low. Market makers, such as HFTs, may face losses during times of unbalanced trading volume, due to adverse selection. Thus, and in order to prevent themselves from potential losses,

HFTs withdraw from the markets during times of high order flow toxicity, instead of providing the much needed liquidity during such events. This action exacerbates the liquidity mismatch and hence increases the VPIN further resulting in a downward spiral until it causes a trading halt (Easley, Lopez de Prado and O’Hara (2010)).

In Figure 9 below, Easley, Lopez de Prado and O’Hara (2010) illustrated the VPIN in relation of the E-Mini price on May 6, 2010. The red dashed line represents the confidence interval of the VPIN (CDF) as a measure of how unusual the level is compared to the normal distribution of the E-Mini²⁴.

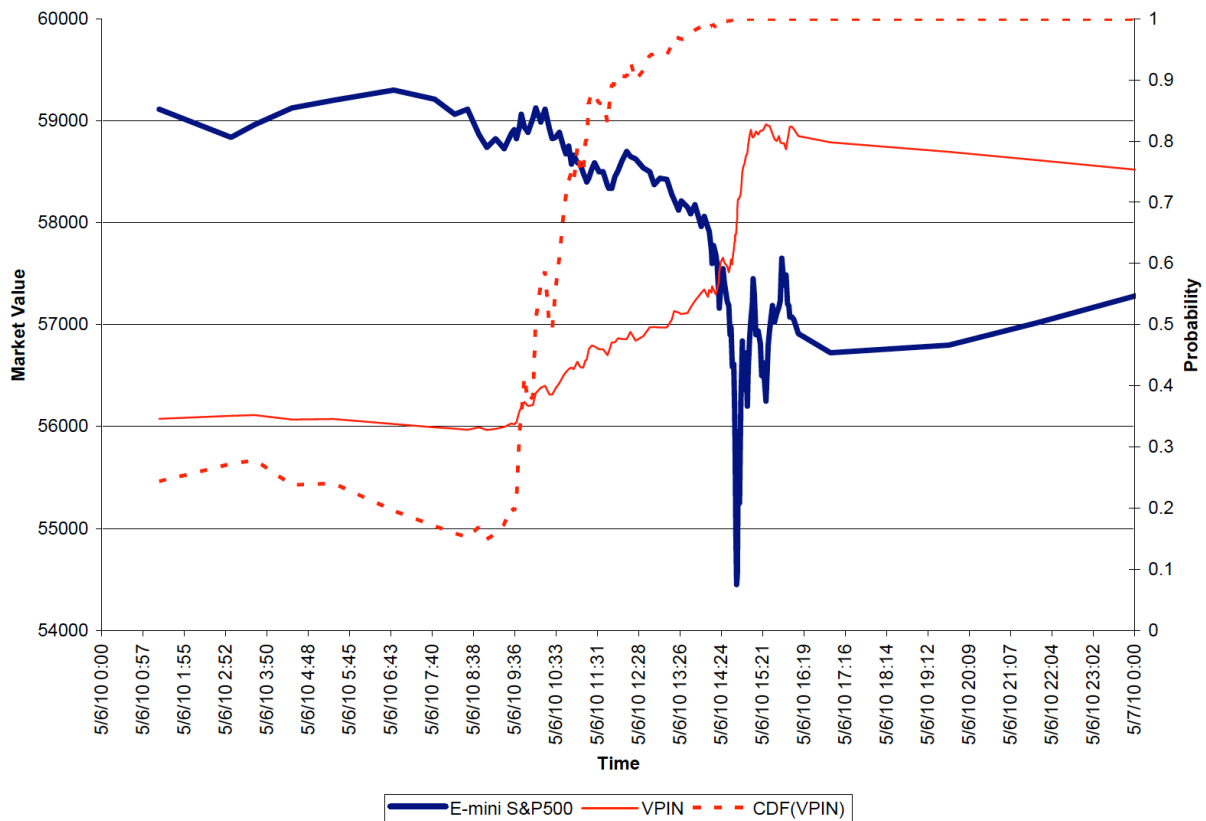


Figure 9: E-Mini S&P 500 Futures' VPIN metric on May 6, 2010
 Source: Easley, Lopez de Prado and O’Hara (2010), p.8

By 11:55 a.m. on May 6, 2010, the VPIN metric exceeded the 90% confidence interval, which itself already represents an alarming level. At 1:08 p.m., it surpassed the 95% level, and two minutes before the crash occurred (according to the CFTC/SEC timeline), at 2:30 p.m., the metric reached its all-time high. Thus, Easley, Lopez de Prado and O’Hara (2010) suggest to use the VPIN as a reliable indicator for upcoming market crashes due to volume imbalances. In their opinion, the “Flash Crash” could have been avoided by recognizing the approaching liquidity crisis, and could have been prevented by market makers providing more liquidity in the market, instead of disposing their inventories and exacerbating the crash.

²⁴ For Example, $CDF(0.44) = 0.8$ means that in 80% of all VPIN metric estimates, the VPIN value is below 0.44.

3.4.3 Menkveld and Yueshen (2016)

Albert Menkveld and Bart Yueshen (2016) analyzed the data of *W&R*, as well as public data on all trades and quotes of the SPY and E-Mini markets around the time of the “Flash Crash”. Their study inspected the “cause-and-effect relationships” of the trade events across markets in 25 milliseconds granularity. They discovered that the beforementioned large fundamental trader overpaid for his demand of liquidity and immediacy due to broken cross-market arbitrage.

Firstly, Menkveld and Yueshen (2016) computed the price *W&R* had paid for immediacy which resulted in a range between \$98.6 million and \$117.3 million (by applying different benchmarks), representing 2.34% or 2.77% of the roughly \$4.1 billion overall paid, respectively. According to historical equity-market evidence, the price for immediacy is more than three times higher than the upper bound of historical price pressure, which accounts for 0.78%. Thus, the large fundamental trader clearly overpaid for his transactions on May 6, 2010, compared with historical standards.

Menkveld and Yueshen (2016) also found that the timing of the trades contradicted the price pressure, since during the minute of the most precipitous decline in E-Mini price on May 6, 2010, the corresponding algorithm contributed only 4% of net sales. Furthermore, *W&R*'s trading program sold most of the E-Mini contracts after the trading halt, as the price rebounded, which contradicts the fact of paying a threefold average for price immediacy.

Further analysis by Menkveld and Yueshen (2016) showed that *W&R* overpaid due to failed cross-market arbitrage. One minute before the said halt, cross-market arbitrage, which usually is a fundamental source of liquidity, suddenly vanished, until eight minutes after the halt. During this time frame, opportunities for arbitrage between the SPY and E-Mini market of up to 3.5% existed.

An additional contribution of the paper is the characterization of the sell algorithm used by *W&R*. The algorithm followed a 9% volume target and turned more aggressive, the more he fell short of the volume target. It sold about 50% of the contracts passively, which means consuming the bid orders of other market participants. The participation algorithm only sold aggressively if prices just rose or the bid depth was too large (Menkveld and Yueshen (2016)). This observation contradicts the analysis of *Nanex*, which concluded that the participation algorithm used on that day was only acting passively, and thus, could not have consumed any market liquidity (Nanex (2010d)).

3.4.4 Aldrich, Grundfest and Laughlin (2016)

The “Deconstruction” of the “Flash Crash” by Aldrich, Grundfest and Laughlin (2016) was the first analysis of the full order book activity of the E-Mini and SPY markets on May 6, 2010, at millisecond granularity. The authors claimed to have found new evidence by observing the day's activity in more detail than ever done before. The main focus of the paper lies on the

analysis of the FINRA's "Trade Reporting Facilities" (TRF) price oscillation. All TRF are obliged to transmit the transacted prices to the "Consolidated Tape System"²⁵ (CTS) in order to consolidate all transactions, their volume and prices, and provide real-time data for exchange-traded securities. Aldrich, Grundfest and Laughlin (2016) stated that during the "Flash Crash", caused by the traffic volume, TRF feeds were delayed. Thus, many trading firms downsized their trading activity due to data integrity problems induced by the delayed TRF feeds. They proved that the TRF oscillation had a significant impact on the E-Mini prices before and during the "Flash Crash". Many market participants stopped trading due to the price oscillations, consequently, resulting in a collapse of liquidity in the SPY. The E-Mini market was not affected in a similar manner by the price oscillations because of its greater depth of market. Nevertheless, bid-side cancellations incremented fiercely, which, in addition to the selling pressure originated by the sell algorithm of *Waddell & Reed*, led to a price drop in the E-Mini.

By 2:46:45 p.m. of May 6, 2010, the liquidity in the SPY market reappeared simultaneously as the TRF oscillations vanished, and not directly after the CME halt, as explained in the CFTC/SEC report (SEC (2010b)). Additionally, Aldrich, Grundfest and Laughlin (2016) analyzed the above introduced Sarao case (see section 3.3 of this thesis) and did not negate that Sarao conducted a felony or created artificial prices and volatility to appear in the market. But they concluded "[...] that the Flash Crash was not a reasonably foreseeable consequence of Sarao's illegal spoofing activity..." (Aldrich, Grundfest and Laughlin (2016), p.57). The authors even suggested that even without Sarao's involvement in the market during May 6, 2010, the "Flash Crash" could still have occurred. The paper terminates, by appeal to the SEC and CFTC, to invest more in data integrity and prevent market participants from manipulating the markets by anomalous trades which lead to confusion during the "Flash Crash" (Aldrich, Grundfest and Laughlin (2016)).

4. Analysis and Comparison

This part consists of the writer's own analysis on the different explanation approaches, aiming to evaluate the plausibility of the particular argumentations and models used to describe the market events of May 6, 2010. Furthermore, it will address whether Navinder Sarao single-handedly was significantly responsible for causing the "Flash Crash".

4.1 Early Theories

For the sake of completeness, the analysis includes the plausibility of the early theories although two out of three were disproven shortly after they were released.

²⁵ The Consolidated Tape System receives and collects all prices for equity transactions across all SEC regulated exchanges. It consolidates the actual prices and disseminates the uniform data back to the market (Aldrich, Grundfest and Laughlin (2016)).

Financial terrorism, meaning the causation of turbulences in the financial markets by enemies of the western society, has been declared as incorrect by the regulators as they did not find any evidence of vast participation of these enemies during the period of turmoil (McTague (2010)). Nevertheless, I admit, that the threat of financial terrorism is present. Foreign groups with immense financial back up and the technological know-how could be capable of disrupting financial markets on short-term using illegal trading strategies. Considering, how much havoc an erroneous algorithm can cause (e.g. Knight Capital's algorithm in 2012 (Jones (2013))), it is easy to recognize the potential threat of financial terrorism.

The claim, a trader of Citibank entered a sell offer of P&G shares worth \$16 billion instead of \$16 million by mistake, which is also known as the fat finger theory, was proven wrong. Easley, Lopez de Prado and O'Hara (2011) showed that the E-Mini price dropped before the P&G share price declined significantly, which consequently refutes the plausibility of the theory.

Price reporting delays, in my opinion, have played a crucial role in triggering the market events of May 6. The extraordinarily high trading volume during the day caused an overload of exchanges' and ATSS' price reporting systems, which resulted in delays of up to five minutes (Flood (2010)). These delays affected the markets immediately, as concerns about data integrity rose and hence some automated trading systems paused or even halted their activity which contributed to a further withdrawal of liquidity. Interviews with market participants who withdrew their liquidity from the markets during the crash revealed that price delays contributed significantly to their decision (SEC (2010b)). The claim of NYSE's Ray Pellecchia, that these delays definitely did not contribute to the market events (Flood (2010)) are wrong in my opinion, for the reason that the withdrawal of liquidity and the increased uncertainty about the markets was fueled by the price reporting delays.

4.2 Plausibility of the CFTC/SEC Joint Report

The official report released roughly four months after the "Flash Crash" (SEC (2010b)) seemed to contain a plausible explanation at first glance. But further investigation conducted by academics dismantled the reliability of the report.

In my opinion, the most crucial factor of a successful investigation is the completeness and availability of the underlying data. Especially in the high-frequency environment, trading data must be at least on a millisecond basis in order to compile meaningful results. The report about the preliminary findings, released twelve days after the "Flash Crash", lacked of available data and thus was not able to conduct a convincing conclusion (SEC (2010a)).

Over the following months, the SEC and CFTC have gathered full order books from NYSE, NASDAQ and BATS, representing approximately 90% of the executions on May 6. The data provided consisted of order book summaries on a minute-by-minute basis for all underlying securities traded at the exchanges. Most of the analysis is focused on trades at

per-second granularity (SEC (2010b)). This fact reduces the significance of the report, as most incidents examined later were based on millisecond-data analysis (Aldrich, Grundfest and Laughlin (2016); Menkveld and Yueshen (2016); Nanex (2014)). Also the fact, that throughout the whole report, no word on significant spoofing activity was mentioned undermines the plausibility of their findings. The unavailability of necessary millisecond data makes it doubtful to come up with a persuasive explanation.

The joint report did detect a lack of liquidity provision during the turbulent minutes in the markets caused by the high level of uncertainty. High trading volume, combined with already nervous market conditions was correctly identified as a contributor to vulnerable markets.

The large selling program of W&R, which increased selling pressure, was allegedly significantly responsible for the following market crash (SEC (2010b)). Despite these allegations, the first interview was held with people actually involved in the execution of the sell algorithm (ZeroHedge (2011)) not until two weeks after the release of the official report (CFTC (2010)). This leads me to conclude, that the SEC and CFTC conducted their official version without having extensive insight in the mechanics of the blamed algorithm. Moreover, until today, I was unable to find any evidence of legal proceedings against W&R, which allegedly have caused the turmoil in financial market with their erroneous sell algorithm. Conclusively, I assume that the regulators doubt the plausibility of their own explanation, caused by disagreeing comments of inter alia the CME Group (2010), wherefore no legal investigation has been initiated.

