



UNIVERSITA' DEGLI STUDI DI PADOVA

**DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI
"M.FANNO"**

**CORSO DI LAUREA MAGISTRALE IN
ECONOMICS AND FINANCE**

TESI DI LAUREA

**"The High Frequency Trading Phenomenon and its Influence on Capital
Markets: Evidences from the Pound Flash Crash"**

RELATORE:

CH.MA PROF.SSA BALDAN CINZIA

LAUREANDO: CAVESTRO STEFANO

MATRICOLA N. 1103591

ANNO ACCADEMICO 2016 – 2017

Il candidato dichiara che il presente lavoro è originale e non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere.

Il candidato dichiara altresì che tutti i materiali utilizzati durante la preparazione dell'elaborato sono stati indicati nel testo e nella sezione "Riferimenti bibliografici" e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo alla pubblicazione originale.

Firma dello studente

TABLE OF CONTENTS

INTRODUCTION	9
1. THE BIRTH OF HIGH FREQUENCY TRADING	11
1.1. Revolution of Financial Markets	11
1.2. Evolution of Electronic Trading	11
1.3. Regulatory Developments in the Capital Markets	13
1.4. Electronic Trading Drivers	17
1.4.1. Smart Order Routing (SOR)	17
1.4.2. Direct Market Access (DMA) and Sponsored Access (SA)	18
2. HIGH FREQUENCY TRADING DEFINITIONS AND FEATURES	21
2.1. The need of defining High Frequency Trading	21
2.2. Definitions of Algorithmic Trading	22
2.3. Characteristics of Algorithmic Trading	23
2.4. Definitions of High Frequency Trading	26
2.5. Characteristics of High Frequency Trading	27
2.5.1. Low-Latency Requirements	29
2.6. How to Identify High Frequency Traders?	32
3. HIGH FREQUENCY TRADING STRATEGIES	35
3.1. Introduction to HFT Strategies	35
3.2. Algorithmic Trading Strategies	36
3.2.1. First Generation Execution Algorithms	37
3.2.2. Second Generation Execution Algorithms	38
3.2.3. Third Generation Execution Algorithms	38
3.2.4. Fourth Generation Execution Algorithms	39
3.3. High Frequency Trading Strategies	39
3.3.1. Liquidity Providing Strategy	40
3.3.1.1. Spread Capturing	41
3.3.1.2. Rebate Driven	41
3.3.2. Statistical Arbitrage Strategy	42
3.3.2.1. Market Neutral Arbitrage	43
3.3.2.2. Cross Asset, Market & ETF Arbitrage	43
3.3.3. Directional Strategy	44
3.3.3.1. Momentum Trading	44
3.3.3.2. Liquidity Detection Strategy	45

3.3.3.2.1. Quote Matching	46
3.3.3.2.2. Momentum Ignition	47
3.3.4. Structural Strategy	48
3.3.4.1. Latency Arbitrage	49
3.3.4.2. Flash Trading	51
3.4. Considerations on High Frequency Trading Strategies	52
4. HIGH FREQUENCY TRADING AND MARKET QUALITY	53
4.1. Introduction to Literature Review	53
4.2. Market Quality	54
4.2.1. Liquidity	55
4.2.2. Price Discovery	57
4.2.3. Volatility	58
4.3. Theoretical models on market quality	59
4.4. Empirical studies on market quality	61
4.4.1. HFT direct approach	61
4.4.2. HFT indirect approach	65
4.5. Summary of Academic Literature Review	67
5. RISKS OF HIGH FREQUENCY TRADING	69
5.1. Introduction to Risks of High Frequency Trading	69
5.2. Two-Tiered Markets	69
5.3. Phantom Liquidity	70
5.4. Market Manipulation	72
5.5. Operational Risk	74
5.6. Systemic Risk	75
5.7. Flash Crashes	76
5.7.1. The Flash Crash of May 6, 2010	77
5.7.2. Mini-Flash Crashes	82
6. HIGH FREQUENCY TRADING REGULATION	87
6.1. Regulatory Activity and Discussion	87
6.2. HFT Regulation in the US	87
6.3. HFT Regulation in the European MiFID II	91
6.4. Ideas and Proposal Measures	95
6.5. Considerations on HFT Regulation Developments	98
7. EMPIRICAL ANALYSIS OF THE POUND FLASH CRASH	99
7.1. Introduction of the Pound Flash Crash	99

7.2. The GBP Foreign Exchange Market	100
7.3. A Flash Event in Three Stages	102
7.4. Descriptive Analysis of the Pound Flash Crash	106
7.5. Analysis of Liquidity during the Pound Flash Crash	111
7.6. Triggers, Vulnerabilities and Amplifiers of the Pound Flash Crash	115
7.7. The Sterling Context prior the Flash Crash	118
7.8. The Lasting Impact of the Pound Flash Crash	120
7.9. Conclusions about Pound Flash Crash	123
CONCLUSIONS	127
FIGURES AND TABLES	129
REFERENCES	131

INTRODUCTION

Over the past decades, capital markets have changed drastically the way financial trades are carried out. Nowadays, outcry markets conducted by men wearing in color-coded jackets that ran, yelled, and gesticulated in order to stand out amongst the crowd to transfer information across the stock exchange floor, do not exist anymore.

Conversely, the most of trades on securities markets are now dominated by sophisticated operators utilizing super computers with complex algorithms to perform thousands of trades in a matter of milliseconds or even microseconds, before a normal person could even blink their eyes, in a practice known as high frequency trading (HFT).

Such practice is the last step of an evolutionary process which has involved the financial system originated by a series of substantial changes spurred by a favorable regulatory environment, technological innovation concerning electronics and communication networks, and a growing competition among traders and stock exchanges.

In spite of the predominant role of HFT, currently, there does not exist a univocal and globally recognized identification of the phenomenon and this has complicated the acknowledgment of the traders who use it (HFTs) and consequently their regulation. Nevertheless, it is catalogued as a subset of algorithmic trading (AT), a technique that uses pre-set algorithms to make decisions such as timing, pricing, execution and modification of orders in total autonomy without human intervention.

There are unique aspects that make distinguish HFT from other trading activity. One of these features is represented by the employment of specific infrastructure aimed at limiting as much as possible the trading latency, i.e. the timing of processing, analyzing, and sending of orders, which constitutes a significant competitive advantage with respect to traditional traders.

In addition, it is not recognized in a specific trading strategy but HFTs can utilize their technological advances in terms of computer software, hardware, and ultra-low latency networks by applying a wide range of different trading strategies with the aim to make profit.

The global presence of HFT on the securities markets and its rapid development have given rise to a heated debate among financial operators and regulatory authorities regarding the potential consequences that such trading practice may entail on the market quality aspects; in particular experts are wondering whether HFT produces more beneficial or detrimental effects concerning liquidity, trading costs, bid-ask spread, price discovery, efficiency and volatility; but since HFT branches off in very different trading applications and HFTs are identified by using different methods, also the answers about market implications result very divergent.

Additionally, HFT has drawn remarkable public attention after the notorious “Flash Crash” of May 6, 2010 and subsequent observations of market failures, which has put under the magnifying glass some specific aggressive high frequency practices which demonstrated to have provoked an important risk for the integrity and stability of capital markets.

These observed phenomena led in the following years to a series of regulations targeted to strongly discipline the HFT world and heavily deter the most contested trading activities.

The fundamental aim of this work is, therefore, to describe, analyze and study the high frequency trading phenomenon in most of its aspects, as detailed as possible, and by maintaining an impartial position about it. This is made through the reporting of relevant literature in subject, both on the side of the supporters and of the detractors, whom have dealt with the effects of HFT on capital markets dynamics. Furthermore, we are going to examine an observation of a *flash crash* phenomenon that hit the British Sterling and from which it is supposed that HFT has played an important role on its appearance.

This present work is organized in the following way. Chapter 1 describes the revolutionary process occurred in the capital markets in the last years, both in terms of technological adoptions and regulatory developments, with the implementation of a series of projects and initiatives that resulted a very favorable context for the birth and proliferation of HFT activities.

Chapter 2 details the specific characteristics of HFT and AT and determines the different approaches used for identifying HFTs.

Chapter 3 lists the wide range of trading strategies that ATs and HFTs may implement, some of these are not new for the trading system, while others are exclusively carried out by HFTs. Different trading strategies implies also different assessments of behavior of HFTs.

Chapter 4 reports the relevant studies about HFT and its impact on market quality measures, such as liquidity, price discovery and volatility. These studies may be in form of theoretical and empirical methodology, with different estimations and approaches for the identification of HFT.

Chapter 5 debates the potential risks coming from HFT activity. Among these, we report phenomena of market abuse and observations of market failures, such as flash crashes.

Chapter 6 is a review of the current regulation activity on HFT subject. It includes also a discussion of potential rules that may be adopted to strongly limit HFT activity.

Finally, Chapter 7 examines an empirical observation of a flash crash occurred on Sterling trading. We analyze the specific dynamics of the phenomenon, the potential causes and effects on future trading. In addition, we wonder whether the behavior of HFT has played a role in that specific situation.

1. THE BIRTH OF HIGH FREQUENCY TRADING

1.1. Revolution of Financial Markets

Advances in technology have totally changed the way in which transmission and execution of orders are performed on financial markets. A fundamental role is attributed to the “*electronic revolution*”, that through the emergence of innovative practices both for trading Exchanges and market traders, has led to many advantages for financial trading including the reduction of order execution time and trading costs.

In this scenario, new trading disciplines were born representing the iteration between two worlds: the finance and the technological innovation. The former introduces the object of their core activity, whilst the latter specifies the way in which such activity is exerted.

Sophisticated trading techniques, such as algorithmic and high frequency trading, are the last result of a series of evolution processes happened in the last years involving the financial system; trading evolution takes into account changes occurred by means of: communication networks, electronics, innovative technologies such as algorithms, and regulatory developments.

In this chapter, we will describe the principal events that have revolutionized the structure and the functioning of the capital markets, identifying the steps and causes that have contributed to the affirmation of alternative trading techniques, including the high frequency trading phenomenon.

1.2. Evolution of Electronic Trading

Securities trading methods and techniques have come a long way since the early 18th century when they were carried out in the first US financial market established on Wall Street.

For many years, securities markets were represented by a physical location called “*trading floor*” where buyers and sellers met in order to execute their trades¹.

Up to 1960s, trade information was spread in human fashion, typically in person or through telegraph and telephone.

Thanks to the emergence of electronics and computerized systems, there has been an important shift in how securities are traded on the capital markets.

The use of information technology for order flow in the financial markets has started about 40 years ago in 1971, when the National Association of Securities Dealers (NASD) introduced a

¹ Anuj Agarwal (2012), *High Frequency Trading: Evolution and the Future*, Capgemini, 4

market-making system assisted by computers, by creating its own automatic quotation (AQ) system and therefore giving life to the National Association of Securities Dealers Automated Quotation (NASDAQ), the world's first electronic stock market².

A few years later, the New York Stock Exchange (NYSE) introduced a procedure known as “*Designated Order Turnaround*” (DOT) system in 1976, and later Super-DOT in 1984. Such procedures allowed the transmission of orders to the trading post in an electronic manner; that is, orders appeared on a special workstation, called “*Display Book*”, which enabled each specialist located on the floor of the Exchange to execute orders for the market³.

These first electronic systems have permitted to consistently speed up the dissemination of information about orders, compared to the slower physical meetings on the trading floor and telephone communications.

The 1980s saw the beginning of decline of floor-based trading caused by the advent of the first fully-electronic financial markets and the performing of particular trading strategies such as “*Program Trading*”, loosely defined by the NYSE as the placing of orders to buy or sell 15 or more stocks valued above \$1 million total. This strategy gained popularity for trades between the S&P 500 equity shares and the futures markets because “program traders” could buy or sell stock index futures contracts, such as the S&P 500 futures, and at the same time sell or buy a portfolio of up to 500 stocks at the NYSE matched against the futures trade; with the aid of computers, this program could be pre-programmed to enter automatically into the NYSE's electronic order routing system at the time when the futures price and the stock index were far enough apart in order to make a profit⁴. Nevertheless, this type of stock index arbitrage strategy was later condemned by some as leading to the “*Black Monday*” stock market crash in 1987⁵.

As it regards the use of electronics in the implementation of orders, in 1985, NASDAQ established an automatic order execution platform only for small volume orders called “*Small Order Execution System*” (SOES). This practice enabled small traders to execute orders automatically without the intermediation of a dealer and by means of this, very aggressive traders called “*SOES bandits*” realized several trades per day in order to profit by establishing positions before most market makers have updated their quotes and lay off them at favorable prices⁶.

² Black, Fischer (1971), *Toward a Fully Automated Stock Exchange*, Part 1, Financial Analysts Journal 27, 29-34

³ Jerry W. Markham & Daniel J. Harty (2008), *For Whom the Bell Tolls: The Demise of Exchange Trading Floors and the Growth of ECNs*, 33 Iowa J. Corp. L., 897-898

⁴ Dean Furbush (2010), *Program Trading*, Concise Encyclopedia of Economics

⁵ Liz Moyer & Emily Lambert (2009), *Wall Street's New Masters*, Forbes, Sept. 21

⁶ Jeffrey H. Harris & Paul H. Schultz (1998), *The Trading Profits of SOES Bandits*, Journal of Financial Economics Vol. 50, No. 2, Oct.

In 1992, the Chicago Mercantile Exchange (CME), a financial and commodity derivative Exchange, launched the first global electronic trading platform called Globex⁷. This fully electronic trading system allowed market participants to trade futures contracts through a computer software program over a network connected to a financial intermediary while they were sitting in office thousands of miles away from the exchange.

While electronic trading in the eighties was dominated by trading on NASDAQ and NYSE, the 1990s saw the emergence of other electronic trading venues called “*Electronic Communications Networks*” (ECNs). An ECN is a type of computer network that facilitates trading of financial securities outside of the regulated Stock Exchanges⁸.

Subscribers to ECN can enter orders electronically (e.g. using algorithms) into the network via a custom computer terminal and then the system will automatically match and execute contra-side orders; in case of mismatching, the ECN orders could be posted externally on the regular markets⁹.

The development of ECNs has provided many benefits with respect to the traditional trading venues including the reduction in costs and trading errors, greater speed and efficiency¹⁰.

In 2010, all seven US exchanges offered either fully electronic or a hybrid mix of floor and electronic trading in options¹¹.

Therefore, the process of electronification has been slow and gradual, but the last years has seen an acceleration which has completely changed the financial framework and has led to the birth of new market participants, such as high frequency traders.

1.3. Regulatory Developments in the Capital Markets

The development of new trading methods and the creation of alternative electronic markets would not have been possible without a regulatory environment that has allowed and even encouraged such proliferation.

Regulatory agencies such as the US Security and Exchange Commission (SEC) and the European Securities and Markets Authority (ESMA) have the core objective of protecting

⁷ Pavitra Kumar, Michael Goldstein, Frank Graves, & Lynda Borucki (2011), *Trading at the Speed of Light: The Impact of High-Frequency Trading on Market Performance, Regulatory Oversight, and Securities Litigation*, The Battle Group Issue 02, 2

⁸ Lauren Liebenberg (2002), *The Electronic Financial Markets of the Future, and Survival Strategies of the Broker-Dealers*, 73-75

⁹ *Id*

¹⁰ *Id*

¹¹ See Kumar, *supra* note 7

investors, and maintaining fair and efficient markets by improving market transparency, capital formation and competition¹².

In the US, the late 1990s and early 2000s saw the US Commission introduce a series of reforms aimed at modernizing and regulating the financial markets, which have had a deep impact on the securities system structure by implicitly favoring the implementation and subsequent success of more advanced trading techniques, including algorithmic and high frequency trading. In 1998, the SEC issued the Alternative Trading Systems Regulation (Reg. ATS)¹³, where authorized and resolved any concerns about alternative trading systems.

An alternative trading system was defined by Reg. ATS as a “*non-exchange trading venue that performs as a Stock Exchange by matching buyers and sellers of securities but does not set rules for subscribers like regulated venues*”. An example of alternative trading system is an ECN.

One of the main weaknesses of the first ECNs in the early nineties consisted in the presence of a sort of “*Chinese Wall*” between ECNs and regulated markets inasmuch as, trading orders issued to ECNs could not be routed to Exchanges¹⁴. Such market split implied the risk for investors to receive a worst price on ECNs with respect to regulated Exchanges.

This market inefficiency has been resolved by the Reg. ATS with the introduction of a “*Limit Order Display Rule*”, a transparency imposition which obliged market-makers to show traders the best present price in all regulated and alternative markets.

Such regulation has been the precursor of innovative electronic systems (e.g. Smart Order Routing system) able to automatically connect ECNs and Exchanges in a very quick way.

In particular, Smart Order System has been fundamental for the emergence of high frequency trading, which exploits its speediness advantage in order to perform trades specially through inter-market connections.

However, the milestone of the algorithmic and high frequency trading phenomenon dates back on 2005 and it was due to the Regulation National Market System (Reg. NMS)¹⁵.

SEC defined Reg. NMS as “*a series of initiatives designated to modernize and strengthen the National Market System for equity securities*”. The regulation was intended to assure that investors receive the best price executions for their orders by encouraging competition among individual markets and individual traders.

The most notable rules of the NMS are:

¹² U.S. Securities and Exchange Commission website available at: <https://www.sec.gov/about/whatwedo.shtml>

¹³ Alternative Trading Systems Regulation (1998) available at: <https://www.sec.gov/rules/final/34-40760.txt>

¹⁴ Alfonso Puorro (2013), *High Frequency Trading: una panoramica*, Questioni di Economia e Finanza n° 198, Banca d'Italia, 7

¹⁵ National Market System Regulation (2005) available at: <http://www.sec.gov/rules/final/34-51808.pdf>

- i) Access Rule (Rule 610);
- ii) Order Protection Rule (Rule 611);
- iii) Sub Penny Rule (Rule 612).

The Access Rule¹⁶ was intended to create an environment in which there was fair and efficient access to quotations throughout the National Market System by requiring greater linking and lower access fees.

The Order Protection Rule¹⁷ also known as “*Trade-Through Rule*” claimed that any submitted order should be executed at the National Best Bid and Offer price (NBBO).

The Rule 611, hence, ensured that both institutional and retail investors must receive the best possible execution price for orders that can be executed immediately and consequently, it prohibited market operators to execute orders at any inferior price.

Before this rule, brokers were allowed to execute orders just at the venue’s best available price even if another platform displayed a better price; but by means of the Order Protection Rule, whether the best price is a displayed price, it cannot be “trade-through” or, in other words, it cannot be ignored.

The Order Protection Rule implied that each venue must establish and enforce policies to ensure consistent price quotations for all NMS traded financial products, which include the major stock Exchanges as well as the ATS.

The rule was imposed by the SEC to make the financial markets more liquid and transparent and for reinforcing the inter-market connections among venues by routing orders in order to be execute at the NBBO.

Through the Sub Penny Rule¹⁸, SEC, instead, imposed to all American financial markets to adopt the decimal system as units of measure for computing securities price quotations over the unity. The decimalization was intended to make as small as possible (i.e. \$0.01) and univocal, the minimum price movement granted for all stocks traded, called “*tick size*”.

In this way, the bid-ask spread has been tightened by making the financial instruments cheaper and by increasing the opportunities for liquidity provision; this has opened up a new road for the development of very fast automatic systems, able to execute many operations per day, which may offer or ask a tick value slightly above or below of current quotation in order to exploit new trading suitability based on a potential smaller profit but more present and less risky.

¹⁶ Responses to Frequently Asked Questions Concerning Rule 611 and Rule 610 of Regulation NMS available at: <http://www.sec.gov/divisions/marketreg/nmsfaq610-11.htm>, April 4, 2008 Update

¹⁷ *Id*

¹⁸ Responses to Frequently Asked Questions Concerning Rule 612 (Minimum Pricing Increment) of Regulation NMS available at: <http://www.sec.gov/divisions/marketreg/subpenny612faq.htm>, October 21, 2005 Update

Even in Europe, similar dispositions were introduced with the Markets in Financial Instruments Directive (MiFID)¹⁹ issued in 2004 by the co-decision procedure of the Council of the European Union and the European Parliament and the subsequent implementation in 2007 in each European country legislation. The directive's main objectives were to increase competition and consumer protection in investment services. The most relevant rules of MiFID are:

- i) Development of Multilateral Trading Facilities;
- ii) Introduction of Best Execution for investment firms;
- iii) Obligation of Transparency.

MiFID has eliminated the monopolistic concentration of the financial instruments traded on Regulated Exchanges, by enabling the development of new categories of trading venues called Multilateral Trading Facilities (MTFs). A MTF can be seen as the European equivalent to the US Alternative Trading Systems, in particular is defined as a “*non-exchange financial trading venue, alternative to the traditional stock exchange, where securities are traded typically using electronic systems*”. The new entrant MTFs have had a considerable impact on European share-trading by initiating a process of “fragmentation” of the financial markets, where liquidity for one financial instrument is no longer concentrated on one exchange but across multiple venues competing each other²⁰.

The diffusion of MTFs, and in particular thanks to advanced platforms able to increase the speed of execution and charged low commissions, has led to the development of sophisticated traders as high frequency traders that implement their trading strategies based on a competitive and low-latency environment.

The introduction of Best Execution rule referred to the duty for investment firms to perform the best possible execution of their clients' orders, mainly, in terms of better execution price and speed. This rule is the symmetric of the NBBO for the US financial markets²¹.

As it regards transparency, MiFID has required to make available order information such as by displaying trading quotes (pre-trade transparency) and publishing the price, volume and time of all trades in listed shares (post-trade transparency), even if executed outside of regulated Exchanges²².

¹⁹ MiFID available at: <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02004L0039-20110104&from=EN>

²⁰ Carole Gresse (2012), *Market Fragmentation in Europe: Assessment and Prospects for Market Quality*, Driver Review DR19, Government Office for Science

²¹ *Best Execution under MiFID: Questions & Answers*, CESR, 2007 available at: https://www.esma.europa.eu/sites/default/files/library/2015/11/07_320.pdf

²² *Non-Equity Market Transparency: Questions & Answers*, CESR, 2007 available at: http://ec.europa.eu/finance/securities/docs/esme/answerstoquestions_en.pdf

Summarizing, the phenomenon of high frequency trading dates back and owes its success to a set of rules issued by the US before and Europe afterwards.

These rules, implemented with the purpose to increase competition among venues, investors protection and trading transparency, have facilitated the introduction of more modern trading techniques utilizing technologies such as algorithms and trading speed in order to exploit the fragmentation of the trading industry and to enjoy of fully electronic alternative trading platforms in which such high frequency trading techniques may be more easily carried out.

1.4. Electronic Trading Drivers

The emergence of computerization in financial markets has favored the use of specific drivers employed with the aim to facilitate the interaction among traders and trading venues and trading venues among them via an electronically manner, and incentivized by the increasing market fragmentation and the new European and US regulations.

Such drivers have been an important success key for Algorithmic and High Frequency Trading and in particular are represented by technology systems such as:

- Smart Order Routing;
- Direct Market Access and Sponsored Access.

1.4.1. Smart Order Routing (SOR)

The obligation to execute orders at the NBBO in US and at the Best execution price in Europe, in conjunction with the market fragmentation of the trading venues in place during the last years, has led the creation of a technological infrastructure, called Smart Order Routing, able to guarantee to the traders a fast and simple access to the best available price in any market and any time.

This procedure is agreed with the “Order Protection Rule” stating that the markets have to create, maintain and reinforce rules and procedures in order to avoid mechanisms such as the “trade-through” which entail the performing of trades at worst prices than protected prices²³.

The best execution price requirement has resulted in the need to connect regulated markets and ECNs (or MTFs), but also in the development of tools able to analyze prices and the relative quantities of securities on each trading venues.

²³ See note 16

The Smart Order Routing is the system that allows real-time analysis of all traded financial instruments' quotations and, through a set of steps, enables to guarantee the best price execution to the investor²⁴.

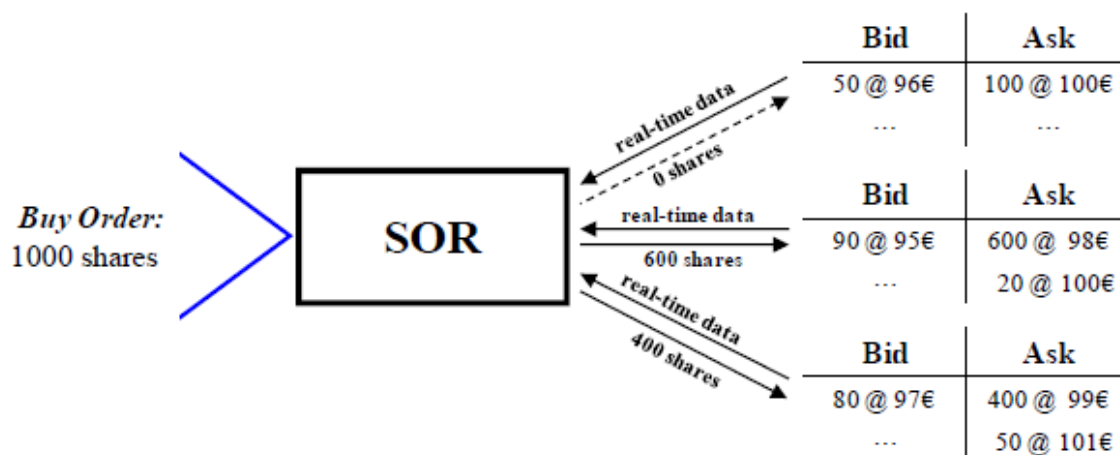


Figure 1: Smart Order Routing - basic principle. Source: Ende and Lutat, 2010.

The figure 1 represents an example of the Smart Order Routing process. An investor issues an order to buy 1000 shares at the current best price (a market buy order). The instrument is currently traded in three different trading venues, each of which has its own trading book where appears the actual bid/ask quotations. The Smart Order Routing compares all markets and, since the incoming order is a buy order, it routes the order where there is the lowest ask price. The first venue has an ask price of €100 for 100 shares, the second venue has an ask price of €98 for 600 shares and €100 for 20 shares, while the third venue has an ask price of €99 for 400 shares and €101 for 50 shares. Therefore, the order will be sent to the second venue and executed for 600 shares at €98 and the remaining quantity will be routed to the third venue where executed for 400 shares at a price of €99. In this way, the investor will obtain an average price of $\frac{600}{1000} * €98 + \frac{400}{1000} * €99 = €98.4$, that is the best possible execution price.