Nevertheless, their argument that the E-Mini selling orders issued by W&R were picked up by HFTs, which started to resell them aggressively to each other and thus exacerbated the price decline (SEC (2010b)), was proven right by further research (Nanex (2014a); Kirilenko et al. (2010); Menkveld and Yueshen (2016)).

4.3 Nanex

The independent research firm Nanex has gained great attention for their analysis of the “Flash Crash” and has been cited in many academic papers ever since. They are known for scrutinizing press releases and reports of regulators by conducting own investigative research within short time²⁶. Nanex’s research is based on profound analysis of real-time data it provides through his own high-performance ticker plant, called NxCore (Nanex (2017)).

The “Flash Crash” analysis report of Nanex did not include any simulation models in contrast to academic research, but focused solely on trading data and quotes. This report showed the difficulties of the exchanges to process the high trading volume, resulting in quote saturation which led to delays in certain feed processing systems. Their analysis further concludes that most stub quotes were executed after the market hit his low point, thus

²⁶ Nanex issued a “Flash Crash” analysis questioning the official report of the SEC and CFTC within two weeks (Nanex (2010b)).

did not cause the “Flash Crash”. Furthermore, Nanex focuses on the negative effect of LRPs during the crash, which allegedly causes stocks after their hit of an LRP to trade at trade at substantially lower prices afterwards. These findings were published three days before the official report of the regulators was released (Nanex (2010a)). In my view, it is impressive, that a small company based in Illinois is able to analyze vast amounts of data in a short period of time and provide such ground-breaking conclusions.

Over the next months, Nanex obtained more data to conduct research on, e.g. all trade executions of W&R. They showed, that the algorithm was well behaved and did not have such an impact as claimed by the official report (Nanex (2010d)). Instead of accusing W&R, Nanex believes, that mainly HFTs have triggered the Flash Crash through their aggressive trading strategies (Nanex (2014)).

In my view, Nanex has discovered many inconsistencies of the official report and plausibly proved the harmlessness of W&R’s sell algorithm in regard of liquidity withdrawal. Their argumentation is based on hard facts and comprehensible. However, the results of Nanex’s investigation must be interpreted with caution, as it is publicly known, that Eric S. Hunsader, founder of Nanex, is a deprecator of HFTs (Mahmudova (2016)). Since his findings are not of academic nature, he does not include such comprehensive description of his models and analysis procedures as compared to academic research papers. This makes it more intransparent to comprehend his conclusions, of which include conspiracy theories about e.g. corrupt exchanges and regulators (Nanex (2014)).

4.4 Plausibility of Academic Research

4.4.1 Kirilenko, Kyle, Samadi and Tuzun (2011, 2014)

Andrei Kirilenko, who was the primary author of the official report of the SEC and CFTC, published a study about the impact of HFTs during the “Flash Crash” together with Kyle, Samadi and Tuzun. Their paper studies the trading behavior in relation with the net holding of different categories of traders during the time period of May 3rd – May 5th in comparison with May 6th. They conclude, that HFTs did not trigger the “Flash Crash”, but their trading behavior exacerbated market volatility (Kirilenko et al. (2011)).

Kirilenko et al. use audit trail transaction-level data from CME Group on all transactions in the E-Mini market between the 3rd and 6th of May, 2010. The study consists of several models comparing trading behavior second-by-second (as well as minute-by-minute). Market participants have been divided into 6 groups based on their holding horizons and trading strategies. Of the 195 intermediaries, the 7% with highest daily trading frequency have been classified as HFTs (equaling 16 accounts)²⁷. The remaining 15227 trading accounts are divided into fundamental buyers (1263), fundamental sellers (1276), opportunistic traders (5808) and small traders (6880)²⁸ (Kirilenko et al. (2011)).

²⁷ According to own calculations: $16/195 = 8.2\%$.

²⁸ For detailed explanation of each category, see Kirilenko et al. (2011) p.12-14

The paper offers interesting findings about the trading behavior of HFTs, Intermediaries and other market participants during normal and turbulent trading periods (Kirilenko et al. (2011)). Nevertheless, in my view, it uses assumptions and conclusions, which do not necessarily contribute to the explanation of the “Flash Crash”.

Firstly, the model is based on second-by-second (or even minute-by-minute) data, which does not offer the necessary profundity a high-frequency environment analysis requires. Especially, while studying trading behavior of HFTs, a granularity of fractions of seconds is required in order to develop a significant conclusion.

Secondly, the assumption that the top 7% (or 8.2%) intermediaries in daily trading volume automatically represent all HFTs seems too far-fetched. Kirilenko et al. (2011) give no reasonable explanation for this separation. There is a high probability, that there are more HFTs among the intermediaries or the opportunistic traders²⁹. A reallocation of the categories would be likely to distort their results significantly.

Another reason to scrutinize the plausibility of the study is the fact, that price reporting delays were not regarded, whether in the assumptions nor in the models. Even the official report by the regulators claimed, that these price reporting delays had an impact in trading behavior of market participants during May 6th (SEC (2010b)).

Furthermore, only completed transactions were observed, although most of the orders during these days have been modified or cancelled (SEC (2010b)).

Lastly, I claim that there is not sufficient causality between the conclusion and the findings of the paper. Kirilenko et al. find that HFTs aggressively trade in direction of price change and their high trading volume, which does not result in significant changes of their inventory, might be misinterpreted by fundamental traders as liquidity. In order to rebalance their portfolios, HFTs might intensify volatility and compete for liquidity. The researchers do not address whether HFTs have triggered the “Flash Crash” or not, but conclude in the abstract, “[...] that HFTs did not trigger the Flash Crash...” (Kirilenko et al. (2011), p.1). Because of the above-mentioned points, I do not acknowledge the plausibility of this theory as explanation for the “Flash Crash”.

4.4.2 Easley, Lopez de Prado and O’Hara (2010)

The paper of Easley, Lopez de Prado and O’Hara uses their measure of VPIN to explain the order flow toxicity, which led to liquidity withdrawal of HFTs and ultimately resulted in the “Flash Crash”. It claims, that the market events were “the result of the new dynamics at play in the current market structure.” (Easley, Lopez de Prado and O’Hara (2010), p.1).

In order to use the model for their analysis, the authors have used historical data over the period between January 1, 2008 and October 30, 2010 to gather a distribution of average VPIN measures, which is close to the log-distribution. On the basis of this historical distribution, the VPIN metrics can show how unusual the level of order flow toxicity is. An

²⁹ Kirilenko et al. (2011) classified opportunistic traders as market participants, which sometimes show trading behavior like intermediaries and at other times like fundamental traders.

abnormally high level of VPIN has been discovered over the whole week previous to May 6, 2010. It has reached its all-time high two minutes before the crash occurred. HFT market makers started to withdraw liquidity instead of providing it, caused by the extraordinarily high order imbalance, which resulted in episodic illiquidity and following the market crash (Easley, Lopez de Prado and O'Hara (2010)).

The model is based on just few assumptions and due to the fact that trades often occur in milliseconds, it uses volume increments defined through historical data analysis (Easley, Lopez de Prado and O'Hara (2010)). Since it uses historical data and only a few minor assumptions, in my opinion, this model is very useful and reliable in a high-frequency environment. Their theory has explained how increasing order flow toxicity can lead to a market crash and gives suggestions on preventive usage of the tool. But the study lacks of explanation on the reason behind the increased order imbalance, which has been at extraordinary level throughout the previous week as well.

Although this model appears very promising as a forecast tool for liquidity-induced market crashes, there is a shortage in similar crashes to examine its reliability. It solely focuses on transaction data, without considering other metrics such as the market conditions prior and during the period examined. Nevertheless, the extraordinary level of order toxicity, which leads to liquidity withdrawal of market makers, who generally serve as liquidity providers and thus trigger a steep price decline (Easley, Lopez de Prado and O'Hara (2010)), seems plausible to me.

4.4.3 Menkveld and Yueshen (2016)

The research paper "The Flash Crash: A Cautionary Tale about Highly Fragmented Markets" by Menkveld and Yueshen (2016) analyzes the market events in regard of broken cross-market arbitrage. They observe, that one minute prior to the market crash, prices in the E-Mini and the SPY were deviating up to 3.5%. This was probably a consequence of the enormous price decline combined with low response times and mistrust of traders towards the reliability of displayed prices. The additional selling pressure due to the selling program of W&R caused a steep decline of E-Mini prices until the trading halt. W&R has overpaid for immediacy and as a result of the broken cross-market arbitrage.

Menkveld and Yueshen (2016) use W&R's trading data, provided by Nanex, as well as all trading and quoting data of the E-Mini and SPY at 25ms time stamp granularity. The data included information consolidator and distributor, which allowed the authors to implement the information delays for low-latency traders in their analysis. They make use of several relevant models in order to analyze inter alia the amount of overpayment by W&R and the aggressiveness of its algorithm.

The paper includes comprehensive explanations of previous researchers, such as Easley, Lopez de Prado and O'Hara (2010), Grundfest, Aldrich and Laughlin (2016) and Kirilenko et al. (2014) in order to provide a plausible explanation of the "Flash Crash". All models and associated variables used are verifiable and the underlying data was detailed enough to support their line of reasoning. Their argumentation, that broken cross-market

arbitrage has triggered the crash, because it led to intensified selling pressure by W&R and mainly other market participants, seems plausible in my point of view. The only point of improvement would be an analysis of the full order book at millisecond granularity, in order to gain even more relevant insight in these high-frequency saturated markets.

4.4.4 Aldrich, Grundfest and Laughlin (2016)

Aldrich, Grundfest and Laughlin (2016) discovered in their research paper a significant relationship between price oscillations, caused by TRF delays, and liquidity withdrawal in financial markets. They further argue, that the impact of large selling orders combined with liquidity withdrawal can cause crashes similar to the “Flash Crash” of May 6th. The increased occurrence of order cancellations joined with the selling pressure, induced by W&R’s sell algorithm, resulted in a steep price decline of the E-Mini. This price drop was then transmitted to other related markets. Additionally, the authors found, that the price rebound was shortly after the trading halt of the CME, caused by the disappearance of the TRF oscillations.

This explanation is by far the most extensive of all analyzed research papers in this thesis. It uses complete full order book data of the E-Mini and SPY at millisecond granularity and not a subsampled version of data as in many other research papers (Kirilenko et al. (2010); Easley, Lopez de Prado and O’Hara (2010)). Additionally, the authors have received priority data from one of the largest equity market participants, which allows the conduct of an extensive infrastructure latency analysis. They use several models and simulations in order to examine inter alia the impact of Navinder Sarao’s spoofing activity, highlight mechanisms, which possibly have caused the “Flash Crash” and the interaction of HFTs and non-HFTs in financial markets (Aldrich, Grundfest and Laughlin (2016)).

Their analysis moreover contained a comparison with August 9th, 2011, which featured similar geopolitically induced conditions for the markets. Furthermore, they regarded the exceptional amount of order cancellations in the E-Mini on May 6th, which seemed to have triggered the TRF oscillations (Aldrich, Grundfest and Laughlin (2016)).

Overall, this explanation approach is, in my opinion, thoroughly plausible, primarily, thanks to the extensive underlying data. Whether a reexamination of a liquidity-induced market crash is investigated at milliseconds or 25 milliseconds granularity can be the decisive factor. Since they have included order cancellations, information delays and prevailing market conditions into their models and simulations, unlike the other academics, the result becomes more closely to reality.

4.5 Conclusion

In general, I have observed, that newer explanation approaches were more plausible than early research papers and theories. As technological analysis tools further develop, researchers are able to process data more in-depth and gain new insights. Future research can be based on previous findings and address newly arisen research questions.

But academic and independent researchers, as well as regulators face the general difficulty, that for post-crash explanations, prevailing market conditions and other circumstances cannot easily be replicated in models. Furthermore, specifications of algorithms are often not thoroughly addressed, which leads researchers to conduct ambitious behavior analysis of the automated trading programs (Kirilenko et al. (2010); Nanex (2010d)).

Extensive research effort in regard to the causes of the “Flash Crash” has achieved great discoveries over the last years. The official report of the CFTC and SEC (2010b) outlined, that liquidity withdrawal of market makers during uncertain market conditions can have a severe impact on price development. Nanex (2010a) proved the influence of quote reporting delays and that W&R’s selling algorithm did not trigger the market crash as previously assumed by the SEC (2010b). As liquidity vanished from the markets, HFTs started aggressively selling their inventories to each other, causing the price drop (Kirilenko et al. (2011)). As Easley, Lopez de Prado and O’Hara (2010) showed, that increasing order imbalances were affecting the liquidity provision of market makers, which resulted in episodic illiquidity. Another contributing factor to the “Flash Crash”, investigated by Menkveld and Yueshen (2016) was the broken cross-market arbitrage between the E-Mini and SPY. Finally, Aldrich, Grundfest and Laughlin (2016) add the impact of TRF price oscillations as a responsible cause for the risen uncertainty among HFTs, which thereupon withdrew their liquidity from the markets. Withdrawal of liquidity in already nervous markets combined with large selling pressure of W&R’s algorithm and other market participants resulted in the price collapse.

As the explanations have shown, HFTs have an immense impact in the financial world. The increasingly accelerating trading strategies and advances in technology also introduces new risks. Nanex (2010a) blames illegal strategies, such as quote stuffing, used by HFTs, as significantly responsible for overloading exchanges’ servers and causing delays. The CFTC accused Navinder Sarao for single-handedly trigger the “Flash Crash” with his spoofing activities (CFTC (2015)).

The next section will provide a more detailed examination on HFTs impact on electronic markets and regulations, in order to prevent unethical/illegal trading strategies.

5. High-Frequency Trading

5.1 Description

“High-Frequency Trading” is a widely used term for trading with the help of powerful computers and sophisticated algorithms, in order to process information as rapid as possible. It is characterized by trading on short-term horizons, high cancellation of orders, high speeds, high turnover rates, and specialized order types. It can be viewed as the algorithmic trading in finance (Lemke and Lins (2013)). Most big banks on Wall Street, as well as many hedge funds, use algorithmic trading in order to outperform their competitors in terms of speed and efficiency. HFTs usually do not hold excessive amounts of capital and abstain from holding their portfolio overnight. They trade large volumes at high speeds with the goal of gaining a fraction of a cent per trade. Thus, in normal functioning markets, their so-called “Sharpe Ratio³⁰” is dozens of times higher than with traditional trading strategies (Aldridge (2010)). In this respect, on February 20, 2015, *Virtu Financial Inc.* (one of the major HFTs in the U.S.) celebrated a near-perfect record of trading six years with only one day of losses (Mamudi and Picker (2015)).

The technology which enables “High-Frequency Trading” was developed in the 1990’s as computers got more powerful and algorithms more complex. In the year 1998, the SEC implemented the “Regulation Alternative Trading Systems” (Reg ATS), which allowed the creation of an alternative to official stock exchanges with the objective of intermediating seller and purchaser of stocks and other securities. ATS are usually regulated as broker dealers, rather than exchanges³¹, and have to be approved by the SEC (Lemke and Lins (2013)). Additional to the eleven equity exchanges in the U.S., there are over 60 alternative trading systems and 13 U.S. options exchanges which trade equity derivatives (Tuttle (2013)). This fragmentation across the whole country has introduced a variety of complexities to the markets. In the search of liquidity across the ATS and exchanges, high speed becomes more valuable since there is no central market anymore. Price differences on the ATS and exchanges therefore offer the possibility of arbitrage (Menkveld and Yueshen (2016)).

Exchanges started to offer HFTs direct feeds to access their trading information faster than usual market participants, as well as co-location of their servers close to the exchange’s in order to trade faster (O’Hara (2014)). With the advancement in technology, more powerful computers and efficient wires allowed HFTs to decrease their trade execution time of several seconds (as of 2000) to milli- and even microseconds³² by 2010 (Haldane (2010)). Latency, which is the time it takes to send data to an endpoint (e.g. an exchange), gained of huge importance among HFTs because every millisecond might decide about wins or losses. In September 2015, a 4’600km long transatlantic cable system between North America and

³⁰ Sharpe ratio measures the risk-adjusted return. It represents the average return earned over a period of time (minus risk-free rate), divided by the standard deviation of the return (Investopedia (2017a)).

³¹ ATS can apply to be regulated like a securities exchange (Lemke and Lins (2013)).

³² 1 millisecond = 1/1000 second. 1 microsecond = 1/1000’000 second

Great Britain, called *Hibernia Express*, was introduced by *Hibernia Networks* (Hibernia Networks (2016)). This cable system reduced the latency from New York to London from 64.8 milliseconds to 59.6 milliseconds. *Perseus Telecom* implemented a new microwave network between London and Frankfurt to reduce the latency from 8.35 milliseconds to below 4.6 milliseconds (O'Hara (2014)). This represents almost speed of light for these distances. It is therefore hard to imagine how such projects, costing millions of dollars, would ever be profitable until one considers the data from the SEC that shows 23% of all cancelled orders and 38% of all cancelled quotes are perceived within <50 milliseconds of their placement (SEC (2014)).

Many exchanges offer their participants different latencies to access their exchange servers. For example, the *Tokyo Stock Exchange* (TSE) offers a standard line which transfers the orders in several milliseconds. Its priority service (in exchange for higher fees) allows to reduce the latency down to 260 microseconds. By co-locating the trading devices next to the TSE's servers, the latency even decreases to 15.7 microseconds (O'Hara (2014)). In June 2010, *Spread Networks* offered an ultra-low latency connection between Chicago and New York City where the two major U.S. stock exchanges are based. The project costs for the fiber cable for a distance of 827 miles ($\approx 1'331$ kilometers), which lowers the round-trip travel time to 13 milliseconds, accounted for roughly \$300 million. The first 200 HFTs willing to sign up for gaining access to this connection paid collectively \$2.8 billion (Tovey (2014)). It is estimated by *Spread Networks* that profits made from price discrepancies between the two cities might add up to \$20 billion per year (Pasqualle (2014)).

The author of this thesis concludes that this arms race of HFTs to gain better access, have faster processors, and more complex algorithms will continue to go on as long it is lucrative. At the moment, it is highly profitable (O'Hara (2014)).

5.2 Strategies of HFTs

The millions of investments in state-of-the-art equipment and priority access to the exchanges is only profitable for HFTs if the right strategies and techniques are used. HFTs are experts in finding new loopholes and mechanisms to gain profit from the markets and slower participants. Some of these strategies will be discussed in the following.

5.2.1 Market Making

Market makers provide liquidity to traders which want to act immediately, by simultaneously posting limit orders at the bid price to buy and at the ask price to sell. Thereby, market makers are able to gain the bid-ask-spread from other's urge for immediacy. There is a possibility to suffer losses while trading with informed counterparties, thus, they must ensure that as much information as currently available is represented in its order prices. HFT market makers use their technical advantage in order to adjust their quotes in response to order submissions or cancellations, and price movements in related ETFs or futures contracts.

Hence, HFTs end up cancelling most of their orders they submit. HFT market makers have widely replaced human market makers benefitting from automated algorithms, mainly because they are less likely to suffer losses by trading with informed counterparties, and due to their lower cost structure thanks to high technology (Jones (2013)).

5.2.2 Arbitrage Trading

Along with market making, HFTs benefit from arbitrage if they spot price discrepancy of the same security on two or more different exchanges (or ATS). Thus, they buy the cheaper security and sell the more expensive one. This strategy does not solely focus on single stocks between exchanges, but furthermore e.g. future contracts on indexes in comparison with the comprised securities within the indexes. For example, *S&P 500* futures are traded at the CME, whereas the SPY is traded at almost every exchange, and ATS in the U.S. and foreign trading venues. The prices of these two instruments move equally and open up possibilities for index arbitrage. If prices for the SPY rise while the futures' price stays the same, a HFT would buy the *S&P 500* futures and sell the SPY contract, which results in small profit on the price differential (Jones (2013)).

Price discrepancies between the exchanges and the ATS lead to profit opportunities for the fastest actor in the market, hence large investments in state-of-the-art technology and server co-location pays off eventually for HFTs, as slower actors will not be able to equalize the prices. The same mechanism works if the *S&P 500* futures price rises more than the underlying shares. HFTs would eliminate the mispricing by buying the component stocks in the correct proportions of the *S&P 500* futures. The gains of these transactions often account for just a few cents per security, but due to the high volumes and (almost) risk-free gains, are still very profitable (Jones (2013)).

5.2.3 Statistical Arbitrage

According to Maureen O'Hara (2014), the difference between traditional market making and HFT market making is that HFTs act across and within markets. HFT market making leverages the concept of statistical arbitrage, meaning to evaluate historical correlation patterns in prices of securities to predict future outcomes on recurring events. If e.g. stock B generally increases after the correlated stock A increases, a HFT would sell stock A and buy stock B in order to gain profits.

Berman (2014) describes the complexity of this process with the example of an equity ETF linked to gold called "GLD SPDR". To take advantage of price deviations, a HFT market maker would be quoting in "GLD SPDR" and gold futures. Since there are 13 other traded products on the exchange which are correlated to gold, whose prices can also diverge from gold prices and each other, the HFT has to place bids and asks for all the 91 different pairs, as well in the gold future and possibly in the cash market.

5.2.4 Key Word Algorithms

News reports can easily move the markets. HFTs have learned to benefit from breaking news on new developments of listed companies. Computer programs parse important news reports within milliseconds after their release, and react by buying or selling the corresponding security accordingly. This technique is also known as “Event Arbitrage”. If for instance the words “Roche” and “lawsuit” or “breakthrough drug” appear in the same paragraph of media, algorithms react immediately and benefit from the following reaction of the markets (Jones (2013)). In this respect, the *Wall Street Journal* exposed that news providers grant early access of data streams of a few seconds to paying traders. Thus, these traders possess an advantage against other market participants due to their early information receipt (Pasqualle (2014)).

5.2.5 Order Flow Signals

According to Jones (2013), some HFT’s strategies are based on so-called “Order Flow Signals”: if a large buy order is submitted to an electronic order book, HFTs might anticipate it as a signal that a large trader has substantial positive information, thus, HFTs eventually buy the shares themselves and profit from the corresponding order flow signal.

Understandably, this is of considerable concern for large institutional traders. If such traders attempt to gradually buy a large amount of e.g. Apple shares, HFTs may be able to recognize the buy pattern and anticipate it by buying Apple shares themselves, and thus, drive up the price for the large institutional trader. The HFT gain from buying the shares at a lower price and selling them afterwards at a more expensive price to the large trader. In order to protect themselves from this disadvantage, large traders often break their trades into very small pieces, using an algorithm, or make use of so-called “Dark Pools”³³. Consequently, these trades often end up as hide-and-seek battles between algorithms (Jones (2013)).

This concern has already been witnessed before the age of algorithms and automated trading. Earlier, institutional traders used brokers at the trading post in the corresponding exchanges to prevent a leakage of information. Studies have shown that the overall institutional trading cost has declined over time, although the HFTs grew more prominently (Jones (2013)).

5.2.6 Unethical/Illegal Strategies

HFTs mostly use subtle strategies which provide them with an advantage in the markets due to their speed of information processing and access to that information. Often, this advantage

³³ The term “Dark Pools” is used to describe private exchanges or forums for trading securities. These “Dark Pools” are not accessible for public investors and are often described by their lack of transparency (Picardo (2014)).

is used to manipulate the markets in favor of the actor, using one or several of the following unethical and/or illegal techniques:

- By using the technique of “**Quote Stuffing**”, a HFT tries to provoke congestion in the market by entering massive numbers of orders in the market, and cancelling them within microseconds. This overload of orders in the market leads to a disadvantage for slow traders who do not possess superior access to the markets. Slow traders in this situation do not have clear view on the current status of trading and are impeded in executing trades. Fast traders can benefit from this impediment by executing profitable trades at the slow trader’s expense (Biais and Woolley (2011)).
- According to Biais and Woolley (2011), a second way to manipulate the market is by so-called “**smoking**” techniques (also known as “**layering**”). HFTs post attractive sell or buy orders in the market, long enough to allure slower market participants, but cancel them immediately after the information has been transferred. In this way, slower traders are tricked into paying more or receiving less, respectively, once their orders reach the market. This technique gained large public attention in 2011, as British regulators fined the Canadian trading firm *Swift Trade Inc.* £8 million for manipulating the markets using layering (Bowley (2011)).
- **Spoofing** is another type of market manipulation that involves placing orders in the electronic market without the intention of fulfilling them, but with the goal of triggering other market participants. This is followed by cancelling the original order and the placement of a new order on the opposite side of the market (FINRA (2012)). In 2010, U.S. President Barack Obama signed off the “Dodd-Frank Act” which declared spoofing as an illegal market manipulation (Dodd-Frank (2010)). By spoofing, traders try to alter the prices of futures artificially, by e.g. entering a small bona-fide sell order and a large amount of non-bona-fide buy orders at higher prices. Consequently, the prices are likely to increase and fill the original bona-fide sell order. Once this has occurred, the large buy orders are cancelled and the spoof immediately takes place in reverse with a small bona-fide buy order and several large non-bona-fide sell orders. If the buy order is fulfilled, the sell orders are cancelled again (Aktas (2013)).
In Figure 10 below, the described spoofing activity is visualized by *HTG Capital Partners*³⁴ and illustrates an example of spoofing. The process of spoofing leads to small profits for the manipulating trader at the expense of other market participants from executing the small orders over a period of time (Aktas (2013)). Since the implementation of the “Dodd-Frank Act”, several HFTs have been pled guilty for spoofing, but proofing this misbehavior has shown to be highly complicated for the regulators (see chapter 6 of this thesis).