The Smart Order Routing system, therefore, constitutes a guarantee for retail investors to receive the current best execution price for an instrument traded.

1.4.2. Direct Market Access (DMA) and Sponsored Access (SA)

In most markets, the trading access is restricted to registered market members; hence, those members are the only ones permitted to perform trading directly by interacting with the order

²⁴ Peter Gomber, Björn Arndt, Marco Lutat, Tim Uhle (2011), *High-Frequency Trading*, Goethe Universitat Frankfurt Am Main, 19-20

book of the trading venues. They are specifically brokers-dealers and market making firms leading to their primary role as market access intermediaries for other investors²⁵.

The Direct Market Access is a service offered by some brokers that enables, through an information technology infrastructure, to sophisticated private traders (i.e. high frequency traders) the placement of buy and sell orders directly on the trading venues' books without being a registered market member²⁶. Thus, high frequency traders utilizing this infrastructure are able to control the way a trading transaction is carried out rather than passing the order to the broker for execution; however, DMA procedure guarantees to broker the conduction of pre-trade risk checks.

Another possibility for the buy side to access a marketplace is granted by the Sponsored Access (SA). A SA is very similar to the DMA except for the fact that the investor does not use the broker's infrastructure but it routes his orders to the execution venue using the broker's membership ID²⁷. Unlike DMA, the broker who enables a Sponsored Access to their clients can conduct pre-trade risk checks only if the execution venue provides such a system and service to the broker, and in this case, we speak of "filtered Sponsored Access".

In the case of "unfiltered (or naked) SA", on the contrary, the broker only receives a drop copy of each order that the investor has sent to the trading venue, and his checks will be done using this information²⁸; therefore, the order execution is not more dependent by the broker's risk checks because it is inspected after submission.

Direct Market Access and Sponsored Access allow to investment firms and private traders similar benefits to those of the membership, but with lower costs and maintaining the anonymity with respect to the market and other participants.

The main advantage of unfiltered Sponsored Access over Direct Market Access refers to the huge reduction in the timing of order submission because of the lack of broker's pre-trade checks; in fact, high frequency traders are more attracted by this type of market access practice in order to perform their trading strategies.

²⁵ See note 24, 9

²⁶ *Id*

²⁷ *Id*

²⁸ *Id*

2. HIGH FREQUENCY TRADING DEFINITIONS AND FEATURES

2.1. The need of defining High Frequency Trading

Before determining the impact of high frequency trading on capital markets, it is necessary to provide a definition of the phenomenon by identifying also its characteristics and strategies. Given that HFT is a trading practice which is relatively recent and it is still evolving, either authors and regulators have different opinions on describing it and consequently it exists a distortion on the classification of the traders that belong to the high frequency trading category. In this chapter, we will provide several definitions in order to have a general view of this innovative trading type and accordingly being able to classify the specific features belonging to high frequency traders.

The first thing to keep in mind is that high frequency trading is a trading method resulting by an evolution and differentiation of other trading techniques.

First of all, high frequency trading belongs to the great family of electronic trading which generally is referred to the ability of transmitting orders electronically, usually via computer networks²⁹.

In addition, high frequency traders make a massive use of automated algorithms for order-execution processes as well as for high-frequency portfolio allocation decisions; this entails that are classified as algorithmic traders³⁰.

Furthermore, high frequency trading is catalogued as a sub-category of algorithmic trading systems, inasmuch as it contains further specific features, for example low-latency order executions, short position holding times and no positions held overnight, which have implied the birth of this new trading practice and the subsequent success.

After these considerations, we can say that High Frequency Trading is a distinct trading practice that is catalogued as 1) an electronic trading that 2) employs algorithms and 3) uses high-frequency infrastructures and strategies.

Figure 2 illustrates the differences among traditional traders, algorithmic traders and high frequency traders in terms of order execution speed and position holding period.

As we note, high frequency trading diverges from other traders because of the less time necessary to execute orders and the lower position holding period.

²⁹ Irene Aldridge (2010), *High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems*, John Wiley & Sons, Inc., 16

³⁰ *Id*

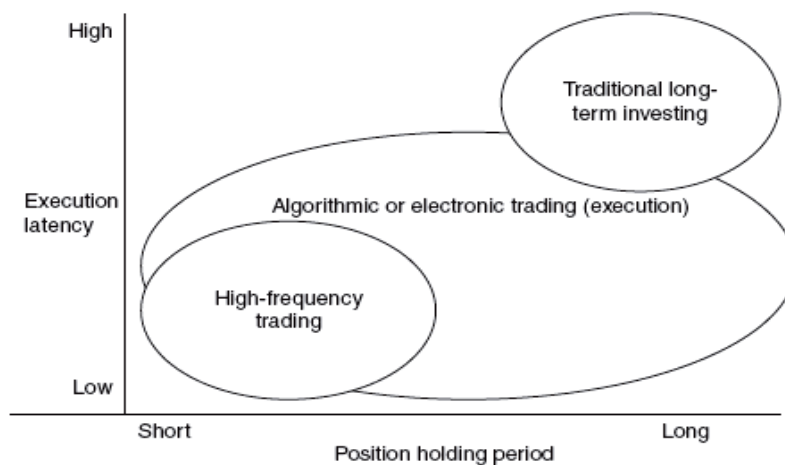


Figure 2: HFT versus AT and traditional long-term investing. Source: Aldridge, 2010

Taking into account these assumptions, we begin to deal with the Algorithmic trading phenomenon in order to frame the starting point of the high frequency trading definition and understand the differences that incur between the two trading techniques.

2.2. Definitions of Algorithmic Trading

The academic literature on Algorithmic Trading (AT), also called “*algo trading*” or “*black box trading*”, is quite extensive due to the importance role that is playing in the trading system.

The definitions range from the very general “*computerized trading controlled by algorithms*” (J. Prial, 2007), or that according to which: “*in algorithmic trading (AT), computers directly interface with trading platforms, placing orders without immediate human intervention. The computers observe market data and possibly other information at very high frequency, and, based on a built-in algorithm, send back trading instructions, often within milliseconds*”. (A. Chaboud, 2009).

A legal definition of Algorithmic Trading (AT) technique, is provided by the article 4(1) (39) of MiFID II³¹ that defines it as: “*a trading in financial instruments where a computer algorithm automatically determines individual parameters of orders such as whether to initiate the order, the timing, price or quantity of the order or how to manage the order after submission, with limited or no human intervention, and does not include any system that is only used for purpose of routing orders to one or more trading venues or for the processing of orders involving no determination of any trading parameters or for the confirmation of orders or the post-trade processing of executed transactions*”.

³¹ MiFID II available at: <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014L0065&from=EN>

Therefore, by definition, especially MiFID, an AT is not solely identified as an automatic execution and electronic routing of orders (like “traditional” electronic trading) but further, trading is carried out by specific algorithms (called *algos*) which, through pre-set trading instructions, automatically determine when, where and how to execute orders in the marketplace without human intervention, so by extremely speeding-up the transactions and by dropping the emotional and behavioral component of the human traders.

Trading algorithms generally specify timing, pricing, quantity and routing of orders by continuously monitoring market conditions with the purpose to promptly react and exploit the current trading situations.

Definitions provide a general classification of several characteristics belonging to algorithmic traders; some of these are familiar for the high frequency traders, too.

2.3. Characteristics of Algorithmic Trading

As high frequency trading is a subset of algorithmic trading family, there are common characteristics that are valid for both trading techniques.

Common characteristics of AT and HFT
1) Used by professional traders
2) No human intervention
3) Pre-designed trading decisions
4) Observing market data in real-time
5) Determining individual parameters of orders
6) Automated order submission
7) Automated order management after submission
8) Direct market access

Table 1: Common characteristics of AT and HFT. Source: Gomber, 2011.

The main common characteristic of HFT and AT is represented by the use of fast computers and complex ad-hoc algorithms to automatically make trading decisions, submit orders, and manage those orders after submission³².

High Frequency Traders (HFTs) and Algorithmic Traders (ATs) are characterized for the employment of informatics devices extremely sophisticated in terms of hardware and software. Software can be “in-house” where HFT and AT firms utilize huge investments in technologies

³² Terrence Hendershott, Ryan Riordan (2011), *Algorithmic Trading and Information*, 2

and informatics engineers in order to implement copyright algorithms that are used exclusively by them; software “tailor-made”, that is designated ad-hoc for HFT and AT firms needs by a third programmers; and software “out of the box” namely commercial applications with very low customization and cheaper compared the others software categories³³. By means of such algorithms, HFTs and ATs can execute complex computations and issue orders toward the trading platforms in split seconds.

They leverage technology and algorithms from end-to-end of the investment chain, from market data analysis and the operation of a specific trading strategy to the generation, routing, and execution of orders and trades³⁴.

In order to be able to elaborate and analyze millions of data, perform appropriate trading strategies and be competitive, algorithmic traders, and subsequently high frequency traders, have continuously needed to heavy investments in the strengthening and updating of their trading algorithms; indeed, it is in place a real “*algos war*” among the various algorithmic traders, where the competition is based on the development of the “perfect” algorithm able to be more profitable with respect to the competitor traders³⁵. The battle for supremacy in algorithmic execution uses an assortment of mathematics, programming, communications and computing hardware capabilities due to regularly adjust algorithmic code to reflect the subtle changes in the dynamic market and to the possible reverse engineering carried out by rival firms³⁶.

As it regards the automated management of orders after submission, it refers that already implemented orders might be modified or cancelled out by ATs and HFTs if in the meantime the underlying’s market conditions have changed or even just for testing the trading reactions of other traders. Such practice is considered one of the most controversial and criticized issue emphasized by the other market participants, who argue that the intensive submission and right after cancellation of orders may cause a significant breakdown of the market stability.

The last important common feature of ATs and HFTs invokes the ability of issuing orders toward the various trading venues via a direct market access and/or a sponsored access by exploiting the technological infrastructure of their brokers, which permits them to execute orders directly without passing to broker and without being a member of the trading venues and then reducing the costs and the timing of trading.

³³ Valeria Caivano, Salvatore Ciccarelli and Giovanna Di Stefano (2012), *Il Trading ad Alta Frequenza: Caratteristiche, Effetti, Questioni di Policy*, CONSOB Discussion Papers No. 5, 8

³⁴ Anuj Agarwal (2012), *High Frequency Trading: Evolution and the Future*, Capgemini, 6

³⁵ Scott Patterson (2013), *Dark Pools: The Rise of A. I. Trading Machines and the Looming Threat to Wall Street*, RH Business Books.

³⁶ Charles Duhigg (2006), *Artificial Intelligence Applied Heavily to Picking Stocks*, N.Y. TIMES

On the other side, there are characteristics of AT that are not commonly associated with the HFT methods.

Exclusive characteristics of AT
1) Agent trading
2) Minimize market impact (for large orders)
3) Goal is to achieve a particular benchmark
4) Trading holding periods of days/weeks/months
5) Executing an order through time and across markets

Table 2: Exclusive characteristics of AT. Source: Gomber, 2011.

ATs are typically Agent traders, it essentially means that they don't trade for their own account purposes but on the contrary, they perform trading activities on their client's behalf in exchange of a trading commission; therefore, ATs are a type of intermediaries that exploit their technological advances, as using the algos, in order to execute their clients' orders granting them the best execution which may be reached in terms of speed and price.

Thus, the principal goal of using non-HFT algorithmic trading techniques concerns the achievement of a specific target set by the client (e.g. buying a determined number of securities shares in a limit time at a specific price) by means of the intelligent implementation of orders through time and across venues thanks to the exploitation of their modern equipment. About that, one of the most used AT activities is the execution of large orders on behalf of institutional investors, such as mutual fund portfolios; specifically, if an investor decides to sell (buy) a large block of stocks and submits the whole block for execution all at once, it should significantly decrease (increase) the price in order to encourage other market participants to absorb the increased supply (demand)³⁷ but alternatively, investor can use algorithms that take the block of stocks and slice it into many small orders that are submitted one by one into different trading platforms for execution.

This trading procedure usually requires to establish or liquidate positions with time horizons of days or even weeks due to the need to alleviate the impact of a large order submission in the market.

Nevertheless, a wide variety of algorithms could be used: for example, some look for arbitrage opportunities, including small discrepancies in the exchange rates between three currencies;

³⁷ James Angel, Lawrence Harris, Chester S. Spatt (2010), *Equity Trading in the 21st Century*, Marshall Research Paper Series, 37-38

some seek optimal execution of large orders at the minimum cost; and some seek to implement longer-term trading strategies in search of profits³⁸.

2.4. Definitions of High Frequency Trading

High-frequency trading (HFT) represents the most sophisticated technological trading technique employed in the financial activities; since it is a phenomenon that is newer with respect to AT, the concerning literature is scarcer and evolving.

There is not a universal and univocal way to define it but it exists a general consensus in the considering HFT as a trading practice (and not a single trading strategy) conducted through supercomputers that give the capability to execute a high number of trades within milliseconds or microseconds³⁹.

More specifically, HFT can be defined as a form of AT characterized by high speeds, high turnover rates and high order-to-rate ratios that leverages high-frequency financial data and electronic trading tools⁴⁰.

US and EU regulators are trying to furnish a more specific identification of the HFT phenomenon, given the great importance that is covering on the securities market landscape. In fact, by most accounts, it has grown substantially over the past ten years; in the early 2000s, HFT still accounted for fewer than 10% of equity orders, but this proportion was soon to begin rapid growth. According to data from the NYSE, trading volume grew up by about 164% between 2005 and 2009 for which high-frequency trading might be accounted⁴¹.

Estimates hold that it accounts for roughly 55% of trading volume in US equity markets and about 40% in European equity markets and more considerably, HFT has grown in futures markets, e.g. foreign exchange and interest rate futures with ceilings of 80% of volume⁴².

U.S. Securities and Exchanges Commission defines High Frequency Trading on its Concept Release on Equity Market Structure⁴³ (2010) as “*a practice that refers to professional traders acting in a proprietary capacity that engage in strategies that generate a large number of trades on a daily basis and characterized by:*

³⁸ Alain Chaboud, Benjamin Chiquoine, Erik Hjalmarsson, Clara Vega (2009), *Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market*, Board of Governors of the Federal Reserve System International Finance Discussion Papers n° 980, 1

³⁹ Rena S. Miller, Gary Shorter (2016), *High Frequency Trading: Overview of Recent Developments*, Congressional Research Service, 1

⁴⁰ See note 29

⁴¹ Charles Duhigg (2009), *Stock Traders Find Speed Pays in Milliseconds*, New York Times

⁴² See note 35

⁴³ SEC Concept Release on Equity Market Structure (2010), 45-46, available at: <https://www.sec.gov/rules/concept/2010/34-61358.pdf>

- 1) *the use of extraordinarily high-speed and sophisticated computer programs for generating, routing and executing orders;*
- 2) *the use of co-location services and individual data feeds offered by exchanges and others to minimize network and other types of latencies;*
- 3) *very short time-frames for establishing and liquidating positions;*
- 4) *the submission of numerous orders that are cancelled shortly after submission; and*
- 5) *ending the trading day in as close to a flat position as possible (that is, not carrying significant, unhedged positions over-night)”.*

Article 4(1) (40) of European MiFID II⁴⁴ describes a high-frequency algorithmic trading (HFAT) technique as “*an algorithmic trading technique characterized by:*

- 1) *infrastructure intended to minimize network and other types of latencies, including at least one of the following facilities for algorithmic order entry: co-location, proximity hosting or high-speed direct electronic access;*
- 2) *system-determination of order initiation, generation, routing or execution without human intervention for individual trades or orders; and*
- 3) *high message intraday rates which constitute orders, quotes or cancellations”.*

Both MiFID and SEC attribute to HFT specific features that are not present in the definition of AT; such exclusive characteristics represent the main hallmark and competitive advantage of HFT with respect to other traders.

2.5. Characteristics of High Frequency Trading

In order to better understand the definitions of HFT provided by regulators, it is necessary to extrapolate the main characteristics specifically related to HFT identification.

They can be summarized in:

Specific characteristics of HFT
1) Proprietary trading
2) Low-latency requirements
3) Use of co-location/proximity services and individual data feeds
4) Very high number of orders
5) Very short order holding periods (seconds/minutes)

⁴⁴ See note 31

- | |
|---|
| <ul style="list-style-type: none"> 6) Profit for buying and selling 7) Very low margins per trade 8) Focus on high liquid instruments 9) No overnight positions (flat position) 10) Very high order-to-trade ratio |
|---|

Table 3: Specific characteristics of HFT. Source: Gomber, 2011.

HFT deeply diverges with respect to the more general AT, either in terms of technology adopted and trading strategies used.

Unlike ATs, the most of HFTs are Proprietary traders: they are private firms utilizing only their own capital for their trading activities⁴⁵. This implies that the trading strategies of the HFTs are not associated to the will of the client but they are defined by themselves with the only purpose to profit.

The fact that this trading practice is properly called high frequency trading entails that HFT activities are based on a huge number of orders issued in the marketplace in a unit of time (which can reach to more than 5,000 per second)⁴⁶. For this reason, although HFTs are a minority of total traders, HFT constitutes most of the total trading volume of the venues where it is employed; in the United States in 2009, HFT firms represented 2% of the approximately 20,000 total firms operating, but accounted for 73% of all equity orders volume⁴⁷.

HFTs are not only identified by the high volumes and high speed of executions, but also by the high percentage of modifications and cancellations of orders; HFT management after order submission is significantly more aggressive compared non-HFT firms, due to the more frequency of such adjustments. The increasing number of cancellations of orders already submitted leads to a high order-to-trade ratio, that represents the percentage of how many orders are issued with respect to the number of orders effectively executed. Estimates claim that almost all the submitted orders by HFT systems do not reach an actual execution; only 1% of the total orders would result in a trading contract⁴⁸.

Another typical characteristic of HFT activities is constituted by the engagement of trading positions on financial instruments for a holding period that does not exceed the trading day⁴⁹; HFTs prefer to liquidate their entire portfolios on a daily basis rather than carrying positions overnight especially for two reasons: the former is represented by the high levels of risk of

⁴⁵ Larry Harris (2003), *Trading and Exchanges: Market microstructure for practitioners*, Oxford University Press

⁴⁶ Valeria Caivano, Salvatore Ciccarelli & Giovanna Di Stefano (2012), *Il Trading ad Alta Frequenza: Caratteristiche, Effetti, Questioni di Policy*, CONSOB Discussion Papers No. 5, 8

⁴⁷ Rob Iati (2009), *The Real Story of Trading Software Espionage*, The TABB Group

⁴⁸ Andrea Greco (2011), *Giro di Vite Consob sul Trading ad Alta Frequenza*, la Repubblica, 4 November

⁴⁹ See note 46

overnight positions because of the more volatility favored by the extension of global trading activities to 24-hour cycles; while the latter is concerning to the potential smaller profitability due to the interest that has to be paid, referred to as an overnight carry rate (usually slightly above LIBOR rates), in order to maintain overnight positions⁵⁰.

HFTs move in and out of short-term positions at high volumes and high speeds (high-turnover rate) looking at capturing only a fraction of a cent in profit on every trade⁵¹ but considering that they trade with an exorbitant frequency, even a very small but almost always present profit per trade can result in a substantial overall gain.

The International Organization of Securities Commissions (IOSCO)⁵² noted that those engaging in HFT typically utilize strategies for both buy and sell trades in extremely large number of stocks, options, futures, currencies, exchange trade funds, as well as other financial instruments but HFTs show a preference on trading in a very liquid markets and financial instruments; in effect, the HFT operations require the possibility to close the holding positions very quickly. The more liquid financial instruments are those from which you can rapidly disinvest because have a market capable to absorb large orders with a little economic impact⁵³. Summarizing, the main characteristics of HFT and related success factor concern the ability of performing its trading operations with an extremely speed and consequently with a high frequency which are fundamental components of all HFT strategies, but how are HFTs able to be very much faster than other types of traders? The answer is recognized in the so-called low-latency requirements, that are fulfilled in particular thanks to the infrastructure employed by HFTs, both in terms of hardware/software capabilities and ultra-low latency network systems, in order to minimize as far as possible, the timing of incoming, elaboration of information and the execution and the transfer of orders up to time intervals of milliseconds and even microseconds.

2.5.1. Low-Latency Requirements

In the capital market context, latency is referred as the necessary time to the implementation of a series of operations needed to transform an economic decision in an actual trading (order execution)⁵⁴.

⁵⁰ See note 29, 2

⁵¹ *Id*

⁵² Regulatory Issues Raised by the Impact of Technological Changes on Market Integrity and Efficiency, IOSCO (2011) available at: <http://www.iosco.org/library/pubdocs/pdf/IOSCOPD354.pdf>

⁵³ See note 46, 8

⁵⁴ Alfonso Puorro (2013), *High Frequency Trading: una panoramica*, Questioni di Economia e Finanza n° 198, Banca d'Italia, 10-11

There are several aspects of the trading process in which the latency has to be taken into consideration:

- i. the speed with which HFTs receive the information from the market and process it (analysis of the data and operational reaction). For example, the ability to analyze in real time the bigger possible amount of data and be able to transform the flow of information in investment choices;
- ii. the time that elapses between the processing of the data and the transmission of the order to the broker;
- iii. the time between receiving the order by the broker and sending it to the trading venue;
- iv. the period that the order takes to arrive on the market by the time it is released by the broker;
- v. the time between the reception of the data by the market and the disclosure of the data to all participants in the market⁵⁵.

A high-frequency system requires the shortest possible time to cover all the steps of the trading process; a time-interval that the technological process has led until few milliseconds (low-latency).

The first step of the trading process is accomplished by high-speed computers and complex algorithms that enable of rapidly compiling market information and making instant decisions with no real-time human intervention.

HFT firms use modern multi-core processors which make data processing much faster as there are exactly several processors working on different tasks at the same time, resulting in a significant increase in the overall system speed⁵⁶.

In addition, HFTs avail themselves of totally automated algorithms which, if well-designated, are able to take trading decisions in the same time as the information is analyzed.

As we have seen in the first chapter, HFTs can overstep the third and the fourth point of the trading process through the direct market access or the sponsored access granted by their broker and therefore by minimizing the concerning latency.

The last issue of the latency process, associated to HFT liability, regards the time that elapses from the trading decision to the transmission of order toward the trading venue. It objectively depends by two physical factors: the distance between the two points of connection and the transfer speed.

⁵⁵ *Id*

⁵⁶ *See* note 29, 13

In order to minimize this type of latency, HFTs have improved their network systems and have co-located their proprietary servers.

There is a physical limit of reducing the distance latency given by the speed of light, about 300,000 kilometers per second or equivalently 3 microseconds per kilometer, that expresses the theoretical top speed reachable in nature. Such transmission speed is potentially accessible and approachable considering a linear and without physical obstacles data travel from a starting point to an arrival point.

About that, common network connections utilized by the normal users are not suitable for HFT purposes inasmuch the tortuous route needed to connect all the private houses would slow down the transmission speed and it may run into frequent connection stability problems; for this reason, HFTs have used huge investments in ultra-low latency private infrastructures able to connect directly their computer servers towards the several trading venues located worldwide. Michael Lewis on its book “Flash Boys”⁵⁷ tells about the construction of an important network system infrastructure called “Spread Networks” that aimed at offering internet connectivity between Chicago and New York at ultra-low latency (very close to the speed of light), high bandwidth and high reliability using dark fiber, privately operated fiber network that connects directly point-to-point.

“Spread Networks” was installed in 2009, running 827 miles from Chicago (home to Chicago Mercantile Exchange, where futures and options are traded) to Carteret, New Jersey (home to the NASDAQ data center) along a route as close to straight as possible; the estimated round-trip along the dark fiber line from Chicago to Carteret would be reduced from 17 to 13 milliseconds. By this project, HFTs engaging such network could enjoy small but important reductions in latency that helped them to close trades before their competitors, in particular through low-latency arbitrage strategies, wherein searching out price discrepancies between future contracts in Chicago and their underlying equities in New York.

Another important aspect, regarding latency network, is concerning the necessity to plot out certain optimal points between geographically separate exchanges, in order to obtain the least amount of time for information to travel between them. For instance, the optimal point to exploit the price difference between the NYSE and the LSE was found to be a spot in the mid-Atlantic Ocean. According to some researchers and given that infrastructure in the form of undersea data cables already exists, such floating trade centers could be a possibility in the future⁵⁸.

The other fundamental component of reducing latency which is peculiar of the HFTs entails the so called “co-location”.

⁵⁷ Michael Lewis (2014), *Flash Boys*, W. W. Norton & Company

⁵⁸ See note 29, 13-14

Intuitively, transfer latency does not only depend by the engineering used to transfer information the fastest possible but the travel time is also influenced by the distance among two connection points. This means that in a hypothetical race between two operators using the same private cable network and competing in the same market for the same instruments, the one located closest to the trading venue server will always arrive first with respect to the farther one.

Hence, in order to be faster than competitors, HFT firms have decided to locate own computer servers (or their broker server) in the same premises where an exchange's computer servers are housed; for example, Wall Street servers are located in Carteret, NJ and in the proximity many HFT firms have rent a slot called "rack" where positioning their proprietary servers⁵⁹. This enables HFT firms to access into trading book of the exchange where are co-located always before the rest of the investors.

Co-location has become a very lucrative business for some exchanges, which charge HFT firms millions of dollars for the privilege of "co-location access". For the huge demand of co-location, some stock exchanges have expanded their data centers substantially, for instance the NYSE Euronext data center in Mahwah, NJ is almost 398,000 square feet compared to 46,000 square feet of the old NYSE building⁶⁰.

HFT firms can exploit the proximity factor even contemporaneously in more different exchange platforms around the world as in NY, London, Singapore through a multiple co-location infrastructure.

If the trading platform does not offer this type of commercial service, a third party can substitute such business via a "central proximity hosting", in which the hired spaces don't belong to the trading venues but in the same way are located very close to them⁶¹.

2.6. How to Identify High Frequency Traders?

From an analytical perspective, the absence of a unique definition of HFT phenomenon makes it difficult to achieve a precise identification of HFT firms. The literature employs a number of approaches to identify HFTs leading to different results in assessing the level of HFT activity in the securities markets.