³⁴ *HTG Capital Partners LLC*, a Chicago-based trading firm, filed a report of 6’960 instances in which it has allegedly been tricked by “spoofers” in the markets between 2013 and 2014 (Leising, Rojanasakul and Pearce (2015))

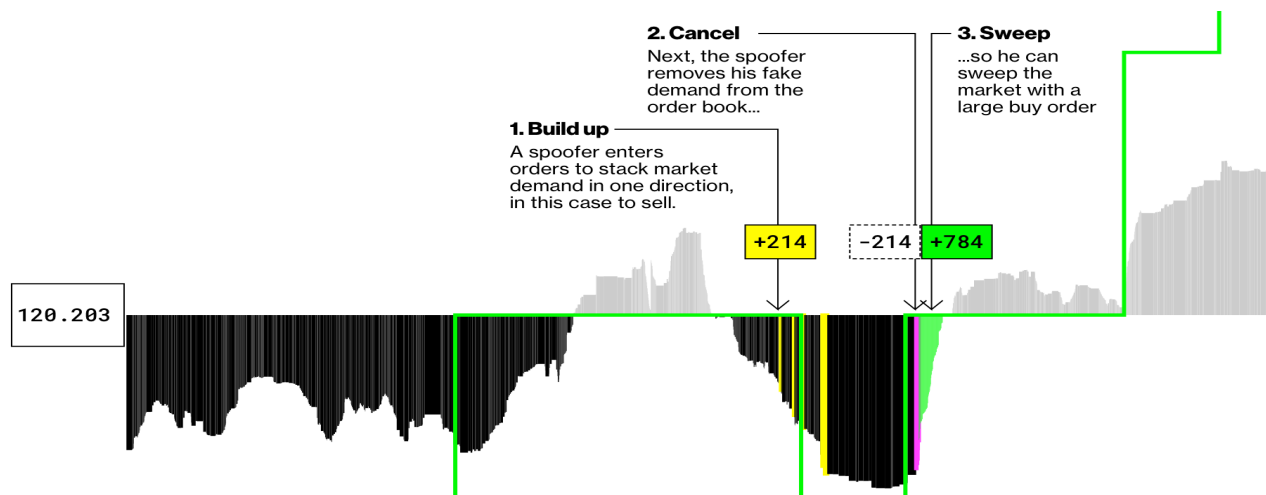


Figure 10: HTG's Spoofing Illustration
Source: Leising, Rojanasakul and Pearce (2015), p.6

5.3 Benefits and Drawbacks for the Markets

Due to the global exposure of the “Flash Crash”, high-frequency trading has gained great attention from the media, regulators, and academic researchers. Despite this, no general agreement has been achieved whether HFTs are more beneficial or harmful for financial markets (O’Hara (2014)). In the following sections, the author will therefore provide arguments for both points of view.

5.3.1 Benefits of High-Frequency Traders in Financial Markets

HFTs act as market makers in financial markets. Thanks to their above mentioned sophisticated technology, discussed premium access to market data, and described lightning-fast reaction through sophisticated algorithms, they are able to **make market prices more efficient**. Several academics concluded that HFTs improve market liquidity, which reflects in narrower spreads and, thus, lower costs of trading for all market participants (Jones (2013)³⁵; Brogaard, Hendershott and Riordan (2014)). Due to the automation of HFTs, costs for intermediation are lower and can be passed on to small investors as narrower bid-ask-spreads, as well as decreased commissions. This leads to lower overall costs of equity capital for firms which trade on financial markets (Jones (2013)). As Malkiel Burton (2009) interprets it, reduced transaction costs enable more people to get involved in financial markets, and the decreased information distribution costs improve the trading environment.

During the last 30 years, trading costs have constantly decreased and this decline seems to accelerate progressively (Angel, Harris and Spatt (2011)). In their field study,

³⁵ In the author’s opinion, it is important to mention that Jones’ (2013) research activity was supported by a grant from *Citadel LLC*, a leading HFT firm.

Malinova, Park and Riordan (2013) provide empirical evidence that since the entrance of HFTs in the *Toronto Stock Exchange*, **retail trading costs have fallen** undeviatingly.

Jovanovic and Menkveld (2011) analyzed the *Chi-X* market's reaction to the entrance of a HFT market maker into the trading of Dutch securities in July 2007. Compared to Belgian stocks, which were not affected by HFTs by that time, the effective bid-ask-spreads of Dutch stocks got 15% narrower and adverse selection reduced by 23%, while volatility remained on a constant level. Hasbrouck and Saar (2013) concluded that **HFTs decrease the short-term volatility, spreads, and depth of order books** in financial markets due to their automation of market activities.

In their investigation about the impact of HFTs on market quality and investor welfare, Rojcek and Ziegler (2016) found that under symmetric information, **market quality is always improved by HFTs**. Under asymmetric information, this is only the case if competition among HFTs is strong enough. Furthermore, investor welfare is not negatively affected by the entrance of HFTs, but reduces the overall welfare of slow speculators³⁶.

5.3.2 Drawbacks of High-Frequency Traders in Financial Markets

The development of financial markets regarding high-frequency trading has not solely had the above mentioned positive consequences, as several academic papers and studies have showed. As mentioned above as one of the main benefits, HFTs improve the markets liquidity. Unfortunately, this statement cannot be generalized. Dalko (2017) e.g. claims that research which concludes the improvement of liquidity through the evolvement of high-frequency trading is based on post-trade data that does not include order cancellations; nor do they explain the strategies used by HFTs. **Most of HFTs' liquidity provision is in the stocks with the greatest capitalization** which already have been overly liquid in the pre-HFT era and, thus, do not depend much on further liquidity (Ye, Yao and Gai (2013)).

Since no one places and cancels orders faster than HFTs, it is difficult to determine where liquidity exists across the fragmented markets. As a consequence, the resulting **uncertainty introduces new profit opportunities for HFTs** within and across markets. Roughly 98% of all orders are cancelled, mostly by HFTs. The midpoint of bid- and ask-price serves as the indicator for the current unconditionally expected price of a traded stock. Due to numerous cancellations, revisions and resubmissions in the order book, quotes consequently are flickering which implicates uncertainty about the actual current price level. This quote volatility results in an **increased execution risk for non-HFT traders** and lower quote informational content (Hasbrouck and Saar (2013)).

HFTs are able to anticipate and respond to changing quotes at the *NASDAQ* within two to three milliseconds (Hasbrouck and Saar (2013)). Market participants often criticize the speed in which new quotes are entered or cancelled, because it creates a deceitful sense of overpriced supply or demand for a security. Golub, Keane and Poon (2012) call these unreal

³⁶ Rojcek and Ziegler (2016) use the term "speculators" to describe traders with no private valuation of an asset, which also includes HFTs.

data **“Fleeting Liquidity”**. Furthermore, it is claimed that HFTs often duplicate their amount of limit orders on several occasions in order to increase their probability of execution which results in an **overestimation of available liquidity** by regular investors (Van Kervel (2012)).

The advantage of speed has been of great consideration for academics while researching in the field of high-frequency trading. Cohen and Szpurch (2011) e.g. discovered that **HFTs are able to front-run³⁷ slow traders**, provided that the fast trader knows the buy or sell-intention of the slower trader. This leads to risk-free gains for the fast trader at the expense of the slower trader. According to Tong (2015), there is significant evidence that **HFTs’ cause an increase in transaction cost for institutional investors**. It is shown empirically that fast liquidity providers use their latency differences against slow liquidity demanders in order to gain minute profits, which accumulates to an estimated \$233 million at slow traders’ expenses per year (McInish and Upson (2012)).

Another drawback of HFTs in financial markets is the **frequent occurrence of market manipulation**. “Predatory” activities such as the above mentioned spoofing, quote stuffing and layering lead to an unethical and illegal advantage over slow traders, and **less market efficiency** (O’Hara (2014)). Not every HFT uses manipulative trading strategies, but they cause way more damage than manipulative trading strategies of non-HFTs (Dalko (2017)). Only market participants with enough financial resources are able to hide their actions or use sophisticated algorithms in order to even the odds of distribution (Pasqualle (2014)).

Biais, Foucault and Moinas (2011) claim that the superior speed of HFTs leads to an **adverse selection** and, again, to a **higher probability of systemic risk events**. The transformation of “human-machine” to “all-machine” driven markets is characterized by innumerable “Black Swan” events³⁸ at ultrafast duration (Johnson et al. (2012)). Market crashes and “Black Swan” events are often provoked by market instability, caused by a certain behavior of HFTs that leads to **episodes of extreme illiquidity and fragility** (Jonas (2013); Kirilenko et al. (2011); Easley, Lopez de Prado and O’Hara (2012); Madhavan (2013)). A study of so-called “Mini Flash Crashes”³⁹ in the US equity market during the four most volatile months between 2006 and 2011, conducted by Golub, Keane and Poon (2012) showed the occurrence of 5’140 “Mini Flash Crashes” (2’760 up-crashes and 2’380 down-crashes). The authors concluded that these events are mainly caused by high-frequency activities caused by “Fleeting Liquidity” and HFTs tending to withdraw liquidity during their occurrence.

Severe market instabilities can also occur if **algorithms lack testing and monitoring**. For example, on August 1, 2012, the *Knight Capital Group*⁴⁰ implemented a new

³⁷ The term “Front-Running” describes the technique of a fast trader using his speed advantage in order to clear the limit order book before a slow trader is able to do so. Thereafter, the front-runner sells his stocks to the slow trader at a higher best price, and gains a risk-free profit (Cohen and Szpurch (2011)).

³⁸ Finance professor Nassim Nicholas Taleb first characterized the term of a “Black Swan” event. It is described as an occurrence that deviates beyond from what is normally expected (Investopedia (2017b)).

³⁹ “Mini Flash Crashes” are divided into upwards and downwards crashes. *Nanex LLC* identified this occurrence first (Nanex (2010c)), see <http://www.nanex.net/FlashCrashEquities/FlashCrashAnalysisEquities.html> for the exact definition.

⁴⁰ *Knight Capital Group* is one of the largest market makers in U.S. equities (Jones (2013)).

algorithm which was apparently not sufficiently tested in advance. The algorithm went berserk and accumulated large amounts of 148 NYSE-listed stocks over 45 minutes. It had to be shut down manually and caused, including the cost of liquidating the positions, losses for the *Knight Capital Group* of \$440 million and resulted in 75 percent of *Knight's* equity being erased (Jones (2013)).

Biais, Foucault and Moinas (2015) furthermore established, that the amount of **investments into high-frequency trading technology may surpasses to its socially optimal level**. The evolved and above mentioned arms race among HFTs for every millisecond in order to (survive and) stay competitive is consuming billions of dollars. It is estimated that a three milliseconds reduction in communication time between the two finance hubs Chicago and New York is worth about \$500 million (Laughlin, Aquirre and Grundfest (2012)). Faster market participants require adaption of the exchanges to be able to control and monitor the trading activity efficiently. Due to so-called "**High-Frequency Spam**"⁴¹, exchanges may be overwhelmed and the insufficient capacity can cause delays in price data reporting (Hunsader (2011)). To prevent this, exchanges need to improve their infrastructure with large investments, although the effective trading volume did not increase likewise (Rojcek and Ziegler (2016)).

6. Regulator's Responses to the "Flash Crash"

The before discussed and analyzed "Flash Crash" was a clear indicator that the U.S. financial markets are flawed inter alia through the misbehavior of HFTs (Nanex (2014)). The regulators and exchanges consequently implemented specific regulations and supervisory measures in order to prevent recurrence of equivalent events and ensure the orderly functioning of financial markets (Born et al. (2011)). Further, researchers have analyzed and proposed additional regulations in their research papers (Easley, Lopez de Prado and O'Hara (2010); Nanex (2014); Rojcek and Ziegler (2016)) Some of those regulations and measures will be presented in the following.

6.1 Implemented Regulations and Measures since May 6, 2010

After the "Flash Crash", the SEC's and CFTC's advisory committee highlighted several issues which had to be addressed in response to the market events of May 6, 2010. On February 18th, 2011, the white paper "Recommendations Regarding Regulatory Response to the Market Events of May 6, 2010" was released. It consists of recommendations regarding the issues in order to improve market stability and supervision capabilities of the regulators (Born et al. (2011)). In the following, the most relevant and already implemented regulations and measures will be presented.

⁴¹ "High-Frequency Spam" refers to cancelling and changing quotes rapidly, without a change in price or the intention of execution (Hunsader (2011)).

6.1.1 Single Stock “Circuit Breaker”

As a consequence of the 1987’s “Black Monday”⁴² and in order to restore investors’ confidence, market-wide “Circuit Breakers” have been introduced to stock and commodities exchanges. When prices fall too far in a short time period, these “Circuit Breakers” temporarily restrict the trade in stocks, options, and index futures. Market-wide “Circuit Breakers” are focused on price drops in indexes (Cui and Gozluku (2016)).

Around one month after the “Flash Crash” of 2010, the SEC released a statement to “permit [FINRA] trading halts due to extraordinary market volatility” (SEC (2010d), p.1) first applied to the *S&P 500* stocks. These stops of five-minute trading halts occur if an individual stock moves by more than 10% within the last five minutes between 9:45 a.m. and 3:35 p.m. (SEC (2010e)). During the “Flash Crash”, several individual stocks traded at unrealistic prices, far away from their pre-crash price. The SEC broke all transactions which traded 60% above or below their initial price (SEC (2010b)). Thanks to the single-stock “Circuit Breakers”, market participants have the possibility to process available information thoroughly and provide liquidity in order to stabilize the situation. The purpose of the pause is to limit extreme adverse selection through market makers (Jones (2013)).

On August 1, 2012, as *Knight Capital’s* rogue algorithm hit the market at 9:30 a.m., single-stock “Circuit Breakers” did not activate until 9:45 a.m. The algorithm bought large amounts of 148 *S&P 500* listed shares, but price changes were often not sharp or fast enough to trigger “Circuit Breakers”. Between 9:45 a.m. and 10:15 a.m., trading of only five shares was halted for a five-minute period. Although “Circuit Breakers” were active, *Knight Capital* has suffered severe losses during that day (Jones (2013)).

6.1.2 Prohibition of Unfiltered Market Access

On November 3, 2010, SEC Chairman Mary Schapiro announced a rule to prevent so-called “Naked Access” to the markets. It requires “brokers and dealers to have risk controls in place before providing their customers with access to the market” (SEC (2010c), 1st paragraph). Broker-dealers used to provide certain customers with a direct access to an exchange or ATS. These market participants, mostly HFTs and institutions, use the so-called “Market Participant Identifier” (MPID) of the broker-dealers to trade by themselves directly at the exchange or ATS, without being pre-screened by the broker-dealer. This practice is called “Naked Access” and has been prohibited through the new SEC rule. Broker-dealers are now required to “implement certain risk management controls and supervisory procedures to manage the various risks” (SEC (2010c), 4th paragraph). It decreases the likelihood that market participants enter erroneous trades or fail to satisfy regulatory requirements (SEC (2010c)).

⁴² On October 19, 1987, stock markets around the world crashed and major indexes in the U.S. dropped by 30% or more. The DJIA plunged 508 points, or 22.6 percent, which represented its largest single-day drop up to that day (Itskevich (2002)).

6.1.3 Prohibition of Market Maker Stub Quotes

Stub quotes are buy or sell orders entered by market makers so far away from the NBBO that there is no interest of fulfilling them. During the “Flash Crash” of 2:45, stub quotes played a significant role, as buying offers of \$0.01 and selling offers of \$100'000 were entered for single stocks. Since these were the only offers in the order book, on some occasions they were filled and exacerbated the price decline during the crash even more. At the end of the day, regulators had to reversed trades which have been executed at unrealistic prices (SEC (2010b)). As of December 6th 2010, the SEC implemented a new rule that requires market makers to enter quotes of no more than 8% deviation in price to the NBBO (SEC (2010d)).

6.1.4 Consolidated Audit Trail (CAT)

As described in section 3.2.2., the implementation of the CAT is in progress. After six years of negotiations, the SEC finally approved the creation of this powerful market surveillance tool. Through it, regulators will gain more insight in real-time data in order to conduct research more efficiently, identify misbehavior in the financial markets, and reconstruct market events. The gained data sets through the CAT system will contain identities of traders in the equity and option markets. CAT will, once completed, be able to process over 58 billion quotes and orders per day, making it the world's largest storehouse for securities data (Bullock and Stafford (2016)). The new capacity is sufficient to process an estimated 25 billion linked records in just an hour. Thus, it is specified to cope with the enormous amount of data which are generally just cancelled within fractions of a second (Palmer, Sherman, Wang and Just (2015)). An estimated amount of 21'000 petabytes (=21'000'000 terabytes) of data will be collected within the first five years after the establishment of the system (Bullock and Stafford (2016)).

The project's costs attracted public attention. Its implementation costs are estimated at \$2.4 billion initially and continuous costs of \$1.7 billion per year. Big parts will be paid by broker-dealers who will pass the costs on to their customers; the investors (Bullock and Stafford (2016)).

6.1.5 Cancellation Fees

Another enforced regulation in order to improve market quality in the presence of HFTs is a “Cancellation Fee” if a certain ratio of quotes-to-trade is exceeded. On July 2, 2012, the *NASDAQ Stock Market Exchange* implemented a rule change which calculates the ratio of total submitted orders and wholly or partly executed orders. If the ratio exceeds the value of 100, a fee is imposed on the market participant. This rule change is a consequence of rising concerns about unethical trading practices, such as spoofing and quote stuffing (SEC (2012b)).

Hasbrouck and Saar (2013) found that approximately 98% of all orders submitted by HFTs are cancelled. Adaption of limit orders to new information in the financial markets cause a high amount of order cancellations by liquidity providing market makers (including HFTs). In order to establish a true market price for securities, it is essential for HFTs to work with a high number of order cancellations (Blocher et al. (2016)). Thus, Malinova, Park and Riordan (2013) show in an empirical study that after the introduction of “Cancellation Fee” at the *Toronto Stock Exchange*, bid-ask-spreads increased by 9% while quoting activity declined by 30%. Therefore, “Cancellation Fee” may affect the frequency of illegal trading practices, but also increases order execution time significantly due to reduced liquidity in the markets. In general, liquidity provision becomes costlier for all traders in a financial market with “Cancellation Fee” (Rojcek and Ziegler (2016)).

6.1.6 Limit Up-Limit Down Plan

In order to address severe price swings, the SEC approved a new regulation called “Limit Up-Limit Down” (LULD) plan on May 31, 2013. Essentially, it is designed to prevent trades outside of rolling price bands which are above or below a certain percentage level of the average reference price over the last five minutes. The LULD plan denies orders outside of these price bands to appear to the market. Human errors or attempts to manipulate the markets often cause these “unacceptable” orders. Thus, the plan allows to efficiently prevent erroneous or manipulative orders from execution, which could lead to severe price movements in a short time interval (NYSE (2015)). Dalko (2017) criticizes that the LULD is not able to restrict other manipulative HFT activities, such as spoofing. Furthermore, she claims that the plan does not consider reasonable orders placed and immediately cancelled again by HFTs. Therefore, the regulation is able to limit volatile price swings, but contains loopholes which will not block a market crash from occurring.

6.1.7 Compliance and Integrity Improvement

On November 19, 2014, the SEC approved a regulation to improve compliance and integrity in the financial markets. This set of rules became effective on February 3, 2015, and obliges self-regulatory organizations (e.g. exchanges), certain ATS, and clearing agencies to have extensive policies and procedures for their technological systems in place. Additionally, the rules include a framework to help efficiently take action if system issues occur and how to inform the SEC and other market participants about these issues. The implemented rules engage entities to carefully design, develop, control, surveil, and maintain their underlying technological systems which are essential for their operations (SEC (2015)).

6.1.8 CFTC: Disruptive Trade Practices

The CFTC implemented released new rulebook on May 28, 2013, which disposes a guide on disruptive trading practices. It defines and prohibits (under the “Dodd-Frank Act”) any practices that violate bids and offers, demonstrate unethical behavior regarding orderly execution during the closing period, or are of character of spoofing (CFTC (2013)).

6.2 Suggested Regulations

Academics not only focused on explaining and analyzing the market events of May 6, 2010 but also on suggestions for regulations in order to make markets more stable and prevent similar occurrences (Rojcek and Ziegler (2016); Easley, Lopez de Prado and O’Hara (2010)). Often, researchers found HFTs as responsible for destabilizing the market and increasing the probability of extreme situations such as the “Flash Crash” (see section 5.3.2.). Thus, these authors demand more regulation and supervision for HFTs in financial markets. Some of these suggested regulations will be discussed in the following.

6.2.1 VPIN Measure

Easley, Lopez de Prado and O’Hara (2010) suggest the use of the VPIN, a measure to estimate the order flow toxicity (see section 3.4.2.), in order to prevent future recurrences of the “Flash Crash”. The measure serves as tool to estimate the risk of a liquidity-induced market crash. Regulators could use the VPIN measure as a warning sign for the case that order flow toxicity reaches similar levels than those shortly before the “Flash Crash” occurred. Hence, market regulators could interact by slowing down or even halting market activity in order to prevent or mitigate the upcoming crash. High-frequency market makers might remain active as liquidity providers in the market if they use the VPIN measure and thus prevent a liquidity-induced market crash.

6.2.2 Financial Transaction Tax

The implementation of a “Financial Transaction Tax” should lead to higher revenues for regulators and lower the overall amount of speculative trading in financial markets (Jones (2013)). Rojcek and Ziegler (2016) measured the impact of different regulations and identified the “Financial Transaction Tax” as the least harmful for welfare while not significantly diminishing market quality. Under a “Financial Transaction Tax”, market participants must pay a certain percentage of the value of their transactions to the regulator. A transaction tax of 0.0034% is already imposed in the U.S., as well as in roughly 40 other countries around the globe. However, the transaction tax imposed in the U.S. is exceptionally low compared with e.g. Italy (0.12%), Switzerland (0.15%), or France (0.2%). In addition, stamp duty is often levied, but exchanges usually grant rebates for liquidity providers.

However, Ait-Sahalia and Saglam (2014) found that the consequence of “Financial Transaction Taxes” is a reduction of welfare by approximately the amount of taxes imposed, and a decrease of HFTs quoting activity which leads to lower transaction volumes.

6.2.3 Cancellation Fees

As mentioned in section 6.1.5, some exchanges have already implemented “Cancellation Fees” for the excess of a certain quote-to-trade ratio. The *NASDAQ Stock Market Exchange* has e.g. introduced a “Cancellation Fees” after exceeding a quote-to-trade ratio of 100. Gao, Mizrach and Ozturk (2015), who studied U.S. markets between 2008 and 2013, concluded that HFTs’ common practice of rapidly submitting and cancelling quotes leads to higher volatility in prices and increased bid-ask spreads, and thus, is harmful for market quality. Hence, the introduction of “Cancellation Fees” on every exchange might be beneficial for overall market quality and reduce unethical trading strategies, such as quote stuffing and spoofing. In this respect, the *Eurex Exchange* has imposed “Cancellation Fees” for market participants exceeding a quote-to-trade ratio of five (Rojcek and Ziegler (2016)).

6.2.4 Minimum Resting Time

In the U.S., regulators have introduced rules which oblige market participants to solely submit bona fide quotes⁴³ (SEC (2011a; 2011b)). These rules will strengthen FINRA’s ability to prosecute manipulative behavior in the financial markets, but do not include exact specifications on how misconduct can be verified. On October 21, 2013, the *European Parliament* agreed to impose a minimum order resting time of 500 milliseconds under the “*Markets in Financial Instruments Directive II*”⁴⁴ (MiFID II) (Stafford (2013)). Italy has implemented a fee for the cancellation of orders within 500 milliseconds of 0.02%. These changes shall result in dismantling the ongoing arms race between HFTs and adverse selection towards slower market participants (Rojcek and Ziegler (2016)). On the other hand, Harris (2013) concludes that a minimum order resting time will cause more losses for liquidity-providing HFTs, and thus, higher transaction costs for all investors.

6.2.5 Speed Bumps

Similar to minimum order resting time, so-called “Speed Bumps” aim to slow down the arms race of HFTs and represent a latency restriction. In contrast to the minimum resting order, which restricts the modification of the order during a certain time, “Speed Bumps” create a

⁴³ A quote is derived as bona fide if the initial intention of the issuer is to fulfill the quote. If the order is placed in the market with the intention of cancelling it again, it is not derived as bona fide (SEC (2011b)).

⁴⁴ The Markets in Financial Instruments Directive (MiFID) is a regulatory authority, which oversees firms that provide financial services and trading venues for financial instruments (FCA (2016)).

time delay between the submission of the order and the time it arrives at the order book (Rojcek and Ziegler (2016)).

The most prominent example for “Speed Bumps” is *The Investors’ Exchange* (IEX) which was approved by the SEC on June 17, 2016, to become an official national securities exchange (SEC (2016b)). Traders on the IEX experience a quote delay of 700 microseconds through the IEX’ “Point of Presence” which is located 38 miles away from the IEX matching engine. An order is transmitted within 350 microseconds from the “Point of Presence” to the matching engine and requires an additional 350 microseconds to transfer the information to the members of the exchange. This intentional delay prevents individual IEX members from having speed advantages over other traders on the exchange (Levine (2015)).

However, Rojcek and Ziegler (2016) conclude that under asymmetric information, “Speed Bumps” decrease spreads and increase order book depth. They therefore worsen price discovery and lead to an increase in execution time for all traders.

7. Spoofing (Detection)

Throughout the past years, regulatory institutions intensively expanded the rulebooks and their capacity for detection of manipulative trading behavior. The prohibition and prevention of spoofing (also known as “Quote Stuffing”) became their primary focus. Often in cooperation with third parties, regulators are able to verify illegal spoofing activities and prosecute the culprits.

7.1 Regulators

Under the “Dodd-Frank Act”, signed off by President Barrack H. Obama in July 2010, trading practices such as layering, quote stuffing and spoofing were classified as illegal. On Monday 13th September 2010, the FINRA sanctioned the first high-frequency trading firm for one of these manipulative trading strategies. New York-based Trillium Broker Services was fined \$2.26 million for entering numerous non-bona fide orders to generate artificial buy-/sell-side pressure in the NASDAQ and NYSE Arca, which is known as quote stuffing (FINRA (2010)).

Roughly four years later, the first criminal prosecution on spoofing was successfully closed against Michael Coscia, owner of Panther Energy Trading. He was declared guilty for spoofing in the futures market in during several months in 2011. Michael Coscia used an algorithm, designed to submit small bona fide sell orders, followed by large non-bona fide buy orders to create artificial buying-side pressure. In consequence of the large buy orders, other market participants anticipated and filled the small bona fide sell order. Coscia’s algorithm immediately cancelled all outstanding buy orders, reversely placed a bona fide buy order and large non-bona fide sell orders. Through his illegal trading practice, Coscia allegedly profited approximately \$1.5 million (U.S.A. vs. Coscia (2015)).

This pronouncement of judgment is seen as an important step in trading crime prevention. Since the “Flash Crash”, regulators have announced more focus on improving their data collection and analysis technology (with e.g. the consolidated audit trail) to provide better market surveillance (SEC (2016a)). In the months following the conviction of Coscia, further trials regarding spoofing were held against inter alia Navinder Sarao⁴⁵, Igor Oystacher (CFTC (2016)) and Da Vinci Invest Ltd (FCA (2015)).

Although several successes have been achieved by the regulators, it is clear their spoofing detection technology is working to capacity, especially in markets like the E-Mini, where huge amounts of orders and cancellations are submitted. While the CFTC, CME Group and the Intercontinental Exchange (ICE) possess the ability to identify the customers behind futures and swaps orders, it is questionable why not more actions against spoofing are taken. The SEC and FINRA do not have the ability to directly identify customers, which is an aggravating factor in spoofing detection. In order to identify customers behind quotes, the SEC and FINRA need to request the corresponding data from the executing broker or clearing firm. The main challenge for regulators is to justify, whether a market participant solely wanted to change his position or if he attempted to intentionally manipulate the prices on the market (Hess Legal Counsel (2015)).