We describe different approaches used in the literature. These fall into two broad categories:

- 1) direct approach;

⁵⁹ See note 46, 10-11

⁶⁰ NYSE website, "Co-location: Nyse Euronext's US Liquidity Center"

⁶¹ See note 52

2) indirect approach⁶².

The direct approach to identify HFTs relies on the identification of market participants either based on:

- their primary business; and/or
- the use of services to minimize latency⁶³.

The information on HFT activity as primary business is obtained by the trading venues in which market participants operate⁶⁴.

Such method focuses on pure HFT firms which are flagged by trading venues as HFT firms and it does not cover HFT activity carried out by other non-HFT firms, such as investment banks.

Furthermore, primary business method may not include HFT activity by HFT firms that exploit another trading venue membership (direct market access and sponsored access), unless the broker reports the HFT firms as clients⁶⁵.

These considerations translate on an underestimation of HFTs because HFT activities by other non-HFT are not counted and on the other side, it could also exist an overestimation of the phenomenon inasmuch as not all activities carried out by HFT firms may be HFT; however, it is more likely that the underestimation element is dominating.

The second direct method relies on information about the use of low-latency infrastructures, e.g. the use of co-location and proximity services or access to fast data feeds⁶⁶.

Such method does not require any knowledge of the firm's primary business but the HFT identification is only based on the use of services to minimize latency. Incorrectly, this method could include brokers whom trading as agent traders who may use co-location services to offer best execution strategies to their clients; therefore, by basing only on this parameter, it would result on an overestimation of the HFT firms⁶⁷.

Conversely, the indirect approach used to identify HFT firms is based on the operating trading activities performed by market participants.

Examples related on these trading activities are identifications regarding:

- intraday inventory management;
- lifetime of orders;
- message traffic (including order-to-trade ratios);

⁶² Antoine Bouveret, Cyrille Guillaumie (2014), *High-Frequency Trading Activity in EU Equity Markets*, ESMA Economic Report, 6

⁶³ *Id*

⁶⁴ *Id*

⁶⁵ *Id*

⁶⁶ *Id*

⁶⁷ *Id*

- identification of strategies⁶⁸.

Each indirect approach does not have to be taken separately because all reported characteristics are useful to identify HFT activity as a whole.

A peculiar characteristic of HFT refers to the maintaining of an average flat position during all the trading day (i.e. HFTs are neither net sellers or net buyers) and the detention of no-overnight inventories; an indirect identification of HFT is the estimation of intraday inventory management through mock-up data. A market participant will be recognized as HFT firm whether utilizing an intraday inventory management compliant to HFT features.

Nevertheless, identification based on intraday inventory management will tend to identify some HFT strategies (market-making) and may not identify other HFT strategies⁶⁹.

An alternative indirect method consists in looking at the lifetime of orders, i.e. holding period of positions before the order is executed, modified or cancelled.

As reported, HFTs are characterized by the rapidly discharge of opened positions in seconds/minutes compared to other traders. Classification based on lifetime of orders finds several problems such as the difficulty to analyze each order management for all securities traded and for all traders and the definition of a calibrated timing threshold that characterizes HFT activity.

Another proxy is based on OTR ratio but it is rather a measure of message traffic than a measure of HFT because it strongly depends by the different strategies adopted by the HFTs.

The last indirect method of identification is related to the strategies used by HFTs but it is less suited to identify the overall level of HFT activity in a market given the very different types of strategies adopted by HFTs but it may be useful on the determining the distinction between for example market making HFT strategies and opportunistic ones (momentum strategies) by identifying the HFT firms who contribute to market stability⁷⁰.

⁶⁸ *Id*

⁶⁹ *Id*

⁷⁰ *Id*

3. HIGH FREQUENCY TRADING STRATEGIES

3.1. Introduction to HFT Strategies

General information about HFT techniques is not sufficient to give a fair judgement of the behavior of HFT activity. Considering only HFT characteristics from a general perspective, it does not provide a relevant opinion on the effect of HFT phenomenon on capital markets.

For this purpose, it is necessary to identify all the possible trading strategies implemented by HFT operators, inasmuch HFT practice does not represent a unique trading strategy but it constitutes a technological mean used to employ a very diverse range of strategies.

Why is it important to distinguish among different trading strategies performed by HFTs?

Concept Release on Equity Markets of SEC⁷¹ explains that the implementation of a specific strategy implies a diverse effect on market structure; it may be harmful or beneficial for market structure performance and for interests of long-term investors, too. It is possible to reliably identify strategy implications through, for example, metrics such as adding or taking liquidity, or trading with (momentum) or against (contrarian) prevailing price movements.

HFT trading strategies are so heterogeneous because there are different traders utilizing HFT techniques, which usually have very sophisticated ad-hoc algorithms and technology specialized to specific trading strategies.

There is a multitude of different institutions with different business models that use HFT and there are hybrid forms, e.g. broker-dealers which run their proprietary trading books applying HFT techniques⁷². Hence, in the assessment of HFT, it is very important to take a functional rather than an institutional perspective because the main purpose remains the evaluation of the HFT methods, and not of the different types of users⁷³. To achieve a level playing field, all operators that use HFT based strategies should be taken into consideration independent of whether HFT is their core or an add-on technology to implement trading strategies⁷⁴; it includes HFT proprietary firms, investment banks and hedge funds leveraging HFT technology to profit. Despite it exists a very wide range of trading strategies used, the common element of all of them is represented by the HFT ability to issue, modify and cancel out thousands of trading orders at very high speed and by the timing advantage of the low-latency component which allows them to arrive first both in terms of data information and market access.

⁷¹ SEC Concept Release on Equity Market Structure (2010), 46-47

⁷² Peter Gomber, Björn Arndt, Marco Lutat, Tim Uhle (2011), *High-Frequency Trading*, Goethe Universität Frankfurt Am Main, 24

⁷³ *Id*

⁷⁴ *Id*

In fact, HFTs can always beat the competitors regarding the speed and can implement ad-hoc trading strategies made possible by the time priority⁷⁵.

Another important feature of HFT is constituted by the nearly continuous trading activity, which permits to high frequency operators to adapt their systems to the different market developments in real-time and to take promptly decisions about issuing, cancellations and modifications of orders and consequently adopt different trading strategies⁷⁶. For instance, HFT algorithms can timely assess the liquidity of the market and the associated trading risk in order to decide the steps to follow for an appropriate trading strategy.

HFTs may act both actively as “price taker” and passively as “liquidity provider”, that is they can issue immediately marketable orders and orders that are currently not matched but positioned on the trading book waiting for matching.

By means of these considerations, there are many different trading strategies adopted by HFTs; all strategies exploit HFT infrastructure with the aim to make profit but they can have opposed effects in terms of market liquidity intended as the cumulated combination of 1) proportion of buying and selling orders on the trading book (limit orders); 2) proportion of incoming buying and selling orders crossing orders on trading book (market orders) and 3) volumes associated with buying and selling orders⁷⁷. Liquidity instability would result in unfair prices and high volatility.

We start on analyzing algorithmic trading strategies, even if not all AT strategies are necessarily high frequency but it is useful to frame the nature of algorithmic methods employed to perform trading.

3.2. Algorithmic Trading Strategies

AT techniques are an instrument used by professional traders to execute trades of other traders granting them the best execution; then, AT strategies do not pursue a mere profit goal but algorithms are ad-hoc pre-set by ATs for executing precise trading instructions drafted by the client.

As we have seen, most of non-HFT algorithmic strategies aim at minimizing the market impact of large orders by slicing them into several smaller orders and by spreading out across time and venues defined by a pre-set benchmark.

⁷⁵ Alfonso Puorro (2013), *High Frequency Trading: una panoramica*, Questioni di Economia e Finanza n° 198, Banca d'Italia, 12

⁷⁶ *Id*

⁷⁷ See Puorro *supra* note 75, 13

According to a classification drawn up by Almgren (2009)⁷⁸ with further information from Johnson (2010)⁷⁹, there are four generations of algorithms.

3.2.1. First Generation Execution Algorithms

First-generation algorithms focus exclusively on benchmarks that are set up on market generated data and are independent from the actual order book situation at order arrival⁸⁰. In this category, we recognize the following algorithms:

- *Participation Rate Algorithms*: they are programmed to participate in the market up to a predetermined volume; for instance, an algorithm tries to participate by trading up to 5% of the volume in the target instrument, with the purpose large orders harder to be detected for other traders, until it has built or liquidated the desired position⁸¹. Therefore, such algorithms reflect the current market volume in their orders. Variants of these algos may add execution periods during which orders are submitted to the market or maximum volumes or prices.
- *Time Weighted Average Price (TWAP) Algorithms*: they split a large order into slices that are sent to the market in equally distributed time lapses⁸². For example, the algorithm could set to buy 6,000 shares within one hour in blocks of 1,000 shares, resulting in 6 orders for 1,000 shares every 10 minutes.
- *Volume Weighted Average Price (VWAP) Algorithms*: they try to match or beat the VWAP, a price benchmark defined as the average trade price of a specific period where each trade price is weighted by the size of the associated trade and calculated following the formula: $VWAP = \frac{\text{Dollar (Euro) Volume}}{\text{Total Volume}} = \frac{\sum_n v_n p_n}{\sum_n v_n}$ where p_n is the execution price and v_n is the size, over a specified time interval⁸³. Large orders have a great impact on VWAP given that trades are being weighted per their size. VWAP algorithms are constructed basing on the historical volume profiles to estimate the target period volume patterns.

By resuming, the first-generation algorithms are used aiming at reaching a specific trading benchmark, such as volume, trading period and price as figure 3 illustrates.

⁷⁸ Robert Almgren (2009), *Quantitative Challenges in Algorithmic Execution*, Working Paper

⁷⁹ Barry Johnson (2010), *Algorithmic Trading and DMA: An Introduction to Direct Access Trading Strategies*, 4Myeloma Press, London

⁸⁰ See note 72, 21

⁸¹ *Id*

⁸² See note 72, 22

⁸³ Larry Harris (2003), *Trading and Exchanges*, Oxford University Press

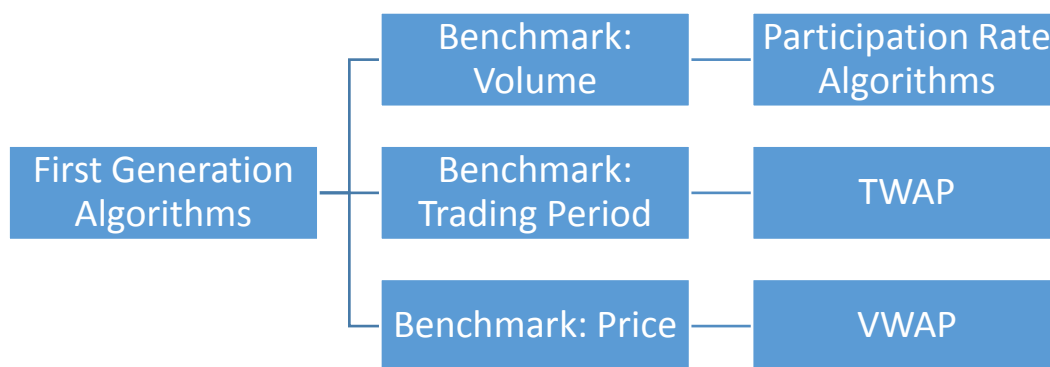


Figure 3: First Generation Algorithms.

3.2.2. Second Generation Execution Algorithms

Second-generation algorithms generally are used to minimize the implementation shortfall indicator⁸⁴.

The quotation midpoint at the time of decision to trade is the price benchmark that algorithms should match or beat.

Implementation shortfall algorithms attempt to minimize the market impact of a large order execution in terms of potential negative price movements during the execution process (timing risk). To hedge against this type of risk, these algorithms predetermine an execution plan based on historical data, and split the order into as many as necessary but as few as possible sub orders⁸⁵.

Sub orders will be scattered over a period which is just long enough to dampen the market impact of the overall order, much shorter than TWAP and VWAP needs⁸⁶.

There is a trade-off between minimizing market impact and timing risk because the necessary time to alleviate the market impact could be long and vice versa, whether the timing is short, the market impact of order is bigger.

3.2.3. Third Generation Algorithms

Third-generation algorithms are more sophisticated than the first and second generation.

Almgren defines them as *adaptive* algorithms because they follow an approach based on re-evaluation and adjustment of their scheduling processes during the execution phase and are not only determined by pre-set programs. This procedure makes such algorithms adaptive to changing market conditions.

⁸⁴ See note 72, 23

⁸⁵ *Id*

⁸⁶ See note 79

3.2.4. Fourth Generation Algorithms

Fourth-generation algorithms, also called *newsreader algorithms*, are very different than previous categories because are not related to the minimization of the market impact for large orders but they are more closely connected to high frequency algorithmic strategies.

These newsreader algorithms have been designed to detect, through statistical methods and text mining techniques, the impact of news announcements on the market⁸⁷. They rely on high-speed market data, further favored by low-latency news feeds provided by exchanges and news agencies⁸⁸.

Non-HFT algorithmic trading strategies are relatively easy to predict and detect, and consequently other participants may take their relative countermeasures to contrast them; HFTs have adopted specific strategies to take advantage of predictable algorithms.

Unlike non-HFT algorithmic traders, HFTs have developed and engaged other trading strategies with respect to simple use of algorithms.

3.3. High Frequency Trading Strategies

The potential number of trading strategies used by HFTs is almost infinite; in fact, HFT represents a powerful instrument in the hands of HFT firms that permits the engagement of very heterogeneous strategies to achieve their profit purposes.

HFT strategies are so various as trading algorithms are as well; strategies reflect the behavior of algorithms that are constantly updated to exploit new economic conditions and then they may assume different conducts depending of current market events.

However, the basis point of all HFT strategies is associated to the employment of their low-latency specific attributes and ad-hoc algorithms necessary to implementation of their awesome speed advantage translated in very fast order execution at a very high frequency.

Basically, most of HFT debate strategies are nothing new and are well known in the market landscape, even if they are being executed using better technology and at greater speed⁸⁹.

Figure 4 shows a classification of strategies carried out by HFTs which is based on complementary information from Concept Release on Equity Markets Structure of SEC (2010), Gomber et al. (2011), and Puorro (2013).

⁸⁷ See note 72, 24

⁸⁸ *Id*

⁸⁹ Sylvain Friederich and Richard Payne (2011), *Computer Based Trading, Liquidity and Trading Costs*, Government Office for Science, 7

Concept Release on Equity Markets Structure identifies four macro-categories of trading strategies employed by HFT firms corresponding to: 1) liquidity providing; 2) (statistical) arbitrage; 3) directional and 4) structural strategies.

In addition, all macro-categories have several sub-categories that may represent very different ways of doing trading with specific effects both in terms of HFT profits and resulting market impact; therefore, it is important to describe each of them to provide a fair assessment of all types of performed strategies.

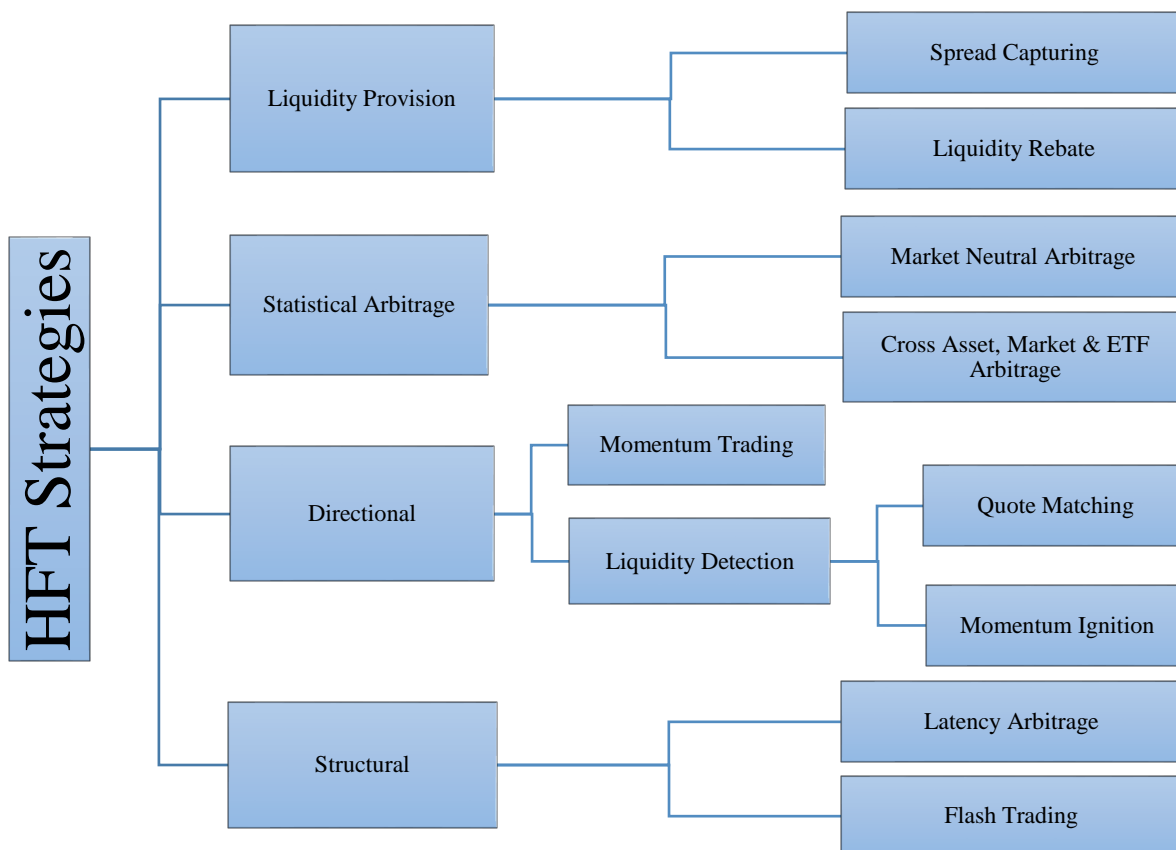


Figure 4: High Frequency Trading Strategies.

3.3.1. Liquidity Providing Strategy

One of the most common HFT strategies is to act as liquidity provider or passive trader.

HFTs can replicate the activities of traditional market-makers, i.e. they place orders to buy (as well as sell) using limit orders that are above (below) the current market price in the case of selling (buying). These orders are called non-marketable because are resting orders that provide liquidity to the marketplace at specified prices.

The great difference with respect to traditional market-makers is that HFTs are not bound to be registered as market-makers of the exchanges and therefore subject to corresponding market-making obligations, such as the continuous two-sided quotes, minimum quantity, maximum

spread and quotation timing⁹⁰; consequently, they can freely choose how to be counterpart of incoming orders and when and whether to operate. About that, it is in place a debate whether HFT firms engaging in market-making activities should submit to traditional market-makers' obligations, also urged by MiFID II that requires to market-making HFTs to undergo the related obligations.

HFT liquidity providers have two basic sources of revenues:

- i) profits from the spread between bid and ask quotations (*spread capturing*); and/or
- ii) incentives from trading venues in the form of reducing transaction fees or even by granting “rebates” (*rebate driven*)⁹¹.

3.3.1.1. Spread Capturing

Spread Capturing is a HFT strategy that is closer to traditional market-making activity inasmuch liquidity provision core activity guarantees a profit from spread between bid and ask prices by continuously buying and selling securities on both sides of the trade⁹². Earnings are provided by lower prices at which market-makers can buy and by higher prices at which they can sell. HFTs do this automatically, by inputting the order limits into their algorithms and letting the computers do the work⁹³.

In this type of activity, HFTs, by issuing thousands of orders on daily basis, will gain a lot of money even if the spread is minimal in fact they are called “*scalpers*”. In addition, low-latency advantages permit to ensure the time priority on trading books, by obtaining that their quotes are inserted systematically before and consequently executed sooner of non-HFT market-makers.

Since they are not subject to market-making obligations, HFT market-makers can freely abandon the market in turbulence periods of high volatility by deeply reducing the associated risk. In fact, in the case of sudden rising of volatility, HFT systems can cancel out their orders very quickly, and reduce the potential losses.

3.3.1.2. Rebate Driven

Rebate Driven strategies are built around particular incentive schemes of certain exchanges and ECNs. To attract liquidity providers and react to market competition resulting by the

⁹⁰ See note 75, 14

⁹¹ See note 72, 25

⁹² Michael J. McGowan (2010), *The Rise of Computerized High Frequency Trading: Use and Controversy*, Duke Law & Technology Review No. 16, 11

⁹³ *Id*

fragmentation of securities markets, some trading venues have adopted asymmetric fees: members removing liquidity from the market (price takers or aggressive traders) are charged for a higher fee for trading, while traders who provide liquidity to the market (price makers or passive traders) are charged for a lower fee or even receive a rebate for trading⁹⁴.

This incentive scheme is finalized to promote the liquidity provision in the marketplace. More competition in liquidity provision implies a cutting down of bid-ask spread and then also of the profits for market-makers on spread capturing; on the other side, liquidity providers will receive a rebate for their liquidity service that will compensate the spread profit reduction. Figure 5 shows the just mentioned revenue sources for HFT liquidity providing strategies.

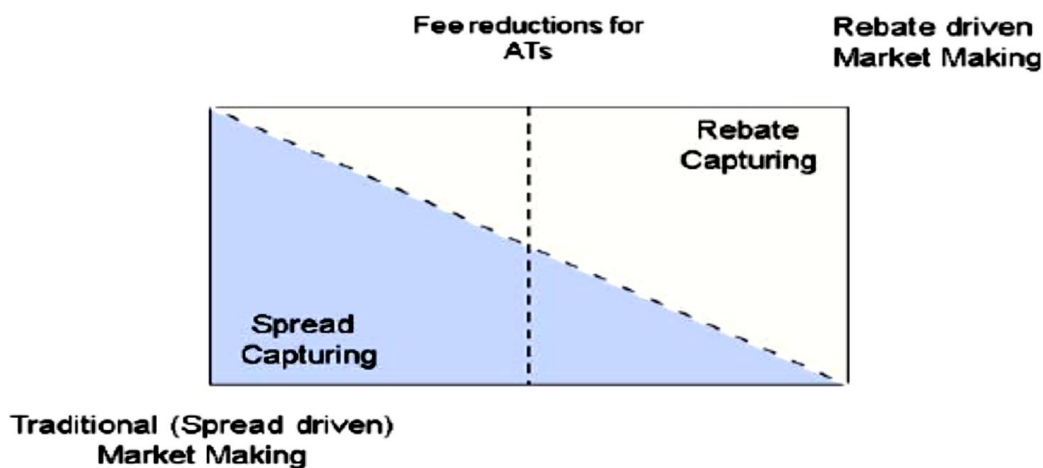


Figure 5: Revenues sources for HFT liquidity providing strategies. Source: Gomber, 2011.

3.3.2. Statistical Arbitrage Strategy

Arbitrage is the practice where a trader takes advantage of price differences of identical or similar financial instruments on different markets or in different forms. Such price differences may occur for different trading reasons but they represent a form of market inefficiencies.

Price discrepancies among identical instruments do not find an economic reason except for a temporary inefficiency in the markets where they are traded and despite the economic theory claims that such inefficiencies are rare and immediately adjustable, opportunities to conduct arbitrage frequently exist but typically are present only for very short periods (fractions of a second) in the modern capital markets⁹⁵.

The presence of several Exchanges and specifically, Alternative Trading Systems (or Multilateral Trading Systems), where same securities are traded, has potentially increased the

⁹⁴ See note 92

⁹⁵ See note 75, 13

possibility to perform arbitrages, since the odds to have different trading prices, even for tiny periods, between correlated instruments is very high.

Statistical Arbitrage is a popular strategy among market participants which is related to the statistical mispricing of one or more assets based on the expected value of these assets calculated through a computational and empirical approach⁹⁶.

Statistical arbitrageurs try to identify similar instruments (depending on common fundamental valuation factors) that are inconsistently priced relative to each other according to their expected value patterns and, once identified, they buy the cheaper and sell the more expensive one⁹⁷. They profit if the cheaper instrument appreciates and the expensive one depreciates, if the cheaper instrument appreciates faster than the expensive one, or if the expensive instruments depreciates faster than the cheaper one⁹⁸.

High frequency systems can recognize securities mispricing through their information technologies without a human intervention and the rapidity factor enables the identification and subsequent exploitation of the arbitrage before other non-HFTs.

Sub-categories of statistical arbitrage are identified in market neutral and cross arbitrages.

3.3.2.1. Market Neutral Arbitrage

Market neutral arbitrage strategy implies the contemporaneous holding of long positions in assets, which are considered undervalued and, at the same moment, of short positions in closely related assets, which are perceived overvalued⁹⁹. The detention of opposed holding positions allows an overall offsetting of market movements, that is, gains and losses compensate each other by obtaining a so called “*delta neutral portfolio*”¹⁰⁰. When prices of correlated instruments normalize in the estimated expected values, HFT firm liquidates the positions to obtain a profit. Unlike traditional pair trading, HFTs use market neutral arbitrage with a portfolio of a hundred or more instruments, that are carefully selected by common broad elements and constantly monitored by algorithmic trading models that apply a frequent portfolio turnover to eliminate the exposure market risks.

3.3.2.2. Cross Asset, Market & ETF Arbitrage

⁹⁶ See note 83

⁹⁷ *Id*

⁹⁸ *Id*

⁹⁹ See note 75, 27

¹⁰⁰ *Id*

Cross Arbitrage is referred to profit from price discrepancies among assets and markets. With current market fragmentation, an instrument is likely traded in more trading venues and each of them can have a specific price. Cross market arbitrageurs can generate profit by buying the asset in the venue where it is valued lower and simultaneously selling it on another venue where it is priced higher¹⁰¹. Likewise, arbitrageurs can profit from inefficiencies across assets: for instance, if an option is priced too high with respect to its underlying, they can sell the option and buy the underlying. In the same way, arbitrage is applicable to exchange trade funds by exploiting price discrepancies between ETF and its underlying.

3.3.3. Directional Strategy

Directional strategies generally involve establishing a long or short position in anticipation of an intra-day price movement of a particular direction, upward or downward¹⁰².

Nevertheless, there may be a wide variety of short-term strategies that anticipate such a movement in prices, because typically the up or down of prices may occur for different trading reasons.

Directional traders usually trade aggressively, by taking market liquidity, and aim at earning profits from short market movements/trends.