Regulators face the challenge of collecting all relevant data across markets and asset classes, being able to analyze it in a timely manner (desirably in real-time) and strictly enforcing the law if misconducts are observed. As mentioned in chapter 6.1.2, the SEC implied new rules for brokers obliging them to conduct risk controls and appropriate surveillance of their customers, instead of granting “naked access”. Brokers can be held responsible, if they grant market access without ensure the required risk controls to prevent market manipulations and detect spoofers (SEC (2010c)).

In April 2016, FINRA has issued so-called cross-market equities supervision “report cards” in order to support firms in identifying and preventing layering and spoofing activities. If the regulator identifies potential manipulative activities from firms or entities the firm is granting market access to, report cards are sent to the corresponding firm. These report cards contain a detailed summary of potential spoofing or layering activities. The ulterior motive of this reporting is not to accuse the firm, but motivate it to review the irregularities. FINRA will continue to surveil the corresponding firm to check whether the potentially unethical trading behavior occurs again and if appropriate, take enforcement action (FINRA (2016)).

7.2 Third Parties

A positive trend is the increasing involvement of third parties in spoofing detection (and manipulative behavior in general). If market participants witness certain market behavior which looks suspiciously like spoofing, they more frequently report or publicize such events to regulators. Most prominent is the independent market research firm Nanex, which inter

⁴⁵ See section 3.3.

alia has specialized on illegal trading strategies of HFTs. On several occasions, Nanex has publicized its own research findings about spoofing occurrences and addressed it to the SEC and CFTC, which led to further investigations by the regulators⁴⁶ (Nanex (2015)).

Financial Tracking Technologies LLC (FTT) is the leading firm for automated compliance solutions and serves hedge funds, money managers, public companies and governmental regulators (FTT (2017)). On January 29th 2016, FTT has announced the launch of a new compliance tool which automatically detects spoofing activities in the financial markets. It recognized artificial price pushing caused by market manipulations and allows FTTs clients to avoid being affected by the negative impact of spoofers (Businesswire (2016)). The author of this thesis argues, that FTTs automated spoofing detection module lowers HFTs incentive to commit market manipulation, thus increases fairness in the financial markets for non-HFT market participants. Additionally, regulators could utilize the service to convict spoofers and simultaneously focus their surveillance capacity on other types of fraudulent behavior.

On some occasions, HFTs raise the regulators' awareness of other HFTs spoofing activity, typically if the accusing HFT has suffered from it. For example, HTG Capital Partners has filed a lawsuit in October 2015 against another HFT for 6'960 instances of spoofing over a period of 20 months (HTG vs. John Doe (2016)). The spoofing of U.S. Treasury futures took place on the *Board of Trade of the City of Chicago* (CBOT), where trading is anonymous and only the CBOT and CME (CBOT's parent company) have access to customer identification. The accused spoofer's identification will stay confidential during the ongoing trial and the debate should be settled within the exchange, according to the CME Group (Zanki (2015)).

7.3 Suggestions for automated Spoofing Detection

As mentioned before, the regulators lack of capacity and tools to detect and prosecute spoofing activity in the financial markets efficiently. Therefore, I will suggest further steps on how to improve the automated spoofing detection of the regulators.

Firstly, regulators must improve their tools for data gathering. The 2016 approved implementation of the Consolidated Audit Trail (CAT) allows the SEC to collect data from all quotes in the equities and options markets, which the SEC oversees, on a daily basis. Customer identification is of great importance in market surveillance, thus the CAT obligates broker-dealers and ATS to report information including cancellations, modification, prices and customer's identification (SEC (2016a)). A resulting problem is that broker-dealers may identify their customers using different names or some with the same name. The use of unequivocal identification numbers such as Social Security Numbers and Employee

⁴⁶ See <http://www.nanex.net/aqck2/4700.html> for further spoofing cases documented by Nanex.

Identification Numbers for trades from legal entities, throughout all broker-dealers, could be beneficial.

Since spoofing takes place in other markets than equities and options markets as well (Nanex (2014)), the information reporting should be extended to futures, which are often targeted by spoofers. This requires full cooperation of the corresponding regulators, such as e.g. the CFTC and the national exchanges.

Regarding the recognition of spoofing in the markets, I suggest the SEC to first analyze datasets of previous spoofing cases and create a “basic spoofing pattern”. This analysis allows to develop a general sequence of order submissions which in the past have led to artificial price adjustments. This basic spoofing pattern will most likely consist of a small bona fide sell (or buy) order, followed by the submission of large non-bona fide orders on the opposite side of the market order book far away from the best price. Note, that the small bona fide order represents the best price in the order book, so that it will be hit first if price changes occur. The markets react due to the large orders, which create an artificial buy-/sell-pressure. Simultaneously, as the market price increases and the bona fide order is filled, the large non-bona fide orders will be cancelled. This process takes place in reverse immediately after the cancellation of the large orders (U.S.A. vs. Coscia (2015)). Nanex illustrates this sequence of trading behavior in figure 11 below on the basis of spoofing activities in the gold futures market.

The illustration shows clearly price movements caused by the submission of large orders. Typical for this technique is the W-movement of the price, which is a consequence of the reversal spoofing activity after the first small bona fide order was filled and the large orders cancelled. The red circles in figure 11 show a large number of buy orders flooding the markets and consequently cause a price incline while almost simultaneously the buy orders are cancelled again (Nanex (2015)).

Regulators need to find these typical patterns, which often are correlated with a high cancellation rate of large orders (Ye, Yao and Jiading (2013)). Thanks to the included customer identification, there is an increased chance of determine the manipulating market participant.

Additional to the pattern recognition, regulators should consider combining their extensive data inflow with the automated spoofing detection module of Financial Tracking Technologies in order to ensure double real-time security and surveillance.

Other similar spoofing scenarios can be identified through further research conducted by the regulators and implemented into their surveillance engines. For example, if more than one market participant is actively involved in spoofing activities in the same market, which could be targeted in further investigation and development of surveillance technology.

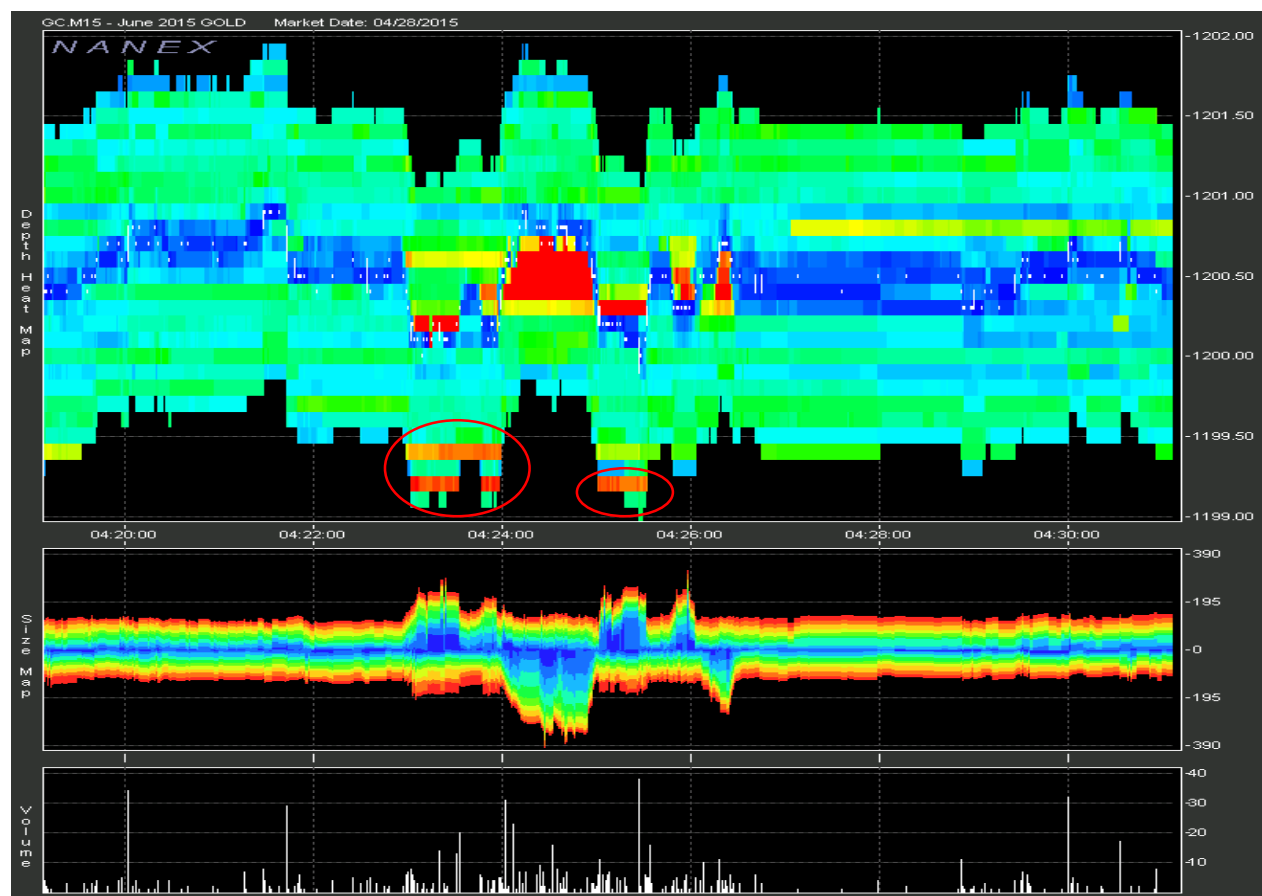


Figure 11: Illustration of Spoofing Activity in Gold Futures Market
 Source: Nanex (2015)

Spoofing does usually not occur uniquely, but often several dozen or hundred times in a short time period (Nanex (2015)). Every time alleged spoofing activity is detected, the regulators should get an automatically generated internal alert to surveil the corresponding market participant's actions more closely. In case of several recurrences of this illegal behavior, the regulating entity should immediately address the broker-dealer and/or customer directly with an activity report, similar to the FINRA's "report cards" (FINRA (2016)), and a warning. The improved presence of regulator's surveillance in the markets will have targeted intimidation effects on spoofers and thus hopefully lead to substantial reduction of illegal market manipulations.

The activity report, which includes a warning, should be stored in the regulator's servers in order to rely on if further misconducts are caused by the individual. The first warning should not be connected to a criminal proceeding. Perhaps an algorithm of the HFT has unintentionally operated in a similar way as spoofers, thus the HFT has the chance to adjust his codes. If the spoofer triggers another alert, he will receive the last and final warning. Legal investigation procedures will be initiated, if the HFT despite previous warnings still continues to manipulate the fair market price.

Through the successful implementation of the abovementioned functions will regulators be able to further increase their automated spoofing detection capacity and ensure fairer markets for all participants.

8. Conclusion

This thesis provides an extensive overview of the different theories about the cause of the “Flash Crash” of May 6th 2010. It concludes that the official report, released by the SEC and CFTC (2010b) is partially flawed and further investigation was needed. Their main claim, that an erroneous sell algorithm went berserk and most significantly triggered the following market events was proven to a certain extent wrong by academics and independent researchers (Nanex (2010a); Easley, Lopez de Prado and O’Hara (2010); Menkveld and Yueshen (2016); Aldrich, Grundfest and Laughlin (2016)). Liquidity withdrawal of market making HFTs turned out to be more detrimental as assumed in advance (Menkveld and Yueshen (2016)). Hence, in order to improve market stability and prevent similar market events in the future, regulations have been implemented to impede HFTs from affecting financial markets negatively.

Another allegation of the regulators referred to Navinder Sarao, who has been accused in 2015 of significantly contributing to the “Flash Crash” through his spoofing activities on May 6th and the previous days (CFTC (2015)). Although it cannot be denied, that he utilized illegal trading strategies, research has showed, that he neither intentionally created the market destabilization nor significantly triggered the “Flash Crash” (Aldrich, Grundfest and Laughlin (2016)).

In conclusion, the already highly volatile markets, due to the tensions of the Greek debt crisis, were further weakened by extraordinary trading volume and low liquidity provision (SEC (2010b)). Price reporting delays of exchanges and ATSS additionally increased market uncertainty. The level of asymmetric information and toxic order flows reached an all-time high previous to the market crash, resulting in reluctance of liquidity providers (Easley, Lopez de Prado and O’Hara (2010)). Market makers withdrew their liquidity when it was needed most to cushion the high sell pressure caused by W&R’s algorithm and HFTs, which traded aggressively with each other in order to decrease their inventories (Kirilenko et al. (2011); Menkveld and Yueshen (2016)). Consequently, their trading exacerbated the price decline and created a “hot potato”-effect (SEC (2010b)). Subsequently, a trading halt was triggered, which calmed the markets, led to the disappearance of price oscillations and allowed liquidity to return (Aldrich, Grundfest and Laughlin (2016)).

Overall, the “Flash Crash” of May 6, 2010 was the result of the interplay of many factors at the same time, thus not a single factor can be held solely responsible. Regulators realized that their rules and surveillance technology had to be improved, which led to the implementation of new regulations and the planning of the CAT (SEC (2016a)). Academics request for further regulations, such as speed bumps, minimum resting time, cancellation fees, transaction taxes and VPIN measurement tool to prevent HFTs from unethical behavior and the causation of future flash crashes.

Lastly, spoofing detection capabilities have increased during the last years, but still need further improvement. This thesis suggests regulators to upgrade their surveillance tool, use basic spoofing patterns and support of third parties, like FTT. Through enhanced

collaboration between regulators and with exchanges, a more efficient and effective spoofing detection tool can be created, which allows to surveil all markets at once. On the long term, this evolvement will rebuild market participants' trust in the capabilities of the regulators.

References

- Acharya, Viral V., 2012, *The Dodd-Frank Act and Basel III: Intentions, Unintended Consequences, and Lessons for Emerging Markets*, Asian Development Bank Institute, October 28th.
- Aït-Sahalia, Yacine and Mehmet Saglam, 2014, High frequency traders: Taking advantage of speed, Working paper, Princeton University and University of Cincinnati.
- Aktas, Deniz, 2013, Spoofing, *Review of Banking and Financial Law* 33, 89-98.
- Aldrich, Eric M., Joseph Grundfest and Gregory Laughlin, 2016, The Flash Crash: A New Deconstruction, Working paper, University of California, Santa Cruz, and Stanford University Law School.
- Aldridge, Irene, 2010, What is High-Frequency Trading, Afterall?, *Huffington Post*, July 8. Retrieved from: http://www.huffingtonpost.com/irene-aldridge/what-is-high-frequency-tr_b_639203.html, December 16, 2016.
- Angel, James J., Lawrence E. Harris, Chester S. Spatt, 2011, Equity Trading in the 21st Century, *Quarterly Journal of Finance* 1, 1–53.
- Bates, John, 2015, Post Flash Crash, Regulators Still Use Bicycles To Catch Ferraris: Blaming the Flash Crash on a UK man who lives with his parents is like blaming lightning for starting a fire, *Traders Magazine Online News*, April 24. Retrieved from: <http://www.tradersmagazine.com/news/technology/post-flash-crash-regulators-still-use-bicycles-to-catch-ferraris-113762-1.html>, December 12, 2016.
- Berman, Gregg E., 2014, What Drives the Complexity and Speed of our Markets?, April 15. Retrieved from: <https://www.sec.gov/News/Speech/Detail/Speech/1370541505819>, December 21, 2016.
- Bethel, Wes and Leinweber, David and Ruebel, Oliver and Wu, Kesheng, 2011, Federal Market Information Technology in the Post Flash Crash Era: Roles for Supercomputing, September 25, Lawrence Berkeley National Laboratory.
- Biais, Bruno and Paul Woolley, 2011, High frequency trading, Unpublished Working paper, University of Toulouse and London School of Economics.
- Biais, Bruno, Thierry Foucault and Sophie Moinas, 2011, Equilibrium high-frequency trading, Unpublished Working paper, HEC Paris, Centre for Economic Policy Research and University of Toulouse.
- Biais, Bruno, Thierry Foucault and Sophie Moinas, 2015, Equilibrium Fast Trading, *Journal of Financial Economics* 116, 292-313.
- Blocher, Jesse, Ricky Alyn Cooper, Jonathan Seddon and Ben Van Vliet, 2016, Phantom Liquidity and High Frequency Quoting, *Journal of Trading* 11, 6-15.

- Bloomberg, 2017, SPDR S&P 500 ETF Trust. Retrieved from:
<https://www.bloomberg.com/quote/SPY:US>, January 27, 2017.
- Born, Brooksley E., John J. Brennan, Robert F. Engle, Richard G. Ketchum, Maureen O'Hara, Susan M. Philips, David S. Ruder, and Joseph E. Stiglitz, 2011, Recommendations Regarding Regulatory Responses to the Market Events of May 6, 2010, Report, CFTC.
- Bowley, Graham, 2011, Clamping Down on Rapid Trades in Stock Market, *The New York Times*, October 8. Retrieved from:
<http://www.nytimes.com/2011/10/09/business/clamping-down-on-rapid-trades-in-stock-market.html?src=un&feedurl=http%3A%2F%2Fjson8.nytimes.com%2Fpages%2Fbusiness%2Findex.jsonp>, November 8, 2016.
- Brogaard, Jonathan, Terrence J. Hendershott and Ryan Riordan, 2014, High-frequency trading and price discovery. *The Review of Financial Studies* 27, 2267-2306.
- Brush, Silla, Tom Schoenberg and Suzi Ring, 2015, How a Mystery Trader With an Algorithm May Have Caused the Flash Crash, *Bloomberg Technology*, April 22. Retrieved from:
<https://www.bloomberg.com/news/articles/2015-04-22/mystery-trader-armed-with-algorithms-rewrites-flash-crash-story>, November 13, 2016.
- Bullock, Nicole and Philip Stafford, 2016, SEC approves vast surveillance system for stock market, *New York Times*, November 16. Retrieved from:
<https://www.ft.com/content/2a428156-ab1e-11e6-9cb3-bb8207902122>, January 2, 2017.
- Burton, Malkiel, 2009, Opinion: High-frequency trading is a natural part of market evolution, *Financial Times*, December 19. Retrieved from:
<https://www.ft.com/content/1513400e-e8cf-11de-a756-00144feab49a#axzz3eFoSJDD4>, December 12, 2016.
- Businesswire*, 2016, Financial Tracking Launches New Compliance Module for Spoofing Detection, January 29th. Retrieved from:
<http://www.businesswire.com/news/home/20160129005737/en/Financial-Tracking-Launches-Compliance-Module-Spoofing-Detection>, November 28, 2016.
- Chevreau, Jonathan, 2010, P&G error started rout but money managers expect "slow but upwards equity markets to continue, *Financial Post*, May 7. Retrieved from:
<http://business.financialpost.com/uncategorized/pg-error-started-rout-but-money-managers-expect-slow-but-upwards-equity-markets-to-continue>, December 6, 2016.
- Chicago Mercantile Exchange Group (CME Group), 2010, *CME Group Statement on the Joint CFTC/SEC Report Regarding the Events of May 6*, October 1. Retrieved from:
<http://investor.cmegroup.com/investor-relations/releasedetail.cfm?ReleaseID=513388>, December 24, 2016.

- Chicago Mercantile Exchange Group (CME Group), 2016, *CME Globex Reference Guide*, September 30.
- Chicago Mercantile Exchange Group (CME Group), 2017, About E-mini S&P 500, Retrieved from: <http://www.cmegroup.com/trading/equity-index/us-index/e-mini-sandp500.html>, January 17, 2017.
- Chicago Mercantile Exchange Group (CME Group), 2017, Stop Logic. Retrieved from: <http://www.cmegroup.com/confluence/display/EPICSANDBOX/Stop+Logic>, January 17, 2017.
- Cohen, Samuel N. and Lukasz Szpruch, 2011, A Limit Order Book for Latency Arbitrage, Working paper, University of Oxford.
- Commodities Futures Trading Commission (CFTC) vs. Sarao, 2016, Case No. 15 cv 3398, November 8.
- Cui, Bei and Arie Eskenazi Gozluklu, 2016, Intraday Rallies and Crashes: Spillovers of Trading Halts, *International Journal of Finance & Economics* 21, 472-501.
- Dakers, Marion, 2016, 'Flash crash' trader takes extradition case to appeal, *The Telegraph*, September 10. Retrieved from: <http://www.telegraph.co.uk/business/2016/09/10/flash-crash-trader-takes-extradition-case-to-appeal/>, November 19, 2016.
- Dalko, Viktoria, 2017, Limit Up-Limit Down: An effective response to the "Flash Crash"?, *Journal of Financial Regulation and Compliance* 24, 420-429.
- Dash, Mike, 2000, *Tulipomania: The Story Of The World's Most Coveted Flower & The Extraordinary Passions It Aroused*, (Random House Inc., New York City, NY).
- Dodd-Frank Wall Street Reform and Consumer Protection Act (Dodd-Frank Act), 2010, Public Law No. 111-203.
- Easley, David, Marcos M. Lopez de Prado and Maureen O'Hara, 2010, Measuring flow toxicity in a high frequency world, Unpublished Working paper, Cornell University and Tudor Investment Corp.
- Easley, David, Marcos M. Lopez de Prado and Maureen O'Hara, 2011, The Microstructure of the Flash Crash, *The Journal of Portfolio Management* 37, 118-128.
- Easley, David, Marcos M. Lopez de Prado and Maureen O'Hara, 2012, Flow toxicity and liquidity in a high frequency world, *Review of Financial Studies* 25, 1457-1493.
- Financial Conduct Authority (FCA), 2015, FCA secures High Court Judgment awarding injunction and over £7 million in penalties against five defendants for market abuse, August, 12. Retrieved from: <https://www.fca.org.uk/news/press-releases/fca-secures-high-court-judgment-awarding-injunction-and-over-£7-million>, November 17, 2016.
- Financial Conduct Authority (FCA), 2016, MiFID II, May 23. Retrieved from: <https://www.fca.org.uk/markets/mifid-ii>, January 26, 2017.

- Financial Industry Regulatory Authority (FINRA), 2010, Sanctions Trillium Brokerage Services, LLC, Director of Trading, Chief Compliance Officer, and Nine Traders \$2.26 Million for Illicit Equities Trading Strategy, September 13. Retrieved from: <http://www.finra.org/newsroom/2010/finra-sanctions-trillium-brokerage-services-llc-director-trading-chief-compliance>, December 12, 2016.
- Financial Industry Regulatory Authority (FINRA), 2012, FINRA Joins Exchanges and the SEC in Fining Hold Brothers More Than \$5.9 Million for Manipulative Trading, Anti-Money Laundering, and Other Violations, September 25. Retrieved from: http://www.finra.org/Newsroom/NewsReleases/2012/P178687?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+FINRANews+%28FINRA+Nsews%29, January 3, 2017.
- Financial Industry Regulatory Authority (FINRA), 2016, FINRA Issues First Cross-Market Report Cards Covering Spoofing and Layering, April 28. Retrieved from: <http://www.finra.org/newsroom/2016/finra-issues-first-cross-market-report-cards-covering-spoofing-and-layering>, December 19, 2016.
- Financial Tracking Technologies LLC (FTT), 2017, About Us. Retrieved from: <http://www.financial-tracking.com/about-us/>, January 14, 2017.
- Flood, Joe, 2010, NYSE Confirms Price Reporting Delays That Contributed to the Flash Crash, *ai5000*, August 24. Retrieved from: https://web.archive.org/web/20100827154851/http://assetinternational.com/ai5000/channel/NEWSMAKERS/NYSE_Confirms_Price_Reporting_Delays_That_Contributed_to_the_Flash_Crash.html, November 27, 2016.
- Gandel, Stephen, 2016, Man Facing 380 Years in Jail Probably Didn't Cause the Flash Crash, *Fortune Magazine*, January 28. Retrieved from: <http://fortune.com/2016/01/28/flash-crash-causes/>, December 12, 2016.
- Gao, Cheng, Bruce Mizrach and Sait Ozturk, 2015, Quote Stuffing and Market Quality, Working paper, Rutgers University and Erasmus University.
- Ghosh, Rex, 2010, The Flash Crash and Financial Terrorism: Could the Nineteenth Street NW Scenario Come True?, *Ecomonitor.com*, October 7. Retrieved from: <http://www.economonitor.com/blog/2010/10/the-flash-crash-and-financial-terrorism-could-the-nineteenth-street-nw-scenario-come-true/>, November 12, 2016.
- Golub, Anton, John Keane and Ser-Huang Poon, 2012, High Frequency Trading and Mini Flash Crashes, Working paper, Manchester University.
- Haldane, Andrew, 2010, Patience and Finance, *Bank of England*, September 9.
- Harris, Lawrence E., 2013, What to do about high-frequency trading. *Financial Analysts Journal* 69, 6-9.
- Hasbrouck, Joel and Gideon Saar, 2013, Low-latency trading, *Journal of Financial Markets* 16, 646–679.

- Hasbrouck, Joel, George Sofianos and Deborah Sosebee, 1993, New York Stock Exchange Systems and Trading Procedures, NYSE Working paper, New York University.
- Hess Legal Counsel, 2015, Spoofing: A Surveillance and Enforcement Challenge for Regulators, May 14. Retrieved from: <https://www.hesslegalcounsel.com/single-post/2015/05/14/Spoofing-A-Surveillance-and-Enforcement-Challenge-for-Regulators>, January 2, 2017.
- Hibernia Networks, 2016, A State-of-the-Art Transoceanic Cable Built for the Digital Age. Retrieved from: <https://www.hibernianetworks.com/network/hibernia-express/>, December 21, 2016.
- Hope, Bradley, 2015, Figure Charged With Flash Crash Role Is 'Old School Point and Click Prop Trader', *The Wall Street Journal*, April 23. Retrieved from: <http://blogs.wsj.com/moneybeat/2015/04/23/figure-charged-with-flash-crash-role-is-old-school-point-and-click-prop-trader/>, January 3, 2017.
- HTG Capital Partners vs. John Doe 1, A, B and C, Defendants, 2015, Case No. 15 C 02129, February 16.
- Hunsader, Eric S., [nanexllc], 2015, This chart exonerates Sarao for causing the flash crash (did he contribute? MAYBE. Cause it? NO) [Twitter Moment], April 24. Retrieved from: https://twitter.com/nanexllc/status/591648730501156865/photo/1?ref_src=twsrc%5Etfw, January 17, 2017.
- Hunsader, Eric S., 2011, Interview by Mike O'Hara: Improving Academic Research into HFT & Fighting High Frequency Spam, *High Frequency Trading Review*, October 26. Retrieved from: <http://www.thetradingmesh.com/pg/blog/mike/read/12977/improving-academic-research-into-hft-fighting-high-frequency-spam>, November 12, 2016.
- Investopedia*, 2016, Circuit Breaker. Retrieved from: <http://www.investopedia.com/terms/c/circuitbreaker.asp>, December 16, 2016.
- Investopedia*, 2017a, Sharpe Ratio. Retrieved from: <http://www.investopedia.com/terms/s/sharperatio.asp>, January 9, 2017.
- Investopedia*, 2017b, Black Swan. Retrieved from: <http://www.investopedia.com/terms/b/blackswan.asp>, January 8, 2017.
- Itskevich, Jennifer, 2002, What Caused the Stock Market Crash of 1987?, *History News Network*, July 31. Retrieved from: <http://historynewsnetwork.org/article/895>, January 8, 2017.
- Johnson, Neil, Guannan Zhao, Eric S. Hunsader, Jing Meng, Amith Ravindar, Spencer Carran and Brian Tivnan, 2012, Financial Black Swans Driven by Ultrafast Machine Ecology, University of Miami and Nanex LLC.
- Jones, Charles M, 2013, What do we know about high-frequency trading?, Working paper, Columbia Business School.