Some directional strategies may be as straightforward as concluding that a stock price temporarily has moved away from its “fundamental value” and establishing a position in anticipation that the price will return to such value¹⁰³ (sentiment oriented trading); conversely, other strategies may be parasitical, by anticipating other traders that will affect prices such as large traders (front-running trading).

3.3.3.1. Momentum Trading

Momentum Trading strategies look for sudden movements of trading prices in a specific direction; momentum trader can foresee the future pricing trend through a technical or news analysis.

Momentum Trading on Technical identifies patterns which indicate that prices differ from their fundamental value; patterns come from quotes, prices and volumes.

¹⁰¹ See note 75, 28

¹⁰² See note 71, 53

¹⁰³ *Id*

Momentum Trading on News is the practice represented by the possibility to take advantage of the effect that news and macroeconomic data may exert on the behavior of financial instrument prices¹⁰⁴.

The fundamental requirement for performing such strategy is constituted by the need of disposing of informatic systems and algorithms, such as *newsreader algorithms*, able of analyzing thousands of data and, at the same time, able of taking the relevant trading operations. In fact, high frequency systems can associate to the incoming news and market data, specific short trading effects, such as the related trading impact on securities markets in terms of increasing or decreasing prices and volatility¹⁰⁵; consequently, HFT momentum traders can adopt specific strategies, before the development of the directional movement, with profit purposes.

This strategy relies more on short-term movements rather than fundamental values of stocks and hence, momentum trader attempts to assess the future micro-trend of a stock, starting from news and/or technical analysis, and depending on predicted price direction, it decides to take an advance long/short position in order to trade out as soon as the prices will go up/down.

3.3.3.2. Liquidity Detection Strategy

An important type of directional strategies is acting as “*liquidity detector*”.

Liquidity detection strategy aims at flushing out order patterns that other market participants, as passive traders, leave in the market, specifically that ones, once launched, allow HFTs to make profit and/or to hedge their trades by being counterpart¹⁰⁶.

Liquidity detector traders act as price takers because they benefit from liquidity that other traders offer in the market.

More specifically, liquidity detection refers to the activity where algorithmic traders analyze through specific mechanisms some key levels detected by algorithms in order to check the presence of particular types of orders in the trading books¹⁰⁷. These orders are typically: stop orders (orders with a stop instruction to stop orders execution until price reaches a stop price specified by the trader) and large orders (orders considered too large to be executed all at once because inclined to affect the liquidity of the market).

¹⁰⁴ See note 75, 16

¹⁰⁵ *Id*

¹⁰⁶ See note 72, 28-29

¹⁰⁷ See note 75, 18

Liquidity detectors are specialized in discovering slices of large orders which may be in the form of hidden orders (orders not fully displayed) and/or orders submitted by an algorithmic execution (such as first-generation algorithms).

Stop orders, as stop loss orders or take profit orders, have a significant disadvantage represented from being easily predictable; for this reason, HFTs, through a technical analysis, can detect the presence and behavior of such orders and whereupon they can take their associated liquidity at low prices¹⁰⁸.

Liquidity detection systems are also called “*algo-searchers*” because their target is to test the presence of such orders, through the issuing of market orders, to trigger their execution and therefore by activating their liquidity on the market.

HFT liquidity detectors can gather further orders information typically by “sniffing out” other algorithms, “pinging” in order books or dark pools¹⁰⁹.

For instance, a pinging strategy permits to discover the full sizes of undisclosed limit orders by submitting large marketable orders with fill-or-kill instructions attached that trade if the size is present, and if it is not, their orders are cancelled out¹¹⁰.

3.3.3.2.1. Quote Matching

A possible way to exploit the liquidity detection of large orders is by means of a “*quote matching*” practice described by Harris (2003).

Quote matchers are front-runners who trade in front of large traders, that they have detected, to try to extract option values of their large orders¹¹¹.

If the detected large order is a limit buy order, quote matcher places its order on the same side with higher limit to position ahead on the limit order book¹¹².

Quote matching strategy permits to profit, if prices move upwards, by the full extent of the price changes; and otherwise if prices fall, quote matcher can hedge its opened position by being counterpart of the large limit order.

Quote matchers profit at expenses of passive large traders: they take the liquidity that otherwise would have gone to the large traders.

This strategy is parasitic trading because HFT quote matchers profit only when they can prey on other traders (large traders); they do not make prices more informative, and they do not make

¹⁰⁸ *Id*

¹⁰⁹ *See* note 102

¹¹⁰ *See* note 83

¹¹¹ *Id*

¹¹² *Id*

markets more liquid and some applications of order anticipator strategies are considered illegals.

3.3.3.2.2. Momentum Ignition

One of the most sophisticated liquidity detection techniques is represented by the so called “momentum ignition”.

Momentum ignition is a strategy in which a HFT firm initiates a series of orders and trades (along with perhaps spreading false rumors in the marketplace) aimed at causing a rapid price move either up or down¹¹³.

Trader, who acts as momentum ignitor, places many orders very quickly (on the same side) to obtain a large position on a security by causing a sudden increase of volatility that may induce other traders into more aggressive trading and/or may trigger the execution of stop orders that emphasize the directional price movements¹¹⁴.

By establishing an early position, HFT firm attempts to profit when it subsequently liquidates the position after it has spurred the evolution of price movement on the expected side¹¹⁵.

Figure 6 shows an example of the implementation of HFT momentum ignition strategy occurred on 13th July, 2012 in Daimler share listed on XETRA (Deutsche Borse AG.).

Exhibit 18: Daimler AG, 13th July, 2012

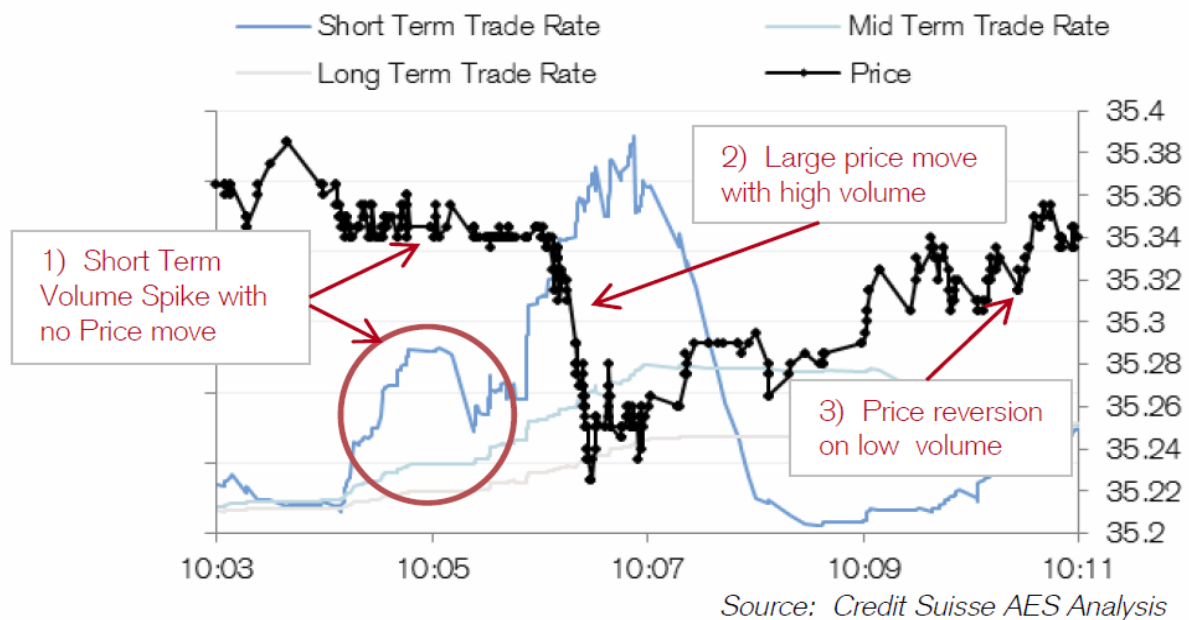


Figure 6: Momentum Ignition Example. Source: Credit Suisse AES Analysis.

A momentum ignition strategy is composed by three phases:

¹¹³ See note 71, 56-57

¹¹⁴ *Id*

¹¹⁵ *Id*

- 1) a sudden spike in volume with no price move;
- 2) a large price move with high volume;
- 3) price reversion and low volume.

In the first phase, momentum ignitor takes a large pre-position with no price change; this ignites other participants to trade aggressively in response, causing a price move and high volume; therefore, he trades out by obtaining a profit and finally, price returns to the pre-ignition value. The phase with high volume and large price move is mainly due to the trigger of stop orders (in the example stop loss order) that once triggered, they release all their intrinsic liquidity that accelerate the price trend; in fact, momentum ignitor, if detects the presence of such orders, can “gunning the market” in order to activate their executions, in particular by igniting other traders to trade more aggressively.

Summarizing, momentum ignitor induces other traders to pursue a precise price direction in order to close its pre-position at a better price.

Momentum ignition strategy is defined as a market manipulation strategy because causes prices to move towards a specific direction inducted by the ignitor and in particular, ignitions provoked by a very high number of orders with many subsequent cancellations, called “spoofing” strategies, are considered very harmful and illegal activities; hence, market regulators are employing a lot of anti-spoofing efforts aiming at limiting induced volatility but it is very difficult to prove that a trader is actually gunning the market¹¹⁶.

3.3.4. Structural Strategy

Structural strategies are activities exclusively carried out by HFT firms, which attempt to exploit structural vulnerabilities in the market or in certain market participants¹¹⁷. For example, by obtaining the fastest delivery of market data through co-location arrangements and individual trading center data feeds, HFT firms may profit by identifying market participants who are offering stale executions at stale prices¹¹⁸.

Strategies in this category are the more controversial and more discussed HFT trading practices because typically are conducted to take advantage of less evolved traders and generally are not beneficial for market quality.

Following, we describe the most widely used structural strategies: latency arbitrage and flash trading strategies.

¹¹⁶ Rena S. Miller, Gary Shorter (2016), *High Frequency Trading: Overview of Recent Developments*, Congressional Research Service, 6-7

¹¹⁷ See note 71, 52-53

¹¹⁸ *Id*

3.3.4.1. Latency Arbitrage

Latency arbitrage is the activity where HFTs attempt to take advantage of small, even tiny, price differences for stocks between various trading venues resulting from an infinitesimal time differences in the trading prices that they report on the same securities¹¹⁹.

Unlike statistical arbitrage, latency arbitrage is applicable solely if you have a system able to satisfy the low latency requirements.

For this reason, latency arbitrage is a natural application for HFTs because requires an ultra-advanced technology to exploit it; in fact, only a high frequency system, thanks to its speed advantage, can recognize an arbitrage opportunity in the same moment when occurs and, once identified, it succeeds to fully take advantage of the spread arbitrage before than other traders can even recognize it.

The identification of latency arbitrage opportunities is purely based on very fast access to market allowed by direct data feeds and co-located infrastructures that minimize as much as possible the reaction times. Spread Networks is a typical example of infrastructure installed for low-latency arbitrage purposes, which permits HFTs to discover discrepancies in the prices between New York and Chicago before all other market participants.

More properly, latency arbitrage exploits disparities in the price at which equivalent securities can be traded in different markets; such disparities can arise in different ways, most directly by the fragmentation of securities markets across multiple exchanges. US securities regulators have attempted to mitigate such fragmentation through the formulation of Regulation NMS, which mandates cross-market communication and the routing of orders for best execution at NBBO; order streams indicating the best buy and best sell available for all trading venues are constantly publicly reported by an entity called the Security Information Process (SIP)¹²⁰.

However, the speed advantage created by co-location, sophisticated technology and direct access to raw data feeds from trading platforms enable HFTs to construct their own trading order books before they are publicly available from the Security Information Processor (SIP) quote¹²¹.

Figure 7 shows the latency differential between a HFT and the SIP system: given order information from exchanges, the SIP takes some finite time, say δ milliseconds, to compute and disseminate the NBBO; while a more computationally advanced trader, as a HFT, can process

¹¹⁹ See note 116, 4

¹²⁰ Elaine Wah and Michael P. Wellman (2013), *Latency Arbitrage, Market Fragmentation, and Efficiency: A Two-Market Model*, University of Michigan, 2

¹²¹ Sal Arnuk and Joseph Saluzzi (2009), *Latency Arbitrage: The Real Power Behind Predatory High Frequency Trading*, Themis Trading LLC White Paper

the order stream in less than δ milliseconds and then, out-compute the SIP to derive the NBBO*, a projection of the future NBBO that will be seen by the public¹²².

By anticipating future NBBO, a HFT system can apply cross-market arbitrage by jumping ahead of incoming orders to pocket a tiny but sure profit.

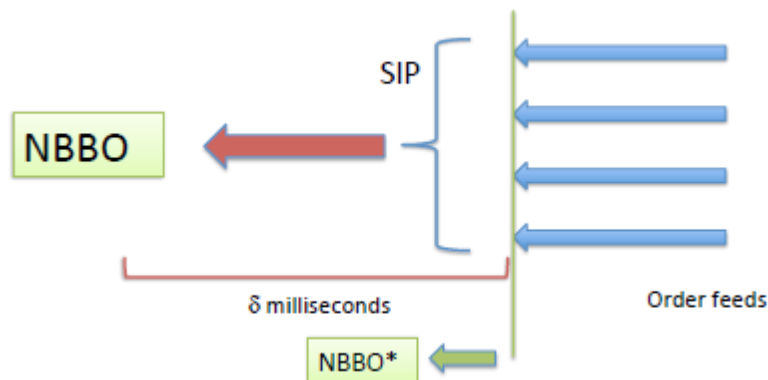


Figure 7: Latency differential. Source: Wah, 2013.

According to Arnuk and Saluzzi (2009), orders placed by institutional algorithms driven by *volume weighted average price (VWAP)* formulas are more subject to latency arbitrage because they adjust automatically if the spread shifts upwards or downwards.

Here's an example of how an HFT trading computer takes advantage of a typical institutional algo VWAP order to buy ABC stock¹²³:

1. The market for ABC is currently \$25.53 bid / offered at \$25.54.
2. Due to Latency Arbitrage, an HFT computer knows that there is a buy order that in a moment will move the NBBO quote higher, to \$25.54 bid / offered at \$25.56.
3. The HFT speeds ahead, scraping dark and visible pools, buying all available ABC shares at \$25.54 and cheaper.
4. The institutional algo gets nothing done at \$25.54 (as there is no stock available at this price) and the market moves up to \$25.54 bid / offered at \$25.56 (as anticipated by the HFT).
5. The HFT turns around and offers ABC at \$25.55 or \$25.56.
6. Because it is following a volume driven formula, the VWAP algo is forced to buy available shares from the HFT at \$25.55 or \$25.56.
7. The HFT makes \$0.01-\$0.02 per share at the expense of the institution.

Therefore, HFT latency arbitrageurs, through their speed advantage, can take advantage of incoming orders that will move the NBBO and by front-running them, they will profit by being

¹²² See note 113

¹²³ See note 114

counterpart; this trading is called predatory because impairs the prices at which other traders (e.g. buy side execution algorithms) are able to trade.

3.3.4.2. Flash Trading

Flash Trading is a very controversial practice that enables HFTs to have a right of first refusal, for specific types of orders, granted by some trading venues.

Flash Trading comes from a loophole of the obligation of “*Trade-Trough Rule*” of NMS, stating that all orders must be executed at the best execution and whether not present in a specific trading venue, such orders must be routed to the trading platform with the current NBBO.

Specifically, once an incoming market order cannot be executed against available liquidity at the marketplace where it is issued, it is flashed to its market participants before directly being routed away¹²⁴. For the duration of the flash (usually few milliseconds) the order is displayed within the marketplace at the current NBBO and if a market participant executes against it, the order is not routed; since such orders are flashed for few milliseconds, HFTs are the only market participants able to accept them.

Figure 8 shows a simple example of the functioning of a flash trading system: a client B sends to the market B a buy market order but it cannot be immediately executed because the marketplace B does not hold the NBBO, and before being routed it towards market A which instead holds the NBBO, it is flashed and executed at the NBBO with the flash trader.



Figure 8: Flash Trading example. Source: Gomber, 2011.

The consequences of the just reported flash trading transaction are positive for Client B that trades its market buy order at the NBBO, for Market B that closes out the trade without routing to Market A, and even for HFT firm that, in the case of acceptance of the flash order, implies

¹²⁴ Austin J. Sandler (2011), *The Invisible Power of Machines: Revisiting the Proposed Flash Order Ban in the Wake of The Flash Crash*, Duke Law & Technology Review No. 003

algorithms have identified a possible trading opportunities to exploit; however, on the other side, Client A sees its limit order that is not filled and Market A does not close a possible trade in its marketplace; therefore, flash trading limits the inter-market connections.

Subsequently, flash trader, after trading against flash order, can take advantage of possible price arbitrage opportunities or other financial benefits, including fees and rebates for using different exchanges or ATSs (since flash trading can be applied with several trading venues).

In this way, HFT firms by exploiting an information advantage, even for a fraction of a second, can obtain a substantial risk-free profit¹²⁵.

Hence, flash trading may be considered as an extreme case of latency arbitrage and furthermore, represents a constant source of other traders' order information¹²⁶.

3.4. Considerations on High Frequency Trading Strategies

Summarizing, HFT represents the usage of very sophisticated technologies, i.e. high frequency systems and algorithms, for applying a wide range of trading strategies and thus, it is not recognized in a single strategy. Traders leveraging HFT can employ the traditional trading strategies (such as market making strategy), likewise proprietary strategies allowed thanks the low-latency requirements (such as low-latency arbitrage strategy).

Since HFT trading strategies are very heterogenous, the assessment of HFT phenomenon should be analyzed considering all possible strategies carried out by HFT firms.

In fact, some trading strategies could likely be beneficial for the whole market system, while others may be considered structurally harmful and aggressive. This distinction is fundamentally important for regulators that should identify HFT implications and subsequently decide on approval, limitation or prohibition of a type of strategy.

US and European regulators must take into consideration a lot of components for a fair judgement of HFT strategies because such decisions contribute to determine the market developments under many aspects such as market liquidity and efficiency of the price formation process.

Therefore, it is important that any strategies that have a negative impact on market integrity and that enable for market abuse are thoroughly investigated; this is especially important if HFT as a technology eases the implementation of these strategies, makes them more profitable or creates an uneven and unfair playing field among market participants¹²⁷.

¹²⁵ See note 72, 18

¹²⁶ *Id*

¹²⁷ See note 75, 30-31

4. HIGH FREQUENCY TRADING AND MARKET QUALITY

4.1. Introduction to Literature Review

High frequency trading is attracting more and more attention not only because it is an important participant on the capital markets considering that almost all trading venues register as the majoritarian activity performed on their marketplace, with trades carried out by HFTs reaching the 50% of the total, but academic and expertise community is further interested in determining the implications and consequences of the usage of HFT techniques on trading market structure in particular after the “Flash Crash” of May 6, 2010 which has put the spotlight on some aggressive HFT practices.

The attention on HFT practice is also manifested by the fact that it is relatively recent and there are no clear ideas on how HFT firms exactly operate, what strategies may employ in different scenarios and what impact they have on the overall market quality. The opaqueness of the knowledge exists because they use proprietary algorithms together with the high speed of execution for their trading and specifically, HFTs are reluctant to disclosure relevant information on how they carried out their trading activities.

For these reasons, the most prominent questions regarding HFT can be summed up to: “Is high frequency trading beneficial or harmful to the financial system as a whole?”. Researchers have tried to answer to this question under different economic aspects, highlighting perspectives mainly focus on market quality parameters, such as bid-ask spread, transaction costs, liquidity, volatility, or price discovery; other researchers concentrate on profitability of HFT firms and their relationship with traditional traders; still others take attention on HFT behavior during a severe market disruption such as the Flash Crash¹²⁸. These market characteristics significantly affect other market participants when making trading decisions¹²⁹.

Broadly, there is not a consolidated and homogenous judgment of HFT because it is a trading technique that may be employed in different ways with very heterogeneous trading strategies. In addition, the fact that HFTs trade in a continuously manner in a very high range of securities, such as equities, bonds, currencies, and derivative products, and in markets dislocated worldwide implies that the evaluation of the phenomenon is limited to the analyzed situation. The empirical evidence on the impact of HFT on markets is rather scarce and incomplete due both to a limited availability of appropriate datasets and to the empirical and theoretical

¹²⁸ US Securities and Exchange Commission (2014), *Equity Market Structure Literature Review. Part II: High Frequency Trading*, 23

¹²⁹ Jonathan A. Brogaard (2010), *High Frequency Trading and its Impact on Market Quality*, Northwestern University

difficulties raised by the exercise; an analysis of the effect of HFT activity on markets faces two fundamental problems: 1) disentangling the impact of HFT from other factors is very complicated, with related endogeneity issue and 2) HFT is employed in many different strategies, each potentially having different impacts on the markets¹³⁰.

Especially concerning HFT strategies which are assessed as predatory trading practices inasmuch structurally prone to negatively affect the markets, the literature is extremely poor.

Nevertheless, even activities that are generally considered beneficial for market quality (e.g. market-making improves liquidity and statistical arbitrage improves price efficiency) are questioned by opponents of HFT because of the way in which they are performed¹³¹.

In fact, some studies report that HFT practice yields global positive effects on the market quality because it brings benefits in terms of liquidity, informational efficiency, and low volatility; others, instead, highlight the possibility of a deterioration of quality values specially, induced by the creation and/or exacerbation of highly turbulent trading session periods with related high probability of systemic risks¹³². HFT may pose, indeed, significant risks for market stability and integrity because the strategies adopted by algorithmic traders are more interrelated than those used by traditional ones¹³³. Moreover, whether market conditions result unstable, HFT attitude can further exacerbate sudden movements in prices provoking disorder in transactions. Discordant evaluations on HFT phenomenon are also due to different models used for estimating HFT activity because HFT firms are not uniformly identified considering that there are several approaches, direct or indirect, for the HFT identification or there are theoretical models that consider HFT firms as informed traders and others as uninformed ones; hence, also the consequences of HFT activity could result contradictory.

For these reasons, reporting several studies performed by diverse authors that consider different market quality aspects results necessary in order to have a larger evaluation of the phenomenon and not restricted to a specific assessment or context.

4.2. Market Quality

Market quality is a broad concept that generally is referred to the ability of a market of performing financial trades in a fair manner and efficiently, by ensuring to traders the best

¹³⁰ Regulatory Issues Raised by the Impact of Technological Changes on Market Integrity and Efficiency, IOSCO (2011)

¹³¹ Sal Arnuk & Joseph Saluzzi (2008), *Toxic Equity Trading Order Flow on Wall Street: The Real Force Behind the Explosion in Volume and Volatility*, Themis Trading LLC

¹³² Valeria Caivano, Salvatore Ciccarelli and Giovanna Di Stefano (2012), *Il Trading ad Alta Frequenza: Caratteristiche, Effetti, Questioni di Policy*, CONSOB Discussion Papers No. 5, 17

¹³³ *Id*

possible execution. Understanding the effects of HFTs on market quality is important, as the latter will ultimately determine long-term investors' welfare and the cost of capital for firms¹³⁴. In this section, for market quality, we take into consideration dynamics such as, liquidity, price discovery and volatility because are the market parameters more considered by the academic literature and empirical studies.

The role of literature is to evaluate the effects of HFT techniques and the behavior of HFT firms on market quality in order to try to discern the beneficial and/or harmful aspects.

Normally, improvements/deteriorations in market quality are recognized by:

- Increasing/decreasing of liquidity parameters;
- Improvement/worsening of price discovery process;
- Lowering/rising of volatility.

4.2.1. Liquidity

Liquidity refers to the ability for market participants to trade large size orders quickly, at a low cost, when they want¹³⁵.

Generally, a liquid market is a market with many bid and ask offers, low spreads and low volatility.

Everyone likes liquidity: traders like liquidity because it allows them to implement their trading strategies cheaply; exchanges like liquidity because attracts traders to their markets; regulators like liquidity because liquid markets are often less volatile than illiquid ones¹³⁶.

Liquidity contains several dimensions inside including: immediacy, width (or market breadth), depth and resiliency. High order turnover velocity and volumes, narrow bid-ask spreads, low transaction fees and resilient prices are all measures which have traditionally been positively associated with liquid markets¹³⁷.

Figure 9 shows the dimensions of market liquidity and their properties in terms of price and quantities for submitted orders. Hence, the market breadth determines the cost of doing a trade including bid-ask difference and trading commissions, the depth specifies the order size at a determined price, while the resiliency illustrates the market reversibility in terms of price and quantity¹³⁸.

¹³⁴ Bruno Biais & Thierry Foucault (2014), *HFT and Market Quality*, Bankers, Markets & Investors n° 128

¹³⁵ Larry Harris (2003), *Trading and Exchanges*", Oxford University Press

¹³⁶ *Id*

¹³⁷ *Id*

¹³⁸ *Id*

Typically, there exists a trade-off among the various liquidity dimensions; for instance, whether a trader searches immediacy by sending a market order instruction, probably will incur in worst prices for a specific quantity, with respect to spend more time in trade execution with a limit order instruction¹³⁹.

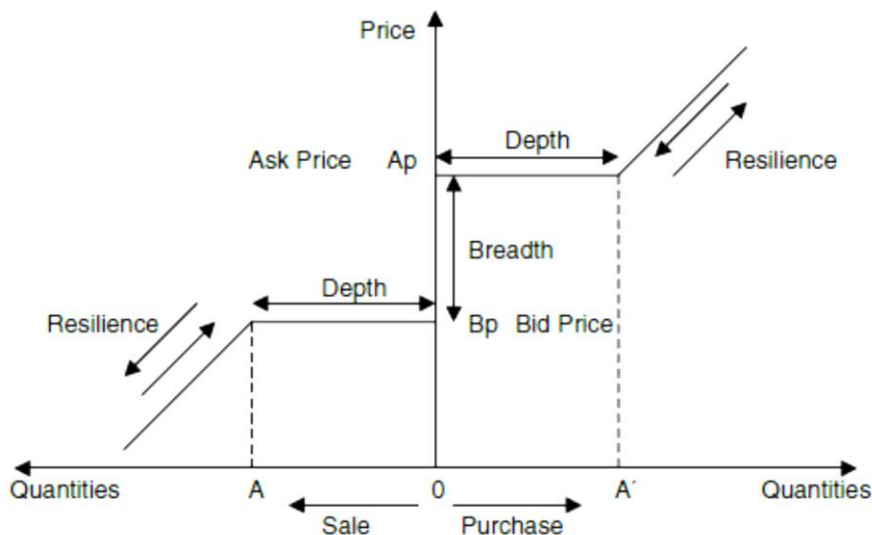


Figure 9: Aspects of market liquidity. Source: Bervas, 2006.

Nevertheless, liquidity constitutes a reliable indicator of market health status but all the dimensions must be overall considered because a sole dimension cannot evaluate the whole liquidity parameter; for example, the only value of trading volumes, referred to depth, is not a reliable indicator of liquidity because excessive volumes on the one side of limit order book may cause a widening of bid-ask spread; and on the other side, the only bid-ask spread is not a fair evaluation of liquidity because if the instruments traded are only one, the market cannot be certainly defined as liquid.

Anyway, the most considered dimension of liquidity, generally, is the bid-ask spread that represents the difference between the prices quoted for an immediate sale (bid) and an immediate purchase (ask) for one security.

Implications of high frequency trading on market liquidity has become subject to a high level of discussions for academics, experts, and regulators.

Firstly, there is to bear in mind that HFT firms require trading in venues that are already enough liquid to be able to quickly enter and exit from trading positions in order to limit their exposure to market risk; but, will markets remain liquid with the entrance of HFT?¹⁴⁰

Literature tried to provide a response to this question and in general, failed to find a significant negative correlation between HFT and market liquidity, at least in a global context.

¹³⁹ *Id*

¹⁴⁰ See note 130

Normally, HFT is commonly believed to supply significant liquidity in the marketplace through an increase of trading volume, and globally by a reduction of implicit and explicit transaction costs for investors; however, these effects may vary significantly depending on whether HFT firms engage primarily in passive activity (market-making contributes to liquidity provision) or aggressive activity (opportunistic strategies absorb liquidity) and, also considering that their behavior can suddenly change because of running market conditions.

In addition, there exists a wide opinion that HFT has globally contributed to add liquidity in a high range of instruments, through different markets and at higher speed.

However, some market participants question whether HFT firms provide liquidity to the market on a consistent basis, i.e. whether they continue to do so during turbulent conditions or whether they withdraw from the market, since HFTs not always rely on contractual market-making obligations, and further, whether certain predatory strategies have a corruptive impact on market liquidity that could lead to an unexpected systemic fragility, such as flash crashes¹⁴¹.

4.2.2. Price Discovery

Harris (2003)¹⁴² defines the price discovery process as the process of determining the price of a stock in the marketplace through the interactions of buyers and sellers which respectively seek the lowest available price and the highest available price; in this mechanism, the best buyers meet the best sellers and the market discovers the price of the negotiated security.

An efficient market should fully reflect all available information into asset prices in order to convey the fair value into their traded securities.

The incoming news and market information then, constantly, condition the dynamics of price formation, consequently, more informed traders or faster as the case of HFTs, can take advantage of such information and assist in improving price discovery by imposing adverse selection costs on uninformed or slow traders¹⁴³.

Historically, financial markets relied on intermediaries, such as traditional market makers, the privilege access to the market in order to facilitate the incorporation of information on prices by providing immediacy to outside investors¹⁴⁴; however, the rise of automation has moved this role on the HFT firms' hands that, unlike traditional human market makers, are not granted

¹⁴¹ *Id*

¹⁴² *See* note 135

¹⁴³ Jonathan Brogaard, Terrence Hendershott and Ryan Riordan (2013), *High Frequency Trading and Price Discovery*, ECB Working Paper Series No 1602

¹⁴⁴ *Id*

privileged access to the market and are not obliged to follow the traditional obligations of market makers¹⁴⁵.

Furthermore, when investors traded stocks on the basis of information about firm fundamentals, in equilibrium, stock prices converged to their fundamental values; anyway, when most trades are based on statistical and often short-lived correlations in stock returns, and investors do not hold stocks for the investment purpose (HFTs typically do not carry any position overnight), the presence of efficient pricing becomes more questionable¹⁴⁶.

Literature defines the contribution of HFT on price discovery and price efficiency as ambiguous.

On one side, an important role that is attributed to HFT firms is that they have contributed to price across different trading venues, a function that is particularly important in a fragmented market environment, through statistical arbitrage activity that facilitates the fair price circulation, if well estimated.

On the other side, without the intervention of HFTs, probably the arbitrage opportunity would be equally exploited, although slightly less rapidly; it is not clear that a decline from, say, 30 seconds to 5 milliseconds in price discovery process is extremely valuable for society¹⁴⁷; while, the technological advantage of HFTs seems to discourage well-informed slow traders to operate on transparent trading venues preferring dark pools¹⁴⁸.

Besides, in situations where HFT firms employ directional strategies, for example, acting as momentum trader on news or even worse as momentum ignitor, their behavior might exacerbate a short-term price movement even though fair value of the security was not been affected by new information, resulting in a worsening of price formation.

4.2.3. Volatility

Volatility is the tendency of prices to unexpectedly change in short time intervals¹⁴⁹. Mostly, prices change in response to new information about security values, because of aggressive behavior of market participants and due to liquidity imbalances between supply and demand sides.

High volatility values are typically connected to instability, uncertainty, and turmoil of the markets and imply the possibility to trade with impaired prices.

¹⁴⁵ *Id*

¹⁴⁶ SEC Concept Release on Equity Market Structure (2010)

¹⁴⁷ See note 134, 12

¹⁴⁸ Valeria Caivano, Salvatore Ciccarelli and Giovanna Di Stefano (2012), *Il Trading ad Alta Frequenza: Caratteristiche, Effetti, Questioni di Policy*, CONSOB Discussion Papers No. 5, 20

¹⁴⁹ See note 135

Speculative traders may benefit from price volatility because represents opportunities to conduct trading arbitrages and other aggressive strategies when prices go to desired direction¹⁵⁰. Regulators are extremely concerned about market volatility and particularly, in cases of extreme volatility where prices suddenly rise and fall dramatically from fundamental values in a matter of seconds.

One of the more used topics by opponents of HFT concerns the assertion that this type of trading entails a negative effect for the volatility of the markets because who engages in HFT exacerbates the price fluctuations in uncontrolled way and helps to provoke the creation of flash crash deriving by volatility burst.

Specifically, the causal link between HFT and volatility seems to act in two ways: on one hand HFT may be more profitable in context of high volatility and, on the other hand, HFTs massive participation may affect volatility and enhance large price variations¹⁵¹.

The academic literature has put many forces in researching a correlation between market volatility and HFT techniques under different points of view; however, finding a correlation is not a simple issue because of reverse causality and HFTs probably act differently in normal times with respect to periods of turmoil.

It follows that results provided by several studies are very discordant; contradictory findings are not surprising because HFT, by nature, is a trading technique that is prone to adapt and react to different situations in different ways and specifically, HFTs may perform, in some circumstances, activities that mitigate volatility, and in others, predatory strategies that may exacerbate price changes.

4.3. Theoretical models on market quality

Academic literature developed several theoretical models aiming at deriving some conclusions about HFT activity and behavior of HFTs in terms of market quality aspects, bearing in mind that usually different model assumptions imply different results, too.

The first theoretical model that faces the impact of HFT on market quality can be found in Cvitanic and Kirilenko (2010)¹⁵² where, with their “*benchmark model*”, simulate an electronic trading market populated by slow traders (humans) with the subsequent entrance of a HFT (machine). The HFT machine is considered as an uninformed trader, following the classical notion that a market maker does not possess any superior information and its distinction, with respect to traditional traders, is only constituted by the speed of execution and cancellation of

¹⁵⁰ *Id*

¹⁵¹ See note 147, 20

¹⁵² Jaksza Cvitanic and Andrei Kirilenko (2010), *High Frequency Traders and Asset Prices*, SSRN

orders. The authors detect that the presence of HFT implies that prices are more concentrated around the mean (i.e. have lower volatility) and they find an improvement of forecastability of transaction prices. Furthermore, they register a rise of trading volume and a reduction in intertrade duration, i.e. the time span between two trades, in direct proportion to the share of humans that change the speed of their orders in the presence of the machine.

Hoffman (2011)¹⁵³ extends a theoretical trading model where investors, mainly due to growing presence of automation, differ greatly in speed with which they can react to the announcements of investment related news and adjust their positions accordingly. Such an advantage allows HFTs to access the best quotes available for the trade, while institutional investors cannot do so. Consequently, this leads to a lower expected profit for institutional investors and their overall expected utility is lower than if they were competing with investors of identical opportunities. Concerning market quality, Hoffman argues that even though the use of HFT algorithms could have some positive externalities, such as improved market liquidity, this comes at the cost to be paid by institutional investors.

Foucault et al. (2011)¹⁵⁴ construct a theoretical model showing that HFTs actually create informational asymmetries and increase adverse selection costs. The authors argue that by being computationally more efficient than other investors, HFTs exploit the advantage that enables them to constantly outperform slower market participants. As a result, HFTs secure higher gains to themselves and at the same time increase adverse selection costs. To avoid informational asymmetries, other investors have to either leave the market or invest in costly technologies similar to ones used by HFTs. It is also argued that in a rare and non-anticipated market event such as the Flash Crash of May 6, 2010, programmed algorithms are slower in adjusting than human investors. Therefore, algorithms can potentially trigger further excessive price changes and harm the market.

Cartea & Penalva (2011)¹⁵⁵ implement a theoretical model assuming three types of traders: liquidity traders (LTs), professional traders (PTs), and high frequency traders (HFTs). Their findings support that HFT increases liquidity in markets and makes trades cheaper for other traders. On the other hand, activity of HFTs seems to increase volatility of stock prices and trading volume; however, this is not due to the higher amount of stock traded, but because of strategies adopted by HFTs.

¹⁵³ Peter Hoffmann (2011), *A Dynamic Limit Order Market with Fast and Slow Traders*, SSRN

¹⁵⁴ Thierry Foucault, Bruno Biais & Sophie Moinas (2011), *Equilibrium High Frequency Trading*, SSRN

¹⁵⁵ Alvaro Cartea & Jose Penalva (2011), *Where is the Value in High Frequency Trading?*, SSRN

4.4. Empirical studies on market quality

The recent evolution of the usage of HFT in capital markets worldwide has induced the proliferation of a lot of empirical studies debating the phenomenon; however, empirical analyses present some issues and considerations that must be taken into account.

Firstly, the potential ideal dataset for an exhaustive empirical analysis should contain information both from trade and order books at the millisecond level and additionally, an identifier of each trader, specifically pointing out the ones trading on high frequency basis. Such data would allow the researcher to observe each HFT's strategy across stocks and over time, and then evaluate the impact of HFT on a market as whole¹⁵⁶.

Nevertheless, such massive amount of information is not always available and anyway, would require a huge computational analysis effort; therefore, empirical studies may be subject to assumptions and HFT proxies in order to perform the research.

Consequently, the lack of an exact measure of HFT activity is one of the causes for the high heterogeneity in the results reached by the empirical economic literature.

Finally, the empirical studies typically examine only a relatively small amount of HFT activity inasmuch the current literature does not reveal a great deal about the extent or effect of the HFT arbitrage and structural strategies because HFT datasets generally are limited to specific markets or products, they provide little opportunity to assess HFT strategies that simultaneously seek to capture price differentials across different products and markets¹⁵⁷.

4.4.1. HFT direct approach

Empirical studies that employ the direct approach utilize proprietary datasets provided by trading venues which identify as HFTs those market participants whose primary business is the HFT carried out on proprietary basis. For instance, researches on US equity markets use data provided by Nasdaq to academics. These data report aggregated trades for 26 firms identified as HFTs by Nasdaq in 120 randomly selected stocks (divided into large, mid, and small capitalization) listed on Nasdaq and the NYSE (over the period 2008 and 2009, and one week of 2010). Nasdaq categorizes a firm as a HFT whether: it engages in proprietary trading only; it uses sophisticated trading tools and co-location services; its net position is often zero; its limit orders tend to be short-lived; and has lower trades per orders ratio. By construction, the Nasdaq

¹⁵⁶ Bruno Biais & Thierry Foucault (2014), *HFT and Market Quality*, Bankers, Markets & Investors n° 128

¹⁵⁷ US Securities and Exchange Commission (2014), *Equity Market Structure Literature Review. Part II: High Frequency Trading*,

sample excludes HFT desks from broker-dealers such as Goldman Sachs, Morgan Stanley, or Merrill Lynch¹⁵⁸.

For evidence purposes, however, HFT direct approach can present some significant shortcomings. First of all, it does not include trading activity carried out by HFT desks of investment banks that may underestimate the phenomenon; and another drawback entails that the use of aggregated data allows to categorize traders only as HFTs or non-HFTs, and for each trade, the data can identify whether the aggressive (liquidity taking) or the passive (liquidity providing) side of the trade was HFT or non-HFT, but cannot distinguish HFTs in different groups according to, for instance, the size of their inventories, or the frequency with which they post competitive quotes etc.; as a result, the inferences about the effects of HFTs' trades or orders on market quality will depend on the predominant strategy in the sample¹⁵⁹.

One of the first empirical analyses, and more exhaustive, is provided by Brogaard (2010)¹⁶⁰ using the Nasdaq sample, where investigates the relationships between HFT activity and the impact it has on the main market quality characteristics such as liquidity, price efficiency and volatility, as well as, assesses other important implications of HFT behavior on market microstructure.

Findings suggest that the overall HFT's impact on the market, in general, has to be considered positive. Specifically, Brogaard finds that HFTs make up a large majority of all trades (77%), and they supply as much liquidity as they demand. They are an integral part of the price discovery process and price efficiency than do non-HFT activity. They tend to make more money in volatile days but they dampen intraday volatility. Furthermore, the author states that: HFTs' strategies are more correlated with each other than are non-HFTs', they tend to engage in a price-reversal (contrarian) strategy, i.e. they buy (sell) stocks whose prices have been declining (increasing) in the last 10 to 100 seconds, they result very profitable, and there is no evidence that HFTs withdraw from markets in bad times or that they engage in abnormal front-running of large non-HFT trades (no quote matching strategies).

With the same dataset, Brogaard, Hendershott and Riordan (2013)¹⁶¹ focus on the implications in terms of pricing process. By applying a state space model, they decompose price movements into permanent and transitory components and relate changes in both to HFTs. The permanent component is interpreted as information, whereas the transitory component as pricing errors or

¹⁵⁸ *Id*

¹⁵⁹ *Id*

¹⁶⁰ Jonathan A. Brogaard (2010), *High Frequency Trading and its Impact on Market Quality*, Northwestern University

¹⁶¹ Jonathan Brogaard, Terrence Hendershott and Ryan Riordan (2013), *High-Frequency Trading and Price Discovery*, ECB Working Paper Series No 1602

noise. The space model incorporates the interrelated concepts of price discovery (how information is impounded into prices) and price efficiency (the informativeness of prices). Results show that HFT aggressive trades (market orders) facilitate price efficiency by trading in the direction of permanent price changes and in the opposite direction of transitory pricing errors, both on average and on the highest volatile days. They further find that HFT aggressive trading imposes adverse selection costs on non-HFT passive traders, raising questions about whether the informational efficiency gains outweigh the direct and indirect adverse selection costs imposed on non-HFTs, given that HFTs predict price changes occurring only a few seconds in the future. Finally, authors find that passive HFT trades (limit orders) are adversely selected, as well as negatively associated with permanent price impact and positively associated with transitory price impact.

Sarah Zhang (2013)¹⁶² studies the role of HFTs and non-HFTs in processing hard and soft information with related implications in price discovery, by using the already mentioned proprietary HFT dataset provided by Nasdaq. In the model, “hard” information is quantitative and refers to index prices, while “soft” information is qualitative and refers to textual news. Evidences show that HFTs dominate non-HFTs in processing hard information shocks and lead price discovery in the short run (within a period of 10 seconds); while non-HFTs analyze better soft information, and facilitate the incorporation into prices for a longer amount of time (up to two minutes).

Other empirical studies utilize non-US markets proprietary datasets. Jarnecic & Snape (2010)¹⁶³ use trade and quote data from the London Stock Exchange on 2009 for their research. Like Nasdaq dataset, this set label all activity by participant type: high-frequency participants, traditional market makers, three types of institutional members (i.e. small, large and investment banks) and retail brokers. Authors detect that HFT activity varies widely across stocks and is more prevalent in large capitalized stocks with high on-market competition, high price volatility and strong off-exchange competition but less prevalent in stocks with high tick sizes and informed order flow. Concerning market quality results, authors indicate that HFT participants both contribute and demand liquidity in almost even proportions, and that their activity is more likely to dampen than increase volatility.

¹⁶² Sarah Zhang (2013), *Need for Speed: An Empirical Analysis of Hard and Soft Information in a High Frequency World*, SSRN

¹⁶³ Elvis Jarnecic and Mark Snape (2010), *An Analysis of Trades by High Frequency Participants on the London Stock Exchange*, Working Paper

Bershova & Rakhlin (2012)¹⁶⁴ examine data on aggregate HFT and long-term institutional investor trading volume that was routed through a single, large multiple-service broker with a significant presence in the Tokyo and London equity markets during the first two quarters of 2010. The authors classify HFT clients as those that use the broker's ultra-low latency infrastructure. The dataset did not permit volume to be classified as aggressive or passive. They show that the increase in short-term intraday volatility and in trading costs due to HFT is more than offset by the narrowing of bid-ask spreads.

Caivano (2015)¹⁶⁵ concentrates on the impact of HFT on stock price volatility over the period 2011-2013 on Italian equity market. Author identifies HFTs according to two methods: the first one, based on public information on the trading strategies of market participants (pure HFTs), the second one includes also the main investment banks since they carry out some proprietary trading with HFT. Potential endogeneity is controlled through an instrumental variable approach (the introduction of a new trading HFT-friendly platform on 2012 known as *millennium*). Empirical findings demonstrate that an exogenous increase of HFT activity has caused a statistically and economically significant increase in price volatility on Borsa Italiana.

A further part of literature has only access to proprietary datasets that identify the wider category of Algorithmic Trading; we report some empirical examples.

Hendershott & Riordan (2011)¹⁶⁶ analyze 30 DAX stocks on the Deutsche Borse in January 2008; the DB provided order data on Algorithmic Trading. Authors find that AT consumes liquidity when it is cheap and provides liquidity when it is expensive. AT contributes more to the discovery of the efficient price than human trading. Authors state that these results demonstrate that AT closely monitor the market in terms of liquidity and information and react quickly to changes in market conditions. Furthermore, they do not find evidence of AT exacerbating volatility during periods of market turbulence.

Chaboud et al. (2009)¹⁶⁷ obtain AT data from EBS on the three most-traded currency pairs: euro-dollar, dollar-yen, and euro-yen from 2006 and 2007. Using a Vector Autoregressive approach, they estimate the contributions of algorithmic trades and human trades to the variance of returns over a 30 minutes horizon. Overall, they find evidence that the presence of more

¹⁶⁴ Nataliya Bershova & Dmitry Rakhlin (2012), *High-Frequency Trading and Long-Term Investors: A View from Buy-Side*, SSRN

¹⁶⁵ Valeria Caivano (2015), *The Impact of High-Frequency Trading on Volatility: Evidence from the Italian Market*, Quaderni di Finanza n° 80, Consob

¹⁶⁶ Terrence Hendershott and Ryan Riordan (2011), *Algorithmic Trading and Information*, SSRN

¹⁶⁷ Alain Chaboud, Benjamin Chiquoine and Erik Hjalmarsson, Clara Vega (2009), *Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market*, Board of Governors of the Federal Reserve System International Finance Discussion Papers n° 980

algorithmic trading is associated with lower exchange rate volatility. On the contrary, human trades contribute more to price discovery than algorithmic trades.

4.4.2. HFT indirect approach

The studies based on the indirect identification of HFT activity estimate HFT firms according to their operational features, through the analysis of data on submitted orders and the speed at which these orders are submitted. The most of HFT proxies are based on: 1) intraday inventory management; 2) lifetime of orders; 3) message traffic; and 4) identification of strategies¹⁶⁸.

One drawback of this approach is that focusing on few features could bring to select only a subset of HFT strategies, resulting in a partial view of the phenomenon; and some proxies may also capture the activity of ATs operating at lower frequencies (like for instance brokers using algorithms to execute orders at low costs for their clients)¹⁶⁹.

Frank Zhang's (2010)¹⁷⁰ investigation represents the first study that indirectly examines the HFT role on capital markets, and in contrast to other studies like Brogaard (2010)¹⁷¹, which analyze intra-day effects of HFT on market quality, Zhang focuses on long-term effects because considered more important for investors. He uses a large sample of US firms during 1985-2009 and classifies investors into three categories: institutional investors, individual investors, and high-frequency traders. Author, specifically, determines HFT as all short-term trading activities by hedge funds and other institutional traders following three assumptions: 1) no-HFT existed before 1994; 2) HFTs do not carry any position overnight; and 3) they have extremely short holding periods. He finds that HFT is positively correlated with stock price volatility after controlling for firm fundamental volatility and other exogenous factors. In more details, HFT is found to lead stock prices overreaction to news as it hinders the incorporation of information about fundamentals into asset prices. The detrimental effects on volatility are higher for large cap stocks and during market turbulences. Moreover, consistently with the hypothesis that HFTs mainly take advantage of large trades typically carried out by institutional investors, stocks with high institutional ownership seem to suffer more the negative impact of HFT.

Kirilenko et al. (2011)¹⁷² provide an analysis of a particular market that was affected by an important disruption in terms of market quality performance, in order to figure out the possible

¹⁶⁸ Antoine Bouveret, Cyrille Guillaumie (2014), "*High-frequency trading activity in EU equity markets*", ESMA Economic Report

¹⁶⁹ See note 156, 13

¹⁷⁰ X. Frank Zhang (2010), *High Frequency Trading, Stock Volatility and Price Discovery*, SSRN

¹⁷¹ See note 160

¹⁷² Andrei Kirilenko, Mehrdad Samadi, Albert S. Kyle, Tugkan Tuzun (2011), *The Flash Crash: The Impact of High Frequency Trading on an Electronic Market*, SSRN

responsibilities associated to HFT. They use audit-trail data and examine the trades in the E-Mini S&P 500 stock index futures market around the period of the Flash Crash on May 6, 2010. Specifically, authors concentrate on the period of extreme volatility occurred in the US financial markets that day, that is, a short period of 30 minutes during which stock market indices, stock index futures, and ETFs registered a decline of unanticipated magnitude (about 9%) and, shortly after, almost totally recovered.

Through their analysis, Kirilenko et al. (2011) define HFTs as intermediaries who 1) individually participate in a very high number of transactions during the day; 2) have a low inventory at the end of the day; and 3) experience relatively low variations in their inventory positions. Findings suggest that HFTs did not trigger the Flash Crash, but their responses led to the creation of a “*hot-potato*” volume effect, as the same positions were passed rapidly back and forth among same HFTs, that exacerbated market volatility, removed the market liquidity, and accelerated the downside pressure.

Hasbrouck and Saar (2010)¹⁷³ investigate the low-latency trading activity and its impact on market quality measures. The dataset used contains order-level Nasdaq data during June 2007, a ‘nominal’ market period, and October 2008, a volatile and uncertain period. The high number of submissions, cancellations and executions determines the proxy used for low-latency trading activity; in addition, authors do not consider agency algorithmic traders as low-latency traders. The empirical findings suggest that low-latency activity improves market quality in both periods, more specifically, short-term volatility is observed to be lower and market liquidity improved, with decreasing quoted and effective spreads and increasing displayed depth in the limit order book.

Hendershott et al. (2011)¹⁷⁴ examine a sample of NYSE stocks and use the rate of electronic message traffic, which includes electronic order submissions, cancellations, and trade reports, as proxy for AT. The data used is from 2001 through 2005 during which in 2003, the NYSE became a fully automated stock exchange. The results suggest that, for large stocks in particular, AT narrows spreads by reducing adverse selection and increasing the amount of information in quotes as compared to trades. These indicate that AT does causally improve liquidity and enhances the informativeness of quotes and prices.

¹⁷³ Joel Hasbrouck & Gideon Saar (2010), *Low-Latency Trading*, SSRN

¹⁷⁴ Terrence Hendershott, Charles M. Jones and Albert J. Menkveld (2011), *Does Algorithmic Trading Improve Liquidity?*, The Journal Of Finance Vol. Lxvi, No. 1 February 2011

Hagströmer and Nordén (2013)¹⁷⁵ use member-level data from NASDAQ-OMX Stockholm for 30 stocks listed on this market in August 2011 (high volatility period) and February 2012 (low volatility period) and categorize HFTs as traders engaging in proprietary trading only and using algorithms in their trading strategies. More interestingly, they decompose HFTs in two subgroups according to strategies: market-makers, who provide liquidity to other participants, and opportunistic traders, who use more directional or arbitrage strategies. They find that market makers constitute the lion's share of HFT trading volume (63- 72%) and limit order traffic (81- 86%) and that have higher order-to-trade ratios and lower latency than opportunistic HFTs. Using an event study based on changes in minimum tick size, authors find that market-making activities are good for the overall market quality in the sense that they reduce short-term volatility and provide liquidity consistently with their limit orders. Conversely, opportunistic HFTs supply significantly less liquidity, especially for stocks with low market capitalization, low trading volume, high volatility, and large spreads, but both statistical arbitrage and momentum strategies mitigate intraday price volatility.

4.5. Summary of Academic Literature Review

Academic literature contributed to develop many researches aiming at assessing the implications of HFT phenomenon in terms of market quality aspects, as liquidity, volatility, and price discovery. These studies spread either in form of theoretical models and empirical analyses.

The theoretical contributions reach often contrasting results strictly depending on the basic assumptions used to develop the model, such as the degree of information attributed to HFT firms or the type of pre-determined intrinsic HFT strategies.

Much more useful appear the empirical studies, that mostly, are designed as event studies that analyze structural breaks introduced either by a change to regulation or by a development of a HFT friendly technology, or focus on particular phenomena of market turbulence¹⁷⁶.

However, even empirical findings turn out very discordant mainly due to different identifications of HFT techniques and participants and different markets and periods taken into account in the analysis.

We register a lack of literature for diverse HFT strategies, especially for structural and inter-market ones, due to the difficulty to obtain relevant data and because of the tough issue of analyzing and processing high-frequency information.

¹⁷⁵ Bjorn Hagstromer and Lars Norden (2013), *The Diversity of High Frequency Traders*, Journal of Financial Markets, 16(4), 741-770.