- Jovanovic, Boyan and Albert J. Menkveld, 2011, Middlemen in limit-order markets, Working paper, New York University and VU University Amsterdam.
- Kirilenko, Andrei, Albert S. Kyle, Mehrdad Samadi and Tugkan Tuzun, 2011, The Flash Crash: The Impact of High Frequency Trading on an Electronic Market, Working paper, Commodity Futures Trading Commission and University of Maryland and College Park.
- Kirilenko, Andrei, Albert S. Kyle, Mehrdad Samadi and Tugkan Tuzun, 2014, The Flash Crash: The Impact of High Frequency Trading on an Electronic Market, Working paper, Commodity Futures Trading Commission and University of Maryland and College Park.
- Laughlin, Gregory, Anthony Aquirre and Joseph Grundfest, 2012, Information transmission between financial markets in Chicago and New York, Unpublished Working paper, University of California, Santa Cruz, and Stanford University Law School.
- Leising, Matthew, Mira Rojanasakul and Adam Pearce, 2015, How to Catch a Spoofer, *Bloomberg*, September 4. Retrieved from: <https://www.bloomberg.com/graphics/2015-spoofing/>, January 5, 2017.
- Lemke, Thomas P. and Gerald T. Lins, 2013, *Soft Dollars and Other Trading Activities* (Thomson Reuters/West, Eagan, MN)
- Levine, Matt, 2015, The 'Flash Boys' Exchange Is Growing Up, *BloombergView*, September 16. Retrieved from: <https://www.bloomberg.com/view/articles/2015-09-16/the-flash-boys-exchange-is-growing-up>, October 12, 2016.
- Madhavan, Ananth, 2013, Exchange-Traded Funds, Market Structure, and the Flash Crash, *Financial Analysts Journal* 68, 20–35.
- Mahmudova, Anora, 2016, This man wants to upend the world of high-frequency trading, *Marketwatch*, February 17. Retrieved from: <http://www.marketwatch.com/story/this-man-wants-to-upend-the-world-of-high-frequency-trading-2016-02-02>, December 19, 2016.
- Malinova, Katya, Andreas Park and Ryan Riordan, 2013, Do retail traders benefit from improvements in liquidity?, Working paper, University of Toronto and University of Ontario Institute of Technology.
- Mamudi, Sam and Leslie Picker, 2015, Virtu Celebrates Another Year Without a Single Day of Losses, *Bloomberg*, February 20. Retrieved from: <https://www.bloomberg.com/news/articles/2015-02-20/high-frequency-trader-virtu-extends-nearly-unblemished-streak>, November 11, 2016.
- McInish, Thomas H. and James Upson, 2012, Strategic Liquidity Supply in a Market with Fast and Slow Traders, Working paper, University of Memphis and University of Texas at El Paso.

- McTague, Jim, 2010, How Foreigners Could Disrupt U.S. Markets, University of Delaware Library, September 11. Retrieved from:
http://green.lib.udel.edu/webarchives/kaufman.senate.gov/press/in_the_news/news/-id=0a760c05-5056-9502-5df6-cc9220d44e1d.htm, November 11, 2016.
- Menkveld, Albert J. and Bart Zhou Yueshen, 2016, The Flash Crash: A Cautionary Tale about Highly Fragmented Markets, Working paper, VU University Amsterdam.
- Nanex, 2010a, Nanex Flash Crash Summary Report, September 27. Retrieved from:
<http://www.nanex.net/FlashCrashFinal/FlashCrashSummary.html> , October 27, 2016.
- Nanex, 2010b, May 6'th 2010 Flash Crash Analysis Final Conclusion, October 14,
http://www.nanex.net/FlashCrashFinal/FlashCrashAnalysis_Theory.html October 27, 2016.
- Nanex, 2010c, Flash Equity Failures in 2006, 2007, 2008, 2009, 2010, and 2011, November 29. Retrieved from:
http://www.nanex.net/FlashCrashEquities/FlashCrashAnalysis_Equities.html, November 30, 2016.
- Nanex, 2010d, May 6'th 2010 Flash Crash Analysis Continuing Developments Sell Algo Trades, October 8. Retrieved from:
http://www.nanex.net/FlashCrashFinal/FlashCrashAnalysis_WR_Update.html, October 27, 2016.
- Nanex, 2014, Reexamining HFT's Role in The Flash Crash, June 12. Retrieved from:
<http://www.nanex.net/aqck2/4650.html>, October 24, 2016.
- Nanex, 2015, Nanex ~ 28-Apr-2015 ~ Spoofing Gold Futures, April 28. Retrieved from:
<http://www.nanex.net/aqck2/4700.html>, January 12, 2017.
- Nanex, 2017, What is NxCore?. Retrieved from: <http://www.nanex.net>, January 27, 2017.
- New York Stock Exchange (NYSE), 2015, Limit Up Limit Down. Retrieved from:
https://www.nyse.com/publicdocs/LULD_Overview.pdf, 11 January 2017.
- O'Hara, Maureen, 2014, High frequency market microstructure, *Journal of Financial Economics* 116, 257-270.
- Palmer, Neil, Michael Sherman, Yingxia Wang and Sebastian Just, 2015, Scaling to Build the Consolidated Audit Trail: A Financial Services Application of Google Cloud Bigtable, Fidelity National Information Services Inc.
- Pasquale Frank, 2014, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press, Cambridge, MA and London, England).
- Pettinger, Teyvan, 2008, Credit Default Swaps Explained, *Economicshelp*, November 11. Retrieved from: <http://www.economicshelp.org/blog/933/finance/credit-default-swaps-explained/>, November 8, 2016.

- Phillips, Matt, 2010, SEC's Schapiro: Here's My Timeline of the Flash Crash, *The Wall Street Journal*, May 20. Retrieved from: <http://blogs.wsj.com/marketbeat/2010/05/20/secs-schapiro-heres-my-timeline-of-the-flash-crash/>, November 3, 2016.
- Picardo, Elvis, 2014, An Introduction To Dark Pools, *Investopedia*, May 6. Retrieved from: <http://www.investopedia.com/articles/markets/050614/introduction-dark-pools.asp>, November 29, 2016.
- Rojcek, Jakub and Alexandre Ziegler, 2016, High-Frequency Trading in Limit Order Markets: Equilibrium Impact and Regulation, Working paper, University of Zurich.
- Securities Industry and Financial Markets Association (SIFMA), 2016, Overview CAT Implementation Timeline. Retrieved from: [http://www.sifma.org/issues/legal,-compliance-and-administration/consolidated-audit-trail-\(cat\)/overview/](http://www.sifma.org/issues/legal,-compliance-and-administration/consolidated-audit-trail-(cat)/overview/), December 15, 2016.
- Simpson, Stephen D., 2010, What Caused The Flash Crash?, *Investopedia*, June 18. Retrieved from: <http://www.investopedia.com/financial-edge/0610/what-caused-the-flash-crash.aspx>, November 27, 2016.
- Stafford, Philip, 2013, Europe agrees on high-speed trading regulation, *Financial Times*, October 22. Retrieved from: <https://www.ft.com/content/810ce436-3b28-11e3-87fa-00144feab7de>, January 22, 2017.
- The Telegraph*, 2016, British 'Flash Crash' trader Navinder Sarao pleads guilty in US to fraud, September 9. Retrieved from: <http://www.telegraph.co.uk/news/2016/11/09/british-flash-crash-trader-navinder-sarao-pleads-guilty-in-us-to/>, December 11, 2016.
- Thompson, Earl A., 2007, The tulipmania: Fact or artifact?, *Public Choice* 130, 99-114.
- Tong, Lin, 2015, A Blessing or a Curse? The Impact of High Frequency Trading on Institutional Investors, Working paper, Fordham University.
- Tovey, Alan, 2014, High-frequency trading: when milliseconds mean millions, *The Telegraph*, April 2. Retrieved from: <http://www.telegraph.co.uk/finance/newsbysector/banksandfinance/10736960/High-frequency-trading-when-milliseconds-mean-millions.html>, December 12, 2016.
- Tuttle, Laura, 2013, Alternative Trading Systems: Description of ATS Trading in National Market System Stocks, SEC Division of Economics and Risk Analysis.
- U.S. Commodity Futures Trading Commission (CFTC), 2010, External Meetings: Meeting with Barclays Capital Inc., October 14. Retrieved from: http://www.cftc.gov/LawRegulation/DoddFrankAct/ExternalMeetings/dfmeeting_101410_174, December 14, 2016.
- U.S. Commodity Futures Trading Commission (CFTC), 2013, Interpretive Guidance and Policy Statement on Disruptive Practices.

- U.S. Commodity Futures Trading Commission (CFTC), 2015, Release: pr7156-15, April 21.
Retrieved from: <http://www.cftc.gov/PressRoom/PressReleases/pr7504-16>,
December 13, 2016.
- U.S. Commodity Futures Trading Commission (CFTC), 2016, Release: pr7504-16,
December 20. Retrieved from:
<http://www.cftc.gov/PressRoom/PressReleases/pr7504-16>, December 29, 2016.
- U.S. Securities & Exchange Commission (SEC) and U.S. Commodity Futures Trading
(CFTC) Commission, 2010a, *Preliminary Findings Regarding the Market Events of
May 6, 2010*, May 18.
- U.S. Securities & Exchange Commission (SEC) and U.S. Commodity Futures Trading
(CFTC) Commission, 2010b, *Findings Regarding the Market Events of May 6, 2010*,
September 30.
- U.S. Securities and Exchange Commission (SEC), 2010c, SEC Adopts New Rule Preventing
Unfiltered Market Access, November 3. Retrieved from:
<https://www.sec.gov/news/press/2010/2010-210.htm>, December 4, 2016.
- U.S. Securities and Exchange Commission (SEC), 2010d, SEC Approves New Rules
Prohibiting Market Maker Stub Quotes, November 8. Retrieved from:
<https://www.sec.gov/news/press/2010/2010-216.htm>, December 4, 2016.
- U.S. Securities and Exchange Commission (SEC), 2010e, Release No. 34-62251, June 10.
- U.S. Securities and Exchange Commission (SEC), 2011a, Release No. 34-65954,
December 14.
- U.S. Securities and Exchange Commission (SEC), 2011b, Release No. 34-65955,
December 14.
- U.S. Securities and Exchange Commission (SEC), 2012a, Transcript: Technology and
Trading Roundtable: Promoting Stability in Today's Markets, October 2.
- U.S. Securities and Exchange Commission (SEC), 2012b, Release No. 34-67292, June 28.
- U.S. Securities and Exchange Commission (SEC), 2014, Equity Market Speed Relative to
Order Placement, Data Highlights 2013-05 and 2014-02, March 19.
- U.S. Securities and Exchange Commission (SEC), 2015, No. 34-73639, February 3.
- U.S. Securities and Exchange Commission (SEC), 2016a, SEC Approves Plan to Create
Consolidated Audit Trail, November 15th 2016. Retrieved from:
<https://www.sec.gov/news/pressrelease/2016-240.html>, January 2, 2017.
- U.S. Securities and Exchange Commission (SEC), 2016b, Federal Register, Vol. 81,
Release No. 34-78101, FR 41142-41170, June 23.
- United States of America (U.S.A.) vs. Michael Coscia, 2015, Case No. 14 CR 551, October
5.

- United States of America (U.S.A.) vs. Sarao, 2015, Case No. 15 CR 75, September 2.
- Van Kervel, Vincent, 2012, Liquidity: What You See is What You Get?, Working paper, Tillburg University.
- Vicci, Gino, 2012, Critics Weigh in on High-Frequency Trading on Flash Crash Anniversary, *Huffington Post*, May 9. Retrieved from: http://www.huffingtonpost.com/gino-vicci/high-frequency-trading_b_1504418.html, December 6, 2016.
- Whipp, Lindsay and Kara Scannell, 2016, 'Flash-crash' trader Navinder Sarao pleads guilty to spoofing, *Financial Times*, November 10. Retrieved from: <https://www.ft.com/content/a321031a-a6cb-11e6-8898-79a99e2a4de6>, December 2, 2016.
- Ye, Mao, Chen Yao and Jiading Gai, 2013, The Externalities of High Frequency Trading, Working paper, University of Illinois and University of Warwick.
- Zanki, Tom, 2015, CME Ordered To Prove Identities Of 'Spoofing' Defendants, *Law360*, September 23. Retrieved from: <https://www.law360.com/articles/706115/cme-ordered-to-prove-identities-of-spoofing-defendants>, January 2, 2017.
- ZeroHedge*, 2011, Zero Hedge Kindly Requests The Immediate Resignation Of Mary Schapiro For Gross Breach Of Professional Responsibilities, October 4. Retrieved from: <http://www.zerohedge.com/news/zero-hedge-kindly-requests-immediate-resignation-mary-schapiro-gross-breach-professional-respon>, December 11, 2016.