¹⁷⁶ See note 132, 16

For this reason, some empirical studies could only discern HFT activity among passive and aggressive, by registering that the former, overall, leads to more beneficial effects to liquidity parameters, while the latter is more likely to improve pricing discovery and efficiency.

Summarizing, the majority of the reported studies carries a positive contribution of HFT on market quality but there are some important exceptions that constitute a strong opposition to HFT.

More in detail, liquidity parameters, such as bid-ask spread, trading volume, and immediacy prove to be improved, except in situations of markets in turmoil where an excessive trading risk is likely to trigger HFT algorithms that automatically decide to stop providing liquidity or even take the remaining one.

Pricing discovery, involving the incorporation of information into prices, is registered to be favored with the introduction of HFT technologies in the markets, although this process might result too fast and aggressive by leading to imposition of heavy adverse selection costs to slower traders that may decide to abandon the market.

More conflicting are the volatility implications; empirical findings agree to maintain that intraday volatility is a fundamental requirement for HFT to trade because they can identify more profit opportunities when prices move more rapidly. The reverse correlation is less distinct inasmuch HFTs may both mitigate the pricing movements or exacerbate it, mainly depending on the predominant adopted strategy and on the current conditions of the markets because, for instance, in a flash crash context, the behavior of HFTs has been demonstrated to generate a “*hot-potato*” effect that led a volatility burst. Conversely, in other quieter situations, HFT empirically alleviated the volatility of prices.

In conclusion, despite empirical evidence generally suggests that HFT tends to improve market quality, there have been many observations that prove some potential detrimental impacts of HFT on market quality, specifically, a number of aberrant stock market behaviors that occurred in the HFT era, such as extreme intraday price volatility, that would seem associated to the proliferation of such trading technique.

Consequently, the next step is to debate about the potential distortive HFT effects on markets, specifically concerning market integrity, and possible systemic risks caused by HFT, by reporting the adopted or potential countermeasures to contrast them, in order to completely evaluate the phenomenon.

5. RISKS OF HIGH FREQUENCY TRADING

5.1. Introduction to Risks of High Frequency Trading

An accurate examination of the role of HFT phenomenon in modern capital markets requires also a deep analysis about potential controversies, risks and new challenges posed by this new form of trading. Such discussion has raised increasing attention for market participants, market operators and especially for supervisory authorities that must strive to decide on potential new provisions which might become necessary if HFT demonstrates to represent a relevant risk for fairness, and market integrity.

Critics and opponents relieved some concerns whether HFT plays a role in exacerbating market fragility; whether it may heighten the market's systemic risk; whether it enhances or harms the quality of those markets; whether certain kinds of HFT strategies may constitute an illegal form of trading; whether HFT helps foster a system of two-tiered trading markets that benefits certain traders at the expense of others; and whether the presence of HFT has been to the detriment of non-HFT investors and investor confidence in the securities markets.¹⁷⁷

5.2. Two-Tiered Markets

A challenge posed by HFT is the need to figure out whether HFT firms' sophisticated speed technologies, essentially represented by low-latency infrastructures, result in an unfair advantage over other market participants, and the extent to which it is allowed that HFTs have information that others do not have.

About that, in order to be faster than other traders, HFT firms often pay for the right to access two pieces of technology for market trading centers: 1) direct market access and 2) co-location, which permit to have, respectively, real-time market quote and trading data fractions of a second before the data reach other investors, and the possibility to minimize transmission times through the right to place their servers in the same data centers in which an exchange's market data systems are located¹⁷⁸.

This has led to charges that HFTs are unfairly advantaged vis-à-vis other traders because an informational advantage of just a fraction of a microsecond can be enough to get a better price, even for a later-placed order¹⁷⁹.

¹⁷⁷ Gary Shorter, Rena S. Miller (2014), *High-Frequency Trading: Background, Concerns, and Regulatory Developments*, Congressional Research Service

¹⁷⁸ *Id.*, 21-22

¹⁷⁹ *Id.*

For some HFT opponents, such private low-latency advantage is believed unlawful, because on one hand, the market regulators forbid firms to release fundamental information to a subset of investors (insider trading) but, on other hand, they allow market centers to sell data feeds directly to certain subscribers, thus creating a tiered system of investors¹⁸⁰.

Proponents, however, say that securities markets have always been characterized by differential on tiered access to securities trades, going back to the time when floor traders had favored access to stock orders¹⁸¹.

Market centers highlight that the benefits of direct feeds and co-locations are available to anyone willing to pay for the services.

Others, however, argue that the cost-benefit trade-off for investing in these tools and capabilities is likely to be much more favorable for organized, strongly capitalized traders, given that the low latency benefits come from very large volumes of trades per day¹⁸².

The competitive advantage becomes even more pronounced when trading venues allow HFT firms to engage in “*flash trading*” activity, inasmuch HFTs have the possibility to see resting orders inside limit order books before all other traders and hence, have a right of first refusal.

Critics of the practice contend this creates a two-tiered market in which HFT firms can unfairly exploit other traders, through a legalized front running conduct¹⁸³.

Market centers offering such service (e.g. Direct Edge), defend their activities affirming that flash technology democratizes access to the non-displayed market and in this regard, removes different “tiers” in market access; additionally, any market subscribers can be a recipient of flashed orders¹⁸⁴.

Regulators never implemented the banning of flash trading, nonetheless, starting from 2011, many exchanges and alternative trading systems have stopped voluntarily to offer this disputable practice¹⁸⁵.

5.3. Phantom Liquidity

It is globally considered and empirically tested that HFT offers substantial liquidity to the marketplace mainly through passive market making activity, which involves the submission of

¹⁸⁰ Joel Hasbrouck & Gideon Saar (2010), *Low-Latency Trading*, SSRN

¹⁸¹ See note 178

¹⁸² *Id*

¹⁸³ Ellen Brown (2010), *Computerized Front-Running*, Counter Punch

¹⁸⁴ Imad Moosa (2015), *The Regulation of High-Frequency Trading: A Pragmatic View*, Journal of Banking Regulation

¹⁸⁵ SEC (2009) available at: <https://www.sec.gov/rules/proposed/2009/34-60684.pdf>

non-marketable orders (i.e. limit resting orders) which offer the opportunity to trade at specified prices.

However, some observers wonder whether market liquidity provided by HFT firms is qualitatively comparable to the liquidity provided by traditional market makers.

A severe criticism about this issue is represented by the assertion that generally HFT liquidity lacks depth because of the relatively small size of quotes and the fact that they have no affirmative market-making obligations, i.e. are not obliged to offer liquidity in various circumstances, like specialist dealers do¹⁸⁶.

This implies that, in the financial jargon, the liquidity offered by HFT is dubbed as “*phantom liquidity*” or “*flickering quotes*”, due to its penchant for rapidly posting and then subsequently cancelling orders¹⁸⁷.

Consequently, the available liquidity for given securities may often be less than what may appear to be the case, in the sense that it could vanish very quickly from limit order trading books during periods of high market stress¹⁸⁸.

After all, even in quiet market scenarios, limit order books can quickly empty and prices can crash simply due to the speed and numbers of orders flowing into the market and due to the HFT ability to instantly cancel submitted quotes¹⁸⁹.

This has generated a negative impact of the market information quality, since the true level of market liquidity is different from its perceived level, inasmuch liquidity could disappear at any moment, forcing other traders to accept worst prices and not those they made into account¹⁹⁰.

Concerning the massive number of quote cancellations performed by HFT firms, observers allege that there may be legitimate reasons for cancelling orders to adjust them to new market conditions (such as in certain HFT arbitrage strategies where the liquidity previously shown completely disappears as soon as the relevant orders from the arbitrage trade are executed on one of the platforms), but entering orders without any actual intention to execute or have them executed is prohibited. If there is not such intention, this means HFT party in question is giving an incorrect or misleading signal for supply of, or demand for, the specific financial instrument¹⁹¹.

¹⁸⁶ Gary Shorter, Rena S. Miller (2014), *High-Frequency Trading: Background, Concerns, and Regulatory Developments*, Congressional Research Service, 18-19

¹⁸⁷ *Id*

¹⁸⁸ Valeria Caivano, Salvatore Ciccarelli and Giovanna Di Stefano (2012), *Il Trading ad Alta Frequenza: Caratteristiche, Effetti, Questioni di Policy*, CONSOB Discussion Papers No. 5, 21-22

¹⁸⁹ Summary Report of the Joint CFTC-SEC Advisory Committee on Emerging Regulatory Issues (2011), 2

¹⁹⁰ Alfonso Puorro (2013), *High Frequency Trading: una panoramica*, Questioni di Economia e Finanza n° 198, Banca d'Italia, 26

¹⁹¹ AFM (2010), *High Frequency Trading: The Application of Advanced Trading Technology in the European Marketplace*, 35-36

Figure 10 shows how is strongly increased the ratio between cancelled orders and actually executed orders in recent years on Nasdaq quotations. The trend of such ratio passed from under 10 at the beginning of 2002 to over 30 by the end of 2009¹⁹².



Figure 10: Cancellation/Execution Ratio. Source: NASDAQ ITCH data.

The boundless use of cancelled orders is reportedly associated with manipulative strategies or, aimed at providing a deceptive framework of the current market conditions in order to capture more profitable trading opportunities.

5.4. Market Manipulation

Market integrity could be undermined by acts of manipulation or attempts thereof, fraud, disruptive trading, unlawful trade practices, such as pre-arranged trading and wash trading. Such risks are not new to financial markets, but automated trading can provide traders with new tools to engage in such unlawful conducts, for many reasons, including creating false impressions of market depth, trading volume, and prices¹⁹³.

In fact, HFT firms may adopt several manipulative strategies that involve the massive and abusive use of quote cancellations, and hidden orders.

Booming manipulative practices include aggressive versions of momentum ignition strategies, where a HFT participant attempts to induce others to trade artificially high or low prices. Examples of this activity consist of “spoofing”, “layering”, and “smoking” strategies¹⁹⁴.

Spoofing consists in the trading practice where the *spoofers*, whose real intent is to sell (or buy) a stock, places a limit order to buy (or sell) that are above (below) the best bidding (asking)

¹⁹² Pavitra Kumar, Michael Goldstein, Frank Graves, & Lynda Borucki (2011), *Trading at the Speed of Light: The Impact of High-Frequency Trading on Market Performance, Regulatory Oversight, and Securities Litigation*, The Battle Group Issue 02, 3

¹⁹³ Joint Staff Report (2015), *The US Treasury Market on October 15, 2014*, 54

¹⁹⁴ Gary Shorter, Rena S. Miller (2014), *High-Frequency Trading: Background, Concerns, and Regulatory Developments*, Congressional Research Service, 23

price, which are not meant to be executed, but the only intention is to scare other traders into buying (selling) at a high (low) price, in order to profit from the bargain prices¹⁹⁵.



Figure 11: Illustration of Spoofing. Source: The Wall Street Journal.

Layering is a form of spoofing in which a trader on one side of the order book inserts multiple hidden orders, while on other side places a large quantity of displayed orders with different price limits. This is designed to create the impression of increasing pressure on other side of the order book. Consequently, HFT can exploit the higher speed to cancel the displayed orders before execution and then, execute the hidden orders at advantageous prices¹⁹⁶.

Smoking strategy involves the placement of particularly alluring quotes, attracting slow market orders, but rapidly revised on to less generous terms, even before slow counterparties in the transaction can be aware of the changing scenario¹⁹⁷.

Concerns are related, also, for the potential implementation of illegal forms of quote matching strategies, which is, an order anticipation strategy directed at extrapolating liquidity by trading ahead of large traders. An unlawful practice is the front running activity which means profiting by placing one's own orders ahead of a large order based on knowledge of that impending order. However, the line between the quote matching applied by HFTs and illegal front-running can be very nuanced; for example, it is prohibited the trading when in possession of material non-public information about a large trader inclined to trade shortly afterwards, but it is allowed the employment of sophisticated pattern recognition software to ascertain from publicly available

¹⁹⁵ *Id.*

¹⁹⁶ *Id.*

¹⁹⁷ Valeria Caivano, Salvatore Ciccarelli and Giovanna Di Stefano (2012), *Il Trading ad Alta Frequenza: Caratteristiche, Effetti, Questioni di Policy*, CONSOB Discussion Papers No. 5, 22

information the existence of a large trader, or the sophisticated use of orders to “ping” different market centers in an attempt to locate and trade in front of large buyers and sellers¹⁹⁸.

Furthermore, HFTs might perform other types of manipulative conducts, which are designed to impede other market participants to trade in a proper way; a widespread form of this conduct is called *quote stuffing*.

Quote stuffing is the activity of quickly entering and then withdrawing large quantities of unwieldy orders in the aim of overloading and generating congestion for other market participants’ systems and therefore, impairing market access for slow traders. This gives free rein to fast traders to execute profitable trades at the expense of the rest of traders¹⁹⁹.

The figure 12 shows two situations of quote stuffing which occurred respectively on 2nd May, 2012 on Heineken shares traded on Euronext (left-hand graph), and on 10th August, 2012 on Telefonica quotations traded on the Bolsa de Madrid (right-hand graph)²⁰⁰. Both examples are characterized by thousands of order submissions and subsequent cancellations on the ask market side²⁰¹.

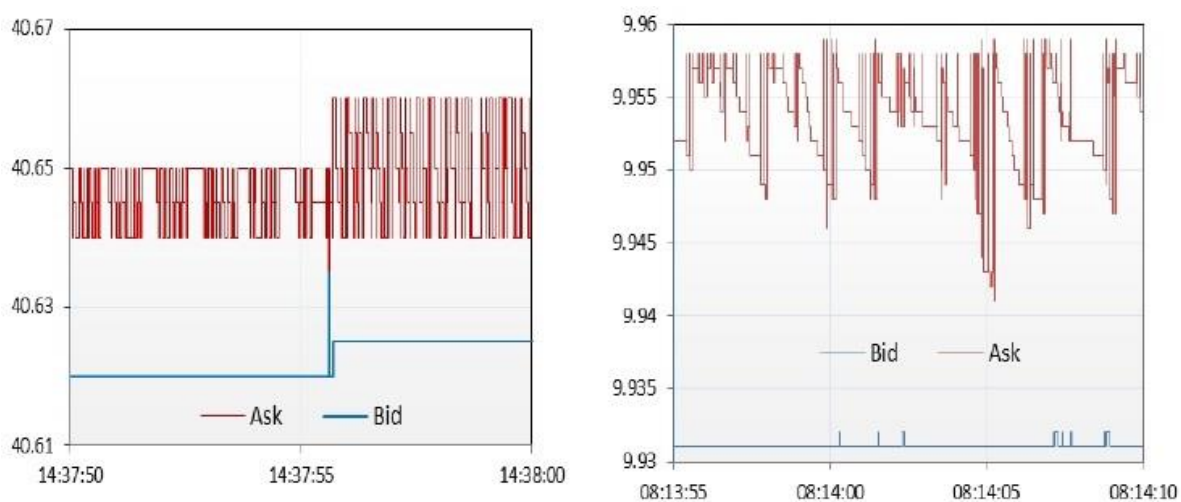


Figure 12: *Quote Stuffing situation. Source: Credit Suisse AES Analysis.*

5.5. Operational Risk

Operational risk is of especial concern with automated trading given the strong use of high execution speed and the management of huge amount of trading data. Operational risks range

¹⁹⁸ SEC Concept Release on Equity Market Structure (2010), 54-55

¹⁹⁹ Bruno Biais and Paul Woolley (2012), *The Flip Slide: High Frequency Trading*, Financial World

²⁰⁰ Alfonso Puorro (2013), *High Frequency Trading: una panoramica*, Questioni di Economia e Finanza n° 198, Banca d’Italia, 21

²⁰¹ *Id*

from malfunctioning and incorrectly deployed algorithms to algorithms reacting to inaccurate or unexpected data²⁰².

The heavy reliance on algorithms for trading decisions and execution may pose serious risk when one or more algorithms behave in an unexpected way. There is the risk that *rogue algorithms*, i.e. algorithms that malfunction and operate in an unintended way, may trigger a chain reaction and withdraw liquidity from the market or impair orderly trading²⁰³.

Such risk is magnified by the dazzling speed at which most algorithmic HFT takes place that may lead one errant or faulty algorithm not only to rack up millions in losses for the executor in a very short period, but even more concerning, may generate irregular securities price movement, through submission of a series of wrong order instructions²⁰⁴.

An example of the crushing effect that may generate a malfunctioning of HFT systems has been demonstrated on August 1, 2012, where an American global financial services firm, Knight Capital Group, deployed untested trading software which sent numerous erroneous orders in NYSE-listed securities into the market, by causing a major disruption in the prices of 148 companies listed, thus, for example, shares of Wizzard Software Corporation went from \$3.50 to \$14.76, in approximately 45 minutes of errant trade executions via a “*rogue*” algorithm, while KCP registered a loss of \$440 million²⁰⁵.

It is therefore in the first place in interests of HFTs that the odds for this type of error are kept as low as possible; however, the quality of monitoring potential wrong order instructions is undermined by the need to operate with exceptional speed; in fact, by using “*naked*” sponsored access, HFTs have available the lowest latency timing, but at the expense of efficient pre-trade checks²⁰⁶.

Therefore, to limit the exposure to erroneous trades and rogue algorithms, it is important that trading venues and intermediaries have systems and controls appropriate to a high frequency environment and that appropriate trading control mechanisms are in place to prevent excessive market movements when errors occur²⁰⁷.

5.6. Systemic Risk

²⁰² Joint Staff Report (2015), *The US Treasury Market on October 15, 2014*, 54

²⁰³ Regulatory Issues Raised by the Impact of Technological Changes on Market Integrity and Efficiency, IOSCO (2011), 29

²⁰⁴ *Id*

²⁰⁵ Gary Shorter, Rena S. Miller (2014), *High-Frequency Trading: Background, Concerns, and Regulatory Developments*, Congressional Research Service, 38-39

²⁰⁶ AFM (2010), *High Frequency Trading: The Application of Advanced Trading Technology in the European Marketplace*, 26

²⁰⁷ See note 203

The use of HFT and, more generally, of AT may be more prone to trigger or exacerbate market systemic risks given that some strategies conducted by algorithms result much more correlated than those used by non-automated traders. A potential concern here is that because of this correlation, shocks that hit a small number of very active HFTs could detrimentally affect the entire market, until to touch also other trading venues since the intense cross-market activity of such traders²⁰⁸. Another criticism is that HFT firms are often very lightly capitalized, a factor that could generate failures, and handling the corresponding counterparty risk could be challenging, because HFTs tend to turn over their positions many times a day by interacting with various traders, while securities trade clearing systems tend to operate at a much slower rate. Combined these elements could generate systemic market disruptions²⁰⁹.

Fueling this thesis, Jain et al. (2016)²¹⁰ provide evidences that the introduction of a high-speed trading platform, nicknamed Arrowhead, launched in January 2010 in the Tokyo Stock Exchange (which reduced the TSE's latency from six seconds to two milliseconds), led to a significantly increase of shock propagation risk by rising both autocorrelation and cross-correlation in the order flow. Specifically, quote-stuffing risk as measured by the quotes-to-trade ratio doubled; systemic risk of events such as flash crashes, also, increased with respect to the pre-high frequency scenario. The systemic risks associated with high frequency trading result from aggressive liquidity demanding behavior, whereas the systemic risks of high frequency quoting emanate from cancellation or absence of quotes from liquidity suppliers. In addition, authors notice, that in accordance with related empirical studies, the new platform improved liquidity parameters, such as spread, depth and the cost of immediacy.

5.7. Flash Crashes

The most frequent criticism and even fiercest about the potential risks derived by HFT activity, alleges that HFT is apt to play a leading role in situations of extreme market fluctuations, in the financial jargon dubbed as "*Flash Crashes*".

Due to these assertions, HFT gained prominence in the media after May 6th, 2010, the day of the occurrence of the Flash Crash, and is increasingly intensified with observations of ongoing mini-flash crashes that frequently have affected the securities and futures markets.

²⁰⁸ Gary Shorter, Rena S. Miller (2014), *High-Frequency Trading: Background, Concerns, and Regulatory Developments*, Congressional Research Service, 27-28

²⁰⁹ *Id*

²¹⁰ Pankaj Jain, Pawan Jain, and Thomas H. McNish (2016), *Does High Frequency Trading Increase Systemic Risk?*, SSRN

5.7.1. The Flash Crash of May 6, 2010

On May 6, 2010, the prices of many US-based equity products and major equity indices in both the futures and securities markets (particularly the Dow Jones Industrial Average, DJIA), which were already down of 4% relative to the previous day close, dropped again a further 5-6% in a couple of minutes before recovering to their previous values²¹¹.

Many of the almost 8,000 individual equity securities and ETFs traded that day suffered declines in price ranging from 5% to 15% before rebounding and recovering most, if not all, of their losses in a short period of time²¹².

Over 20,000 trades across more than 300 securities were executed at prices more than 60% away from their values a few minutes before; these trades were executed at extreme prices, some for pennies and others for 100,000\$ or even more, before bouncing back to their original levels²¹³.

Finally, the turbulent trading day ended with the futures and securities markets suffering a loss of only 3% from their prior day close²¹⁴.

The nearly 1,000-point DJIA's plunge was historical, representing the largest intra-day decline in its history. The whole event had been dubbed the "*The Flash Crash*", given the sudden and striking fall in prices²¹⁵.

Afterwards the event, many questions are arisen, mainly focused on the role of HFT and the behavior of HFT firms during this outstanding occurrence.

On September 30, 2010, the SEC and the CFTC issued a joint report on the market events of May 6, 2010, which clarified the chain of events that led up to the Flash Crash. The report described the trigger, the traders' behavior, and the market framework during the disruptive event.

The report detailed how an undisclosed large institutional trader executed a single trade, which consisted of a large sell order worth about \$4 billion through an automated execution algorithm (but not through HFT) at a time when the markets were already extremely stressed²¹⁶. The order of E-Mini S&P 500 (a stock market index futures contract traded on Globex, the CME's electronic platform) contracts initially exhausted available buyers, including HFTs, who began to aggressively sell them. The report, which largely focused on market structure and liquidity

²¹¹ Report of the Staffs of The CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues (September 30, 2010), *Findings Regarding the Market Events of May 6, 2010*, 1

²¹² *Id*

²¹³ *Id*

²¹⁴ *Id*

²¹⁵ Gary Shorter, Rena S. Miller (2014), *High-Frequency Trading: Background, Concerns, and Regulatory Developments*, Congressional Research Service, 29

²¹⁶ *Id*

concerns, did not place blame on HFT for the crash²¹⁷. Rather it raised questions about the ability of HFT to provide continuous market liquidity. The report also observed that HFTs “in the equity markets, who normally both provide and take liquidity as part of their strategies, traded proportionally more as volume increased, and overall were net sellers in the rapidly declining broad market along with most other participants”²¹⁸.

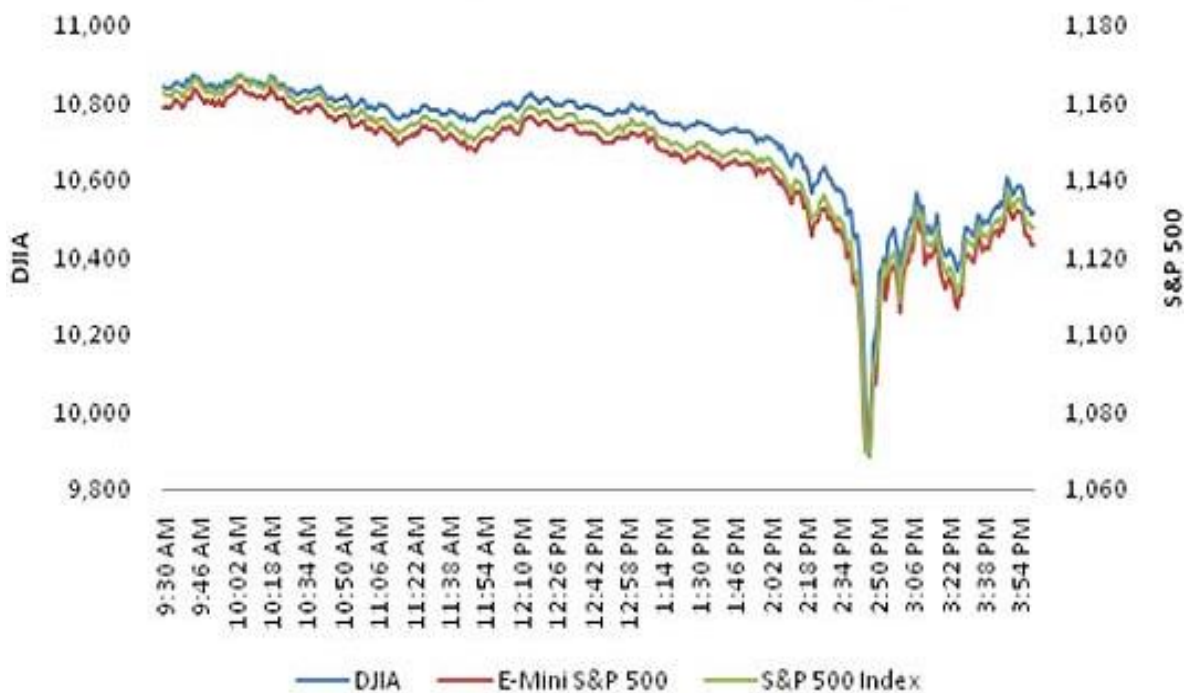


Figure 13: Equity Indices and Equity Index Futures, May 6, 2010. Source: Preliminary Findings of May 6, 2010.

Reconstruction of the facts

The SEC-CFTC report divided the Flash Crash event of May 6 into 5 phases.

During the first phase, from the open through about 2:32 pm, prices were broadly declining across markets. This was due to the disturbing news regarding the European debt crisis; specifically, you noticed a rise of the premiums for buying protection against default (Credit Default Swaps) by the Greek government on their sovereign debt. Furthermore, around 1 pm EST, the Euro started a sharp decline against both the U.S. Dollar and Japanese Yen²¹⁹.

This negative market sentiment was already affecting an increase in the price volatility of some individual securities by reaching levels above the common ones, and by 2:30, selling pressure

²¹⁷ *Id*

²¹⁸ Report of the Staffs of The CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues (September 30, 2010), *Findings Regarding the Market Events of May 6, 2010*, 5

²¹⁹ *Id*, 1

had pushed DJIA down about 2.5% and buy-side liquidity in the E-Mini had fallen by over 50%, and in the S&P 500 SPDR Exchange Trade Fund (SPY) by 20%²²⁰.

In the second phase, from 2:32 through 2:41, the broad markets began to lose another 1-2% triggered by a large fundamental seller who initiated a program to sell a total of 75,000 E-Mini contracts via an automated execution algorithm based on the trading volume (9% execution rate over the previous minute) without taking into consideration neither price nor time²²¹.

The algorithm was executed extremely rapidly, in just 20 minutes, because of high trading volume that affected markets²²².

Three types of buyer absorbed this large sell pressure: HFT firms and intermediaries in the futures markets, fundamental buyers in the same markets and cross-market arbitrageurs who, by purchasing the E-Mini future contracts and selling SPY or individual equities in the S&P 500, transferred this sell pressure to the equities markets²²³.

Therefore, HFTs were the buyers of the initial batch of orders submitted by the sell algorithm and then, they built up temporary long positions of about 3,300 contracts but only temporarily because as well known, HFTs don't take sizable inventory positions in one direction in their common strategies²²⁴.

During the third phase, between 2:41 and 2:45.28, volume spiked upwards and the broad markets plummeted a further 5-6% to reach intra-day lows of 9-10%²²⁵.

In these few minutes, HFTs started to aggressively sell E-Mini contracts to reduce their temporary net long positions, by competing for taking liquidity with the large fundamental seller, who responded to the sudden increased volume by increasing the rate at which it was feeding the orders into the market²²⁶.

Remarkably, from 2:45.13 and 2:45.27, HFTs traded over 27,000 contracts generating a “*hot-potato*” effect due to repeated buying and selling among them but by net buying only 200 contracts. As a result, buy-side market depth in the E-Mini, already very low, fell to less than 1% of its morning level depth and the price of E-Mini plummeted by 1.7% in just 15 seconds²²⁷.

²²⁰ Report of the Staffs of The CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues (September 30, 2010), *Findings Regarding the Market Events of May 6, 2010*, 1-2

²²¹ *Id.*, 2

²²² *Id.*, 2

²²³ *Id.*, 3

²²⁴ *Id.*, 3

²²⁵ *Id.*, 9

²²⁶ *Id.*, 3

²²⁷ *Id.*, 3-4

In the fourth phase, from 2:45.28 to 3:00, broad market indices started to recover while at the same time many individual securities and ETFs underwent extreme price fluctuations and traded at prices as low as one penny or as high as 100,000\$²²⁸.

At 2:45.28, trading on the E-Mini was paused for 5 seconds by a circuit breaker called the CME Stop Logic Functionality in order to prevent a cascade of further price declines. During the halt trading, sell-side pressure was partly alleviated and thus, buy-side interest increased²²⁹.

At 2:45.33, trading resumed and the E-mini as well as the SPY started to recover, while the sell algorithm program continued to execute until 2:51 p.m. as the prices were rapidly increasing²³⁰. Nevertheless, a second liquidity crisis took place in the equities markets caused by the break of automated trading systems used by market makers and other liquidity providers in reaction to the sudden price declines. These built-in pauses are designed to prevent automated systems from trading when prices move beyond pre-defined thresholds. In response to this situation, some of these traders widened their quote spreads, others reduced offered liquidity, and a significant number withdrew completely from the markets²³¹.

Between 2:40 and 3:00, as liquidity completely evaporated in a number of individual stocks, participants issuing market orders found not immediately available liquidity resulting in trades being executed at irrational prices as low as one penny or as high as \$100,000. These trades occurred as a result of so-called “*stub quotes*”, which are generated by market makers at levels far away from the current market in order to fulfill continuous two-sided quoting obligations²³². In the fifth and last phase, starting at 3:00, prices of most individual securities significantly recovered and trading resumed in a more orderly fashion²³³. Indeed, by 3:08, accelerating demand from both opportunistic and fundamental buyers lifted the E-Mini prices back to nearly their pre-drop level²³⁴.

The role of High Frequency Traders

The analysts of SEC and CFTC tried to study the role of the 12 major HFT firms that traded in the equities on May 6, 2010, in order to figure out their degree of participation during the Flash Crash.

²²⁸ *Id.*, 9

²²⁹ *Id.*, 4

²³⁰ *Id.*, 4

²³¹ *Id.*, 4-5

²³² *Id.*, 5

²³³ *Id.*, 9

²³⁴ *Id.*, 16

Based on analysis of Financial Industry Regulatory Authority (FINRA) data, HFT activities increased significantly from 2:43 to 2:46, during the period in which the broad indices were rapidly declining, while in the other minutes from 2:00 to 3:00, they traded in lower scale²³⁵.

Six of the twelve HFTs scaled back their trading when the broad indices hit their lows at about 2:46 pm. Specifically, two HFTs largely stopped trading at 2:47 and remained inactive through the rest of the day; four other HFTs appear to have each significantly curtailed trading for a short period of time from as little as one minute (from 2:46 to 2:47) to as long as 21 minutes (from 2:57 to 3:18)²³⁶.

Globally, HFT firms were primarily sellers on May 6; a portion of this selling of securities could be attributable to cross-market strategies in which they contemporaneously buying a future product and selling ETFs or stocks; in addition, HFTs may have engaged in cross-product strategies of buying ETFs and selling stocks²³⁷.

HFT firms sold more aggressively during the rapid price decline in the period ending 2:45, removing significant buy liquidity from the public quoting markets²³⁸.

Lessons learned after the Flash Crash

One key lesson learned by means of the Flash Crash is that under stressed market conditions, the automated execution of a large sell order can trigger extreme price movements, especially if the automated execution algorithm does not take prices and time into account. Moreover, the interaction between automated execution programs and HFT strategies can quickly erode liquidity and result in disorderly markets. As the events of May 6 demonstrate, especially in times of significant volatility, high trading volume is not necessarily a reliable indicator of market liquidity²³⁹.

In addition, the nature of cross-market trading activity between derivatives and securities markets performed by HFTs may contribute to a propagation of a crash and thus, to systemic risks²⁴⁰.

Another important lesson is that many market participants employ their own versions of a trading pause based on different combinations of market signals; in fact, a market liquidity

²³⁵ Report of the Staffs of The CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues (September 30, 2010), *Findings Regarding the Market Events of May 6, 2010*, 45-48

²³⁶ *Id*

²³⁷ *Id*

²³⁸ *Id*

²³⁹ *Id*, 6-8

²⁴⁰ *Id*, 6-8

crisis can develop if many market participants withdraw at the same time resulting in trades that could be executed at stub prices²⁴¹.

As experienced by the CME's Stop Logic Functionality that triggered a halt in E-Mini trading, pausing a market can be an effective way of providing time for market participants to reassess their strategies, for algorithms to reset their parameters, and for an orderly market to be re-established²⁴².

5.7.2. Mini-Flash Crashes

In the aftermath of the Flash Crash event, several observers, including officials of the CFTC and SEC, noticed that so-called mini-crashes, which are significant and precipitous drops in the prices of individual securities but which do not reach the level of the 2010 Flash Crash, appear to be fairly common and ongoing feature of the current capital market.

Mini Flash Crashes are abrupt and severe price changes involving equities, bonds, foreign exchanges, and derivatives that occur in an extremely short period, typically in milliseconds, with a subsequent rapid recovery.

The discovery of Mini Flash Crashes and related public attention was made possible thanks to the researches of Nanex Llc., a firm employed in the high-frequency trading analysis, that performed a thorough study of the US stock markets in the aftermath of the May 6th Flash Crash, ascribing chiefly these crashes to HFT activity.

Nanex has provided, also, a meticulous definition for a Mini Flash Crash attributing the following conditions:

- (i) It must tick down (up) at least 10 times before ticking down (up);
- (ii) Price changes must occur within 1.5 seconds;
- (iii) Price change has to exceed 0.8%²⁴³.

Such mini-crashes may occur both in the upside and downside, even though the down-crashes are much more concerning for public attention; generally, down-crashes are labeled as "*flash crashes*", while up-crashes are dubbed as "*flash dashes*"²⁴⁴.

Mini Flash Crashes have occurred in the history very frequently with the introduction of automation in the financial markets. Herein, we report some remarkable episodes.

²⁴¹ *Id.*, 6-8

²⁴² *Id.*, 6-8

²⁴³ Anton Golub, John Keane, and Ser-Huang Poon (2012), *High Frequency Trading and Mini Flash Crashes*, SSRN, 5-6

²⁴⁴ *Id.*

The 23th of April 2013, DJIA registered a 143 point fall (about 1%) in few minutes following a false tweet posted by hackers on Associated Press twitter, about a white-house bombing that left Barack Obama injured, shown in the figure 14²⁴⁵.



Figure 14: AP's False tweet.

This dramatic news led automated algorithms' HFT to immediately withdraw buy limit orders and to submit marketable sell orders to forerun the bearish trend. Thus, the Dow Jones plummeted and as soon as the information was denied, HFTs closed their short positions and the market recovered within a few minutes²⁴⁶.



Figure 15: AP Tweet Flash Crash on April 23, 2013. Source: FactSet, MarketWatch.

²⁴⁵ Alfonso Puorro (2013), *High Frequency Trading: una panoramica*, Questioni di Economia e Finanza n° 198, Banca d'Italia, 27-31

²⁴⁶ *Id*

Flash Crashes have affected many of the most capitalized stock equities in the world during the recent years; notorious examples are Google on 22nd April 2013²⁴⁷, which dropped more than 3% in about ¾ of a second before reversing course a second later and Apple on 1st December 2014²⁴⁸, where stock lost over 3% in one minute, falling as much as 6.4 percent in the following minutes before recovering quickly. Both flash crashes were characterized by spike volumes and thinning liquidity during the price downfall where it would seem HFTs have played a determining role.

They hit also US government bonds, which are normally seen as safe and stable investments; on October 15, 2014²⁴⁹, US Treasuries and related markets experienced one of their largest intraday changes in yields in the past 25 years. Yields on 10-year bonds fell by 37 basis points before rebounding quickly.

Gold price experienced a flash crash and a flash dash in the same day on 13th September 2012²⁵⁰, trading was so furious that CME circuit breakers triggered and halted the futures contract twice. Gold underwent additional mini flash-crashes afterwards; on 6th July 2014²⁵¹ for instance, it collapsed by 4% in just under 100 milliseconds and a year and half later, the CME decided to fine a HFT firm for having caused, in that situation, a mass entry of order messages which resulted in “a disruptive and rapid price movement”.

For regulators and academic researchers, the main attention of a Mini Flash Crash episode is constituted by the possible causes that have triggered out its occurrence and the consequent actions taken by market participants during the falling down in prices and the successive recovery, in order to adopt suitable measures to restrain their appearances and to reveal potential unlawful and manipulative behaviors carried out by traders.

Concerning the causes, there is not a univocal explanation about the ignition of a Mini Flash Crash inasmuch a fast dropping in price may be due of several conjunctions.

Golub et al (2012)²⁵² find that they are the result of regulation framework and market fragmentation, in particular due to the aggressive use of Intermarket Sweep Orders (ISO)²⁵³ and Regulation NMS protecting only top of the book; they find strong evidence that mini-flash crashes have an adverse impact on market liquidity and are associated with fleeting liquidity.

²⁴⁷ Nanex, 22 April 2013, *Google Flash Crash*, available at: <http://www.nanex.net/aqck2/4171.html>

²⁴⁸ *Apple tumbles as much as 6 percent in unusual trading*, Reuters December 1 2014, available at: <http://www.reuters.com/article/us-apple-shares-idUSKCN0JF2M420141201>

²⁴⁹ Nanex 15 October 2014, *Treasury Flash Crash*, available at: <http://www.nanex.net/aqck2/4681.html>

²⁵⁰ Nanex 13 September 2012, *Double Flash Crash in Gold*, available at: <http://www.nanex.net/aqck2/3578.html>

²⁵¹ Nanex 6 January 2014, *First Gold Halt of 2014*, available at: <http://www.nanex.net/aqck2/4522.html>

²⁵² Anton Golub, John Keane, and Ser-Huang Poon (2012), *High Frequency Trading and Mini Flash Crashes*, SSRN

²⁵³ An ISO is a limit order which is designed to be executed at one market center exclusively regardless of whether or not other market centers have better quotations published. It is an exception of NMS Rule 6111

In practice this manifest as a burst of trades from one exchange, as remaining orders rips through the book due to the lack of depth of the book protection.

Johnson et al (2012)²⁵⁴ suggest that such crashes are likely be a result of interaction between several trading algorithms, or a positive feedback loop induced by market environment, and are not simply the product of some pathological regulatory rule for crashes.

Some critical observers, such as Nanex firm, have suggested that a contributing factor behind some of these mini-crashes is HFT.

Indeed, HFTs appear to play a relevant role in these mini-flash crashes as happened in 2010 Flash Crash. First, they may all react at the same time to market signals by placing buy or sell market orders that might consume most of current liquidity, by triggering sharp price movements as stop limit orders get hit by market orders from HFTs. Second, their ability to rapidly leave the market and cancel pending orders as soon as market conditions are extremely turbulent, it facilitates the possibility of a sudden market destabilization.

SEC officials have responded that those who “*try to use instances of mini-flash crashes as clear and incontrovertible evidence of the problems with high-frequency trading, high-speed markets, fragility, and impending doom...may be looking in the wrong places*”. Instead, the officials attributed such developments to various kinds of human errors, including inadequate risk management practices in which there has been a “lack of checks and balances”²⁵⁵.

²⁵⁴ Neil Johnson, Guannan Zhao, Eric Hunsader, Jim Meng, Amith Ravindar, Spencer Carran, and Brian Tivnan (2012), *Financial Black Swans Driven by Ultrafast Machine Ecology*, SSRN

²⁵⁵ Gary Shorter, Rena S. Miller (2014), *High-Frequency Trading: Background, Concerns, and Regulatory Developments*, Congressional Research Service, 30

6. HIGH FREQUENCY TRADING REGULATION

6.1. Regulatory Activity and Discussion

Regrettable recent market events such as flash crashes, manipulative conducts and cases of runaway algorithms have elicited regulators to meditate on and to take special notice of HFT. Policymakers started to develop a number of initiatives, consultations and proposals aimed at more strictly regulating HFT. Such task is mainly entrusted to regulatory authorities directed at supervising capital markets in the equities and futures context.

In the US, the SEC is the authority with the assignment to oversee HFT for securities markets and the more limited securities-related derivatives; while, the CFTC supervises trading on futures exchanges such as the Chicago Mercantile Exchange (CME) and International Exchange (ICE)²⁵⁶. Moreover, since capital markets are particularly fragmented and interconnected, regulatory initiatives are often characterized by the coordination and joint activity among regulators of both equity and equity derivative markets.

In Europe, HFT regulation is part of Directive MiFID II²⁵⁷ adopted on 2014 by EU Parliament with the support of European Securities and Markets Authority (ESMA), which, however, needs to be implemented by the individual EU members within 2018 in order to be conform to the different member states' regulatory framework.

6.2. HFT Regulation in the US

In the aftermath of Flash Crash event, SEC and CFTC commenced adopting a number of HFT-related programmatic regulatory initiatives intended to provide for investor protection and to maintain fair, orderly and efficient markets in a high-frequency framework. The developments also help better monitor HFT activity to stem potential disruptive impact on the US trading system.

Co-location

In June 2010, the Commodity Futures Trading Commission (CFTC) proposed a rule which was intended to assure equal and fair access to co-location/proximity hosting services. The CFTC argues that these services offer a significant competitive advantage for high frequency traders

²⁵⁶ Rena S. Miller, Gary Shorter (2016), *High Frequency Trading: Overview of Recent Developments*, Congressional Research Service, 2

²⁵⁷ MiFID II available at: <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014L0065&from=EN>

and therefore have to be equitably accessible. To ensure fair and open access for all traders, the regulator proposed that marketplaces and third parties listing significant price discovery contracts (SPDCs)²⁵⁸ should implement uniform fees for co-location and associated services and that their cost is not used as a means to deny access to some market participants by pricing them out of the market. Furthermore, the CFTC wanted to increase the latency transparency and therefore proposed to make available to the public information about the longest, shortest, and average latencies. Lastly, the regulator's intention was to secure that sufficient co-location space was available and to ensure that shortages in co-location space could not impair fair access²⁵⁹.

Stub Quotes

A Stub Quote is an offer to buy or sell a stock at price so far away from the prevailing market that it is not intended to be executed, as an order to buy at a penny or an offer to sell at \$100,000. During the Flash Crash, market makers used stub quotes to nominally comply with their obligations to maintain a two-sided quotation; however, the sudden loss of liquidity due to traders withdrawing, led to their executions at that time.

Following the Flash Crash episode, the SEC has imposed a ban on placing stub quotes, by preventing that trades are executed at irrational prices and reducing the need that quotations be then broken if the markets become particularly volatile²⁶⁰.

The new adopted rules require market makers in exchange-listed equities to maintain continuous two-sided quotations during regular market hours that are within a certain percentage band of the NBBO²⁶¹.

Circuit Breakers

Circuit breakers (or impediments to trade) are a tool designed to reduce the risk of a cascading price collapse by temporarily halting trading in presence of excessive price volatility.

They were introduced for the first time following the "Black Monday" market crash of October 19, 1987.

²⁵⁸ Significant Price Discovery Contracts are contracts that are linked to existing exchange-traded contracts: are traded on an electronic exchange; and perform a "significant price discovery function". Bracewell and Giuliani (2009)

²⁵⁹ Peter Gomber, Björn Arndt, Marco Lutat, Tim Uhle (2011), *High-Frequency Trading*, Goethe Universitat Frankfurt Am Main, 43

²⁶⁰ *SEC Approves New Rules Prohibiting Market Maker Stub Quotes* (2010), available at: <https://www.sec.gov/news/press/2010/2010-216.htm>

²⁶¹ *Id*

However, before the 2010 Flash Crash, they were exclusively conceived on market-wide basis and triggered only by very large price movements, i.e. they halted trading in all exchange-listed securities throughout the US markets in the case of serious market decline percentage thresholds in reference to DJIA benchmark.

In order to prevent mini-flash crashes and to deal with periodic illiquidity in markets, the SEC and CFTC, therefore, updated the preceding market-wide circuit breakers and introduced a limit up-limit down mechanism for individual securities.

The updated market-wide circuit breakers have reduced the market decline percentage thresholds necessary to trigger a circuit breaker, have shortened the duration of the resulting trading halts, and have changed the reference index used to measure a market decline, passing from DJIA to S&P 500 Index²⁶².

In 2012, the SEC adopted a “limit up-limit down²⁶³” mechanism to address extraordinary market volatility in individual securities. The mechanism is intended to prevent trades in individual exchange-listed stocks from occurring outside of a specified price band, set as a percentage level above and below the average price of the stock over the immediately preceding five-minute trading period. The price limit bands are 5%, 10%, 20% or 75% depending on the price of the stock, and are doubled in size during the often more volatile opening and closing periods of the trading day. Once triggered, trading in a stock will be paused for a minimum of five minutes on all nationwide markets if the NBBO price matches one of the upper or lower band limits for at least 15 seconds²⁶⁴.

Naked Access

Before the Flash Crash, many HFT firms gained special access to securities exchanges through “*naked access*”, a process through which registered brokers allowed the firms to basically piggyback on their direct access to securities market. The arrangement enabled HFT firms to reduce their trade latency while avoiding the various risk checks and capital requirements, which they would have needed to comply with had they been registered traders²⁶⁵.

²⁶² SEC (2011), *SEC to Publish for Public Comment Updated Market-Wide Circuit Breaker Proposals to Address Extraordinary Market Volatility*, available at: <https://www.sec.gov/news/press/2011/2011-190.htm>

²⁶³ SEC (2012), *SEC Approves Proposals to Address Extraordinary Volatility in Individual Stocks and Broader Stock Market*, available at: <https://www.sec.gov/oiea/investor-alerts-bulletins/investor-alerts-circuitbreakersbulletinhtm.html>

²⁶⁴ *Id*

²⁶⁵ Gary Shorter, Rena S. Miller (2014), *High-Frequency Trading: Background, Concerns, and Regulatory Developments*, Congressional Research Service, 32-33

In November 2010, the SEC adopted a “Market Access Rule²⁶⁶”, by obligating registered brokers to put in place risk management controls and supervisory procedures to help prevent erroneous orders, ensure compliance with regulatory requirements, and enforce pre-set credit or capital thresholds, which essentially prohibited HFT firms from receiving naked access.

Regulators’ Access to Information

In 2011, the SEC adopted the “Large Trader Reporting Rule²⁶⁷”, which imposed an identification and reporting obligations for large traders.

Large traders are recognized according to volume and/or value traded, and therefore include also traders employing rapid algorithmic systems for quoting and trading in huge volumes.

Such rule empowered regulators to obtain relevant information in 1) assessing the impact of large traders’ activity on the securities markets; 2) reconstructing trading activity following periods of unusual market volatility; and 3) analyzing significant market events for regulatory purposes²⁶⁸.

This monitoring activity has been further supplemented by the implementation of a “*consolidated audit trail system*”.

The Consolidated Audit Trail System Rule²⁶⁹, adopted in 2012, is aimed at giving regulators the ability to monitor trading activity and analyze atypical events across the fragmented US securities markets.

The rule requires US securities exchanges to establish a market-wide system for collecting and for accurately identifying every order, cancellation, modification and trade execution for all exchange-listed equities and equity options²⁷⁰.

With the audit trail in place, SEC is able to receive real time access to most of the data needed to reconstruct a market dislocation such as a flash crash.

In 2015, the SEC took steps toward a registration requirement for certain HFT brokers-dealers, which requires them to register with the Financial Industry Regulatory Authority (FINRA), a self-regulatory organization which acts as the front-line regulator for broker-dealers²⁷¹.

FINRA’s registrants are subject to examinations, various disclosure requirements, and rules governing various aspects of their conduct.

²⁶⁶ See “Current Regulatory Initiatives”, Rule 15c3-5 – Market Access Rule, NASDAQ, available at <https://www.nasdaqtrader.com/Trader.aspx?id=currentregulatory>

²⁶⁷ *Id.*, Large Trader Reporting

²⁶⁸ *Id.*

²⁶⁹ *Id.*, Consolidated Audit Trail

²⁷⁰ *Id.*

²⁷¹ Rena S. Miller, Gary Shorter (2016), *High Frequency Trading: Overview of Recent Developments*, Congressional Research Service, 11

Systems Compliance and Integrity

In November 2014, the SEC adopted the Regulation Systems Compliance and Integrity (SCI) which created new, enforceable standards for maintaining and testing the trading systems used by securities exchanges and brokers. Specifically, the rules are designed to: reduce the occurrence of systems issues, improve resiliency when systems problems do occur, enhance the SEC's oversight and enforcement of securities market technology infrastructure²⁷².

Anti-Spoofing Efforts

The 2010 Dodd-Frank Act amended the Commodity Exchange Act (CEA) to expressly prohibit certain disruptive trading practices, commonly known as “spoofing”, that represent bidding or offering with the intent to cancel the bid or offer before execution²⁷³.

Applying such a provision on spoofing to the HFT world, however, can be challenging, because usually high-speed computers and algorithms automatically generate many bids and offers in a millisecond, and cancel them quickly, therefore it can be difficult to ascertain at time whether such automated trading practices rise to the level of spoofing.

Nevertheless, the CFTC and the SEC have used their new anti-spoofing authorities in a number of recent enforcement actions.

For example, in 2015, the SEC reached a \$1 million settlement with Briargate Trading LLP, a HFT firm because between 2011 and 2012, Briargate was charged with orchestrating a scheme that involved placing sham trades called spoof orders for the purpose of creating “*the false appearance of interest in stocks*” to manipulate their prices²⁷⁴.

6.3. HFT Regulation in the European MiFID II

The Markets in Financial Instruments Directive II adopted specific rules to subject HFT and other forms of AT to particular scrutiny as they entail some risks to the capital market structure. Specifically, Article 17 sets rules on subject of algorithmic trading, including also the sub-category of HFT.

It introduces requirements to be met by investment firms that apply AT and HFT in relation to internal systems and controls; imposes duties regarding information and records on these parties; and imposes specific requirements on investment firms that apply AT and HFT as

²⁷² SEC Regulation SCI (2013) available at: <https://www.sec.gov/spotlight/regulation-sci.shtml>

²⁷³ See note 271, 7

²⁷⁴ *Id.*, 12

market makers. It sets, further, obligations for brokers that provide direct electronic access to HFT firms.

In addition, article 48 disciplines some responsibilities that trading venues which operate with HFT firms must follow in subject of systems resilience, circuit breakers, fee structure and low-latency infrastructures.

HFT internal systems and controls

MiFID II introduces requirements for internal systems and controls to be met by investment firms that engage in AT and HFT.

The internal systems and controls of the HFT firm must be such that they ensure that the firm's trading systems are resilient and have sufficient capacity, are subject to appropriate trading thresholds and limits, and prevent the sending of erroneous orders or the systems otherwise functioning in a way that may create or contribute to a disorderly market²⁷⁵.

Such investment firm, moreover, shall have in place effective business continuity arrangements to deal with any failure of its trading systems and shall ensure its systems are fully tested and properly monitored²⁷⁶.

The detailed requirements have been set out by ESMA in draft regulatory technical standard²⁷⁷. In particular, it specifies requirements for the resilience of trading systems of investment firms which involves testing the algorithms, monitoring and, where necessary, changing the algorithms used, annual stress testing, incorporating a kill functionality so that all resting orders can be cancelled in the event of an emergency, monitoring for the prevention and identification of potential market abuse, carrying out pre-trade controls on order entry, carrying out post-trade controls.

HFT information and records

Investment firm that engages in AT and HFT has certain duties regarding information and records that have to be communicate to the competent authority of its Member State and to the trading venues at which it is member or participant²⁷⁸.

The competent authority may require the investment firm to provide, on a regular and ad hoc basis, a description of the nature of its AT and HFT strategies, details of the trading parameters

²⁷⁵ Article 17 (1) of European MiFID II

²⁷⁶ *Id*

²⁷⁷ ESMA/2015/1464, *Regulatory Technical and Implementing Standards*, Annex I MiFID II / MiFIR (28 September 2015), 201–46

²⁷⁸ Article 17 (2) of European MiFID II

or limits to which the system is subject, a description of the key compliance and risk controls that it has in place, and information about the testing of its systems²⁷⁹.

Investment firm that, specifically, engages in high-frequency algorithmic trading technique must store in an approved form accurate and time sequenced records of all its placed orders, including cancellations of orders, executed orders and quotations on trading venues and shall make them available upon request²⁸⁰.

HFT market makers

MiFID II introduces specific requirements for investment firms engaged in AT and HFT in pursuance of a market making strategy defined as involving posting firm, simultaneous two-way quotes of comparable size and at competitive prices relating to one or more financial instruments on a single trading venue or across different trading venues²⁸¹.

The investment firm must, taking into account the liquidity, scale and nature of the specific market and the characteristics of the instruments traded, carry out this market making activity continuously during a specified proportion of the trading venue's trading hours with the result of providing liquidity on a regular and predictable basis to the trading venue²⁸².

Nonetheless, the obligation to make public bid and offer prices on a continuous basis is not absolute, in fact in exceptional circumstances the market maker is no longer obliged to do so. These include circumstances of extreme volatility, political and macroeconomic issues, system and operational matters, and circumstances which contradict the investment firm's ability to maintain prudent risk management practices²⁸³.

Direct Electronic Access

MiFID II imposes rules for providers of *direct electronic access* (DEA), regarding controls, information, and relationship with their clients.

An investment firm that provides DEA must have in place effective systems and controls which ensure a proper assessment and review of the suitability of clients using the service, that clients using the service are prevented from exceeding appropriate pre-set trading and credit thresholds, that trading by clients using the service is properly monitored and that appropriate

²⁷⁹ *Id*

²⁸⁰ *Id*

²⁸¹ Article 17 (4) of European MiFID II

²⁸² Article 17 (3) of European MiFID II

²⁸³ Article 17 (7) of European MiFID II

risk controls prevent trading that may create risks to the investment firm itself or that could create or contribute to a disorderly market²⁸⁴.

The DEA provider shall monitor the clients' transactions in order to identify infringements of rules, disorderly trading conditions or conduct that may involve market abuse and that is to be reported to the competent authority²⁸⁵.

The controls applied to sponsored access shall be at least equivalent to those applied to direct market access²⁸⁶. This entails that any form of sponsored unfiltered access is forbidden in the EU, as well.

HFT and trading venues

MiFID II also specifies rules for the trading venues interacting with AT and HFT.

Generally, trading venues must have in place effective systems, procedures and arrangements to ensure their trading systems are resilient, have sufficient capacity to deal with peak order and message volumes, are able to ensure orderly trading under conditions of severe market stress, are fully tested to ensure such conditions are met and are subject to effective business continuity arrangements to ensure continuity of their services if there is any failure of their trading systems²⁸⁷.

Furthermore, they must have in place effective systems to reject orders that exceed pre-determined volume and price thresholds or are clearly erroneous²⁸⁸.

Trading venues must be able to temporarily halt or constrain trading if there is a significant price movement in a financial instrument on that market or a related market during a short period and, in exceptional cases, to be able to cancel, vary or correct any transaction²⁸⁹.

However, if HFT and AT traders are connected to these trading venues, the internal systems and controls are subject to additional requirements; trading venues must have systems to limit the ratio of unexecuted orders to transactions that may entered by a HFT firm, and be able to slow down the flow of orders if there is a risk of their systems capacity²⁹⁰.

Additionally, a trading venue must ensure that its fee structure is transparent, fair and non-discriminatory and that does not create incentives to place, modify or cancel orders or to execute transactions in a way which contributes to disorderly trading conditions or market abuse²⁹¹.

²⁸⁴ Article 17 (5) of European MiFID II

²⁸⁵ *Id*

²⁸⁶ Article 48 (12) of European MiFID II

²⁸⁷ Article 48 (1) of European MiFID II

²⁸⁸ Article 48 (4) of European MiFID II

²⁸⁹ Article 48 (5) of European MiFID II

²⁹⁰ Article 48 (6) of European MiFID II

²⁹¹ Article 48 (9) of European MiFID II

For trading venues allowing co-location services, in order to ensure orderly and fair trading conditions, it is essential that they provide such co-location services on a non-discriminatory, fair and transparent basis²⁹².

6.4. Ideas and Proposal Measures

In parallel with the adopted measures or about to, further discussions have raised the need of more pressing policies to heavily tighten HFT activity, or at least to disincentivize its more aggressive conducts.

Some of such proposals have been implemented by single countries or trade exchanges, registering discordant results in terms of cost-benefit analyses for the entire market.

Below, we report a number of the potential HFT regulatory proposals that have become part of the public policy discourse.

Order Cancellation Fees

Some observers argue that by imposing penalty charges for excessive order cancellations, HFTs would be discouraged from posting orders they do not intend to execute or using cancellations as a part of manipulative strategies, since penalty fees would render such activities unprofitable²⁹³.

Additionally, such fees would discourage the most flagrant excessive cancellations which come along with higher volatility and would result in a consistent reduction in order messages that would prevent overload in the exchange computer systems²⁹⁴.

Opponents to the proposal respond that mostly order cancellations are the result of the rapid reaction to new information and is often a way for HFT market makers to minimize the risks of offering prices to other traders, and therefore, an imposition on that orders would likely reduce the provision of liquidity, thus reducing market depth²⁹⁵.

Both US and EU regulators have called for the imposition of some cancellation fees. For example, MiFID II suggests that regulated markets should impose a higher fee for placing an order that is subsequently cancelled than an order which is executed and to impose a higher fee on participants placing a high ratio of cancelled orders to executed orders²⁹⁶.

²⁹² Article 48 (8) of European MiFID II

²⁹³ Gary Shorter, Rena S. Miller (2014), *High-Frequency Trading: Background, Concerns, and Regulatory Developments*, Congressional Research Service, 34

²⁹⁴ Matt Prewitt (2012), *High-Frequency Trading: Should Regulators Do More*, Michigan Telecommunications and Technology Law Review 19, 131

²⁹⁵ See note 293

²⁹⁶ Article 48 (9) of European MiFID II

CFTC's 2013 Concept Release has solicited greater system safeguards to protect trading against potential abuses or disruption unique to electronic trading. Among such measures, it includes controls related to order cancellation protocols with the introduction of an extra fee for excessive cancellations²⁹⁷.

However, such impositions have been only incompletely enforced in barely some exchanges: on Nasdaq and Direct Edge, on Borsa Italiana and Deutsche Boerse exchanges.

Minimum Resting Times

Under this scheme, HFT limit orders must remain in the trading book and then cannot be cancelled within a pre-determined minimum time span, say 50 milliseconds²⁹⁸.

Some argue that such a requirement would be another means of curbing what many perceive to be the problematic and excessive use of cancelled orders by HFTs.

Minimum resting times obligation is intended to increase the likelihood that viewed quotes being available to trade and therefore to make the order book dynamics more transparent. In addition, longer order exposure times would create more liquidity and would reduce price variance in the market. Lastly, the slow-down of markets might favor participation, especially if some traders (e.g. small retail investors) feel that high speed makes market unfair and hurts market integrity²⁹⁹.

Detractors, however, argue that such a protocol would hinder HFT firms to provide liquidity during times of high volatility because it would be particularly expensive to post limit orders considering the impossibility to cancel "stale" orders. Conversely, this measure would not affect HFT liquidity demanders with the result to attract more aggressive HFTs in the market³⁰⁰.

Affirmative Trade Obligations

Some suggest that consideration be given to imposing certain affirmative trade obligations on these HFT firms who are not registered broker-dealers and thus are not legally obligated to step in and provide needed liquidity, particularly during market disruptions similar to the Flash Crash where the market liquidity is particularly scarce.

An imposition in such sense, it would render, in effect, HFTs as market-makers even though, they unlike specialists do not enjoy the various privileges that come with them, such as the

²⁹⁷ Gary Shorter, Rena S. Miller (2014), *High-Frequency Trading: Background, Concerns, and Regulatory Developments*, Congressional Research Service, 40-41

²⁹⁸ *Id.*, 34

²⁹⁹ Sandrine Jacob Leal and Mauro Napoletano (2016), *Market Stability vs. Market Resilience: Regulatory Policies Experiments in an Agent Based Model with Low- and High-Frequency Trading*, Working Paper, 18-20

³⁰⁰ *Id.*

access to information about order flows, the right to make decisions after others make their decisions, the ability to create the market quote, the ability to create and exercise certain look-back timing options, and the right to collect brokerage commissions from executing system order flow³⁰¹.

Supporters claim that HFTs, thanks to their shorter latency, have already an important competitive advantage and the imposition of an affirmative market-making obligation is only a measure for limiting the potential negative effect that their market withdrawal could generate during extreme volatility events.

Financial Transaction Tax

A financial transaction tax is a fee on HFT trades as a way of limiting that kind of trading and its perceived negative consequences.

MiFID II has called regulated markets for the introduction of a financial transaction tax for traders operating a high-frequency algorithmic trading technique in order to reflect the additional burden on system capacity³⁰².

So far different schemes and levels of taxes have been implemented all over the world. Examples are the stamp duty in the UK, the French financial transaction tax on HFT and the pricing scheme introduced on NYSE Euronext³⁰³.

In 2013, Italy imposed a tax on trades on Italian financial markets that are generated by a computer algorithm that automatically determines the decisions related to relevant orders or metrics. The tax rate was fixed to 0.2% of the transaction value with an additional fee of 0.02% for any portion of changed or cancelled daily orders in less than half a second where the ratio exceed 60% of the total number of submitted orders³⁰⁴.

The tax was reportedly introduced due to concerns that the growth of HFT in Italy could potentially have an adverse impact on the integrity and quality of Italian financial markets, particularly with regard to volatility and liquidity.

Nevertheless, an empirical analysis carried out by Ruhl and Stein (2014)³⁰⁵ showed how an introduction of a transaction tax in Italian financial markets has led to an increase in volatility and quoted spreads, in the opposite direction when compared to the assumed intention of regulators.

³⁰¹ Larry Harris (2003), *Trading and Exchanges: Market microstructure for practitioners*, Oxford University Press

³⁰² Article 48 (9) of European MiFID II

³⁰³ See note 299, 24-27

³⁰⁴ Gary Shorter, Rena S. Miller (2014), *High-Frequency Trading: Background, Concerns, and Regulatory Developments*, Congressional Research Service, 36

³⁰⁵ Tobias R. Ruhl and Michael Stein (2014), *The Impact of Financial Transaction Taxes: Evidence from Italy*, Economics Bulletin Volume 34, Issue 1

A Kill Switch

A kill switch procedure would permit the suspension of an individual HFT firm's trades following erroneous trades or an excessive trading volume. Some regulators have argued in favor of a such a protocol as a way of thwarting large-scale market events in which HFT has played a leading role. In the recent regulation developments, there appeared to be widespread agreement that a kill switch could be useful, but that it would require multiple layers and thresholds to ensure that it would not be used at inappropriate times. There were, however, concerns over how and when such mechanisms would be implemented and whether market stakeholders would be willing to "pull the trigger" during market disruptions³⁰⁶.

Proponents suggest that both HFT market participants and trading venues should have the capability to cancel working orders and interrupt trading activity under certain problematic market conditions. Furthermore, trading platforms should have clear, objective policies and procedures detailing circumstances that warrant use of a kill switch, based upon experience about their HFT market participants' trading style and strategies³⁰⁷.

6.5. Considerations on HFT Regulation Developments

HFT Regulation initiatives carried out by competent authorities both in the US and EU are chiefly aimed at imposing a limitation of the verified and perceived negative effects and risks of such trading activity, as cases of registered manipulative conducts and exacerbation of market volatility. The need of a deeper regulation of HFT phenomenon has been considered necessary for the preservation of market integrity and fairness and for a better investor confidence. However, it must be kept in mind that a too strict HFT regulatory regime could produce, instead, the opposite effect, intended as a strongly deterioration of the beneficial effects in terms of market quality obtained thanks to the proliferation of HFT technologies and strategies.

³⁰⁶ Gary Shorter, Rena S. Miller (2014), *High-Frequency Trading: Background, Concerns, and Regulatory Developments*, Congressional Research Service, 37

³⁰⁷ *Id.*, 41

7. EMPIRICAL ANALYSIS OF THE POUND FLASH CRASH

7.1. Introduction of the Pound Flash Crash

The empirical analysis covered in this work concerns the study of a mini-flash crash phenomenon that affected the British Sterling on Friday 7 October 2016, where the Pound depreciated by around 9% versus the Dollar during early Asian trading in less than a minute, before quickly recovering much of the move. The registered free-fall touched a historical minimum of \$ 1.1450 (according to Thomson Reuters data), which represented the lowest reached level by the British currency since March 1985.

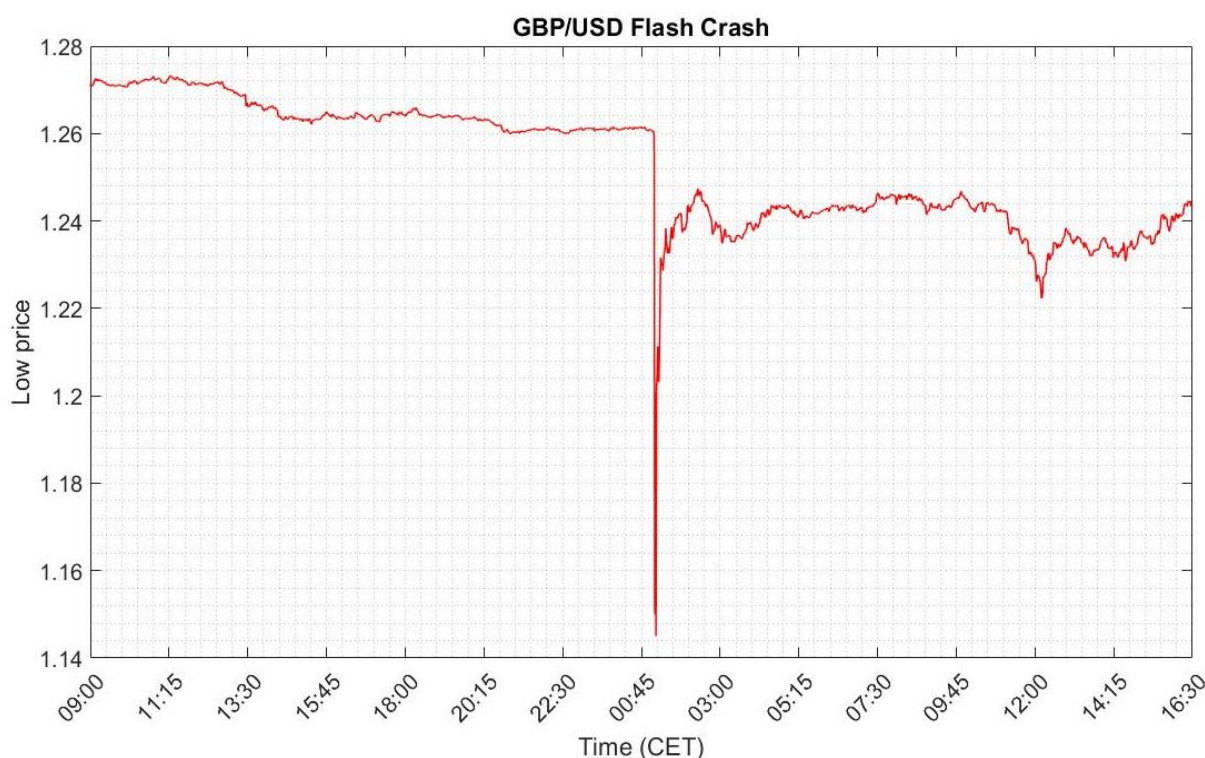


Figure 16: GBP/USD Flash Crash. Source: author's elaboration on Thomson Reuters data.

Figure 16 shows the chart of the continuous minute-by-minute low prices for GBP/USD quotations during 6 and 7 October. It is possible to note the sudden and temporary decline occurred in GBP values against the US Dollar during the overnight period between 6th and 7th of October. Specifically, the exchange rate quotation at 01:08 am CET abruptly collapsed down, passing from a value of 1.2596 to 1.15 in just one minute, a tumble of 8.7 percent magnitude. The following minutes saw a rapid price recovery with the value that fluctuated around at \$ 1.24 for the rest of the trading day. At the end, the British Sterling was down 1.4% as London closed for the weekend.

Such empirical behavior is typical of the mini-flash crashes theory claiming that they occur all of a sudden with a large-scale but with a subsequent quick recovery (a large v-shaped

movement). As other similar occurrences, it cannot be fully explained by incoming news and economic data, but typically, it is characterized by unusual high trading volumes relative to liquidity measures in the event time window. In addition, the speed at which it has unfolded might suggest a determinant role played by automated and high-frequency trading, even though it is not so simple to prove such assertions.

The choice of such event is due to the fact that it occurred recently and for this reason, it was possible to gather some high-frequency data (precisely with one-minute time frame) for an empirical analysis of the phenomenon; in addition, it constitutes a typical example of how mini-flash crashes are becoming very frequent in modern automated financial markets and may hit, in case of particular market and trading conditions, also those markets whose size and liquidity is generally considered as a protection against such events, as the foreign exchange market.

The analysis in this study is mainly focused on a post-examination of the phenomenon. In particular, the specific dynamics of the event, the idiosyncratic factors that have contributed to its occurrence, the general context of the Sterling, and considerations about potential consequences for market participants and market stability are reported.

7.2. The GBP Foreign Exchange Market

Before looking in details at the moves of 7 October, it is necessary to provide context on the foreign exchange market characteristics, how it works and what are the differences with respect to equities markets.

Firstly, foreign exchange market (FX or Forex) is the largest financial market in the world in which participants are able to buy, sell, exchange and speculate on currencies.

It is characterized by a wide fragmentation of trading, with vast majority of trading performed in over-the-counter markets based on cash-for-cash transactions, primarily via risk transfer in principal-to-principal bilateral trading arrangements for: payment for goods and services abroad, investing in overseas assets, hedging foreign currency exposures, and active management of portfolios, including for speculative purpose³⁰⁸.

The FX has enticed retail currency traders from all over the world because of its several benefits. One of the benefits of trading on currencies is its massive trading volume, which covers the largest asset class globally. This implies that Forex traders are provided with high liquidity, then bid-ask spreads are generally tight and levels of volatility are narrow.

³⁰⁸ *The Sterling 'flash event' of 7 October 2016* (2017), Bank for International Settlements, *Appendix A*

Unlike equities markets, Forex trades 24 hours a day, five and half days a week and all over the world, due to the operative continuity incurring from Asian trading opening to US trading closing.

Trading in major currencies is carried out by a multitude of platforms that aggregate and advertise liquidity from providers to consumers, although price discovery is thought to primarily rely on a smaller number of key venues. Some platforms are simply matching buyers and sellers, other operate as exchanges³⁰⁹.

Figure 17 illustrates the market shares for the major FX electronic trading platforms.

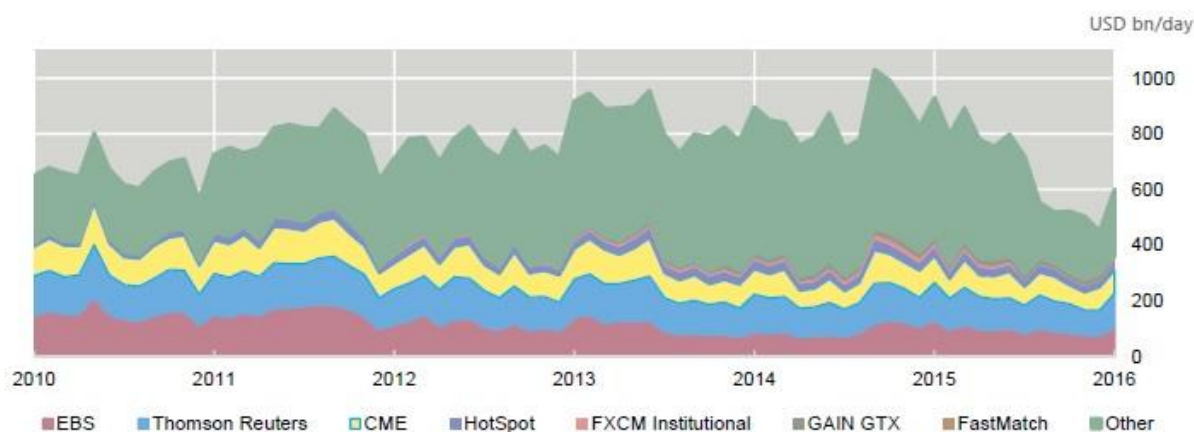


Figure 17: FX spot volume by electronic trading platform. Source: MarketFactory.

There are often differences across platforms in costs, rule books, latencies, and connectivity. Typically, there are no compulsory requirements for spot FX platforms to have trading halt mechanisms: for instance, two of the main spot FX electronic broking platforms (EBS and Thomson Reuters Matching) do have some pre-trade controls, but not circuit breakers, although CME futures exchange does. Each participant is therefore responsible for its own risk management and controls³¹⁰.

Importantly, unlike equities markets, there is no formal obligation for market-makers to provide liquidity. To that end, each price-maker controls its own trading presence, in effect establishing their own bespoke circuit breaker, widening spreads or even withdrawing completely³¹¹.

Regarding the market in question, the GBP/USD is the British Pound and US Dollar currency pair or cross quotation. The currency pair tells how many US Dollars (the quote currency) are needed to purchase one British Pound (the base currency).

Trading on GBP/USD is also known as the “cable”, a term deriving from the 19th century as transactions between the Pound and Dollar were executed via transatlantic cable.

³⁰⁹ *Id*

³¹⁰ *Id*

³¹¹ *Id*

The “cable” value is affected by factors that influence the value of the British Pound and/or the US Dollar in relation to each other. For this reason, the interest rate differential between the Bank of England (BoE) and the Federal Reserve (Fed) affects the value of this quotation. For example, when the Fed intervenes in open market activities to make the Dollar stronger, the value of the GBP/USD cross could decline, *ceteris paribus*.

Another important factor that can affect the value of the currency pair is the political uncertainty; if the governance of a country is at risk of undergoing substantial change, this situation could make investors less likely to demand its currency by preferring the safer one.

7.3. A Flash Event in Three Stages

The Sterling flash crash of 7 October 2016 has been deeply diagnosed by the *Bank for International Settlements* (BIS) who has released a report³¹² into their investigation on the event in order to shed light on the market microstructure dynamics and to try providing an explanation about the nosedive observed in the Pound.

According to the BIS report, the event can be broken down into three different stages.

First, the early phase of the move that lasted for a matter of seconds, during which Sterling depreciated rapidly from 1.26 to around 1.24 against the Dollar in response to a significant selling flow, but in an orderly fashion and with broad participations on key venues.

Second, a period of about ten minutes of extreme dysfunction that saw Sterling fall further, rebound and then trade in a wide range. This phase involved lower volumes and narrower participation, pointing to a greater role for the actions of individual market participants as a driver of the sharp movements.

Third, the gradual recovery over the hours that followed, as liquidity returned to the market.

Stage 1: Shortly after midnight British Summer Time (BST), equivalent to Continental European Time (CET) minus one hour, on 7 October, trading volumes picked up sharply and Sterling began to depreciate against Dollar.

Over a period of around eight seconds, from 00:07:03 to 00:07:11 BST, Sterling fell from 1.2600 to 1.2494, based on the Reuters mid-price³¹³ (figure 18). During this time, there was a surge in volumes (figure 19), with the vast majority of executed trades representing so-called “aggressive” sales of Sterling, pointing to a very significant imbalance in order flow. Despite the magnitude of the move and the volumes transacted, GBP/USD bid-offer spreads remained

³¹² *The Sterling ‘flash event’ of 7 October 2016* (2017), Bank for International Settlements

³¹³ Mid-price is the mid-point between the best bid and ask prices.

little changed until around 00:07:14 BST and measures of the price impact of transactions over this period were relatively low (figure 20)³¹⁴.

At 00:07:13 BST, the Financial Times published an article entitled “*Hollande demands tough Brexit negotiations*” on its website. Market participants noted that this release would have been interpreted as somewhat Sterling-negative by exacerbating the current volatility (figure 18).

Stage 2: At 00:07:15 BST, the CME triggered its velocity logic mechanism, which paused trading for 10 seconds on the futures exchange, in response to the large moves in the preceding two seconds (figure 22). At this point, bid-offer spreads in the spot market widened significantly³¹⁵.

After reaching 1.24 on Reuters, at 00:07:15 BST, GBP/USD accelerated its fall. From this point onwards, and particularly past the 1.22 level, price gapping between trades is increasingly visible (figure 18). And by 00:07:34 BST, 19 seconds later, GBP/USD had reached 1.20 and the move exhausted the resting sterling bids across a variety of electronic trading platform’ order books (figure 21)³¹⁶.

While bids quickly returned to the market, overall depth in the order book remained extremely low for several minutes. Those wishing to trade could only execute in relatively small sizes at prices at a wide spread to the implied mid-price (figure 20). For example, on Reuters prices as low as 1.1491 in GBP/USD traded at 00:07:41 BST, which represented a fall of almost 9% from the pre-event level of 1.26 (figure 18). Other platforms reported transactions at even lower prices³¹⁷.

During this period, it is likely that the trading activity of individual participants could have had a significant impact on market functioning and prices traded, given the lack of depth. Indeed, UK supervisory data point to a significant increase in certain market participants’ share of trading activity as others withdrew, suggesting a role for idiosyncratic factors in driving the extreme dysfunction observed as Sterling traded at levels well below 1.20 against the Dollar. Among such participants, we cannot exclude the possibility of an important role played by automated and high-frequency traders.

It is worth noting that some of the trades executed during this period were subsequently torn up or had their prices revised, but this was often the consequence of bilateral agreements or specific contractual arrangements as there is no single methodology for determining the low in FX markets. Market contacts suggested they did not feel able to co-ordinate in determining a low

³¹⁴ *The Sterling ‘flash event’ of 7 October 2016* (2017), Bank for International Settlements, 5-6

³¹⁵ *The Sterling ‘flash event’ of 7 October 2016* (2017), Bank for International Settlements, 8-9

³¹⁶ *Id*

³¹⁷ *Id*

price (which is relevant for certain derivatives contracts) as they feared this would breach competition and/or conduct requirements³¹⁸.

The minima reached in Sterling were extremely short-lived but the market dysfunction lasted for several minutes registering high levels of volatility and wide spreads.

Between 00:08:00 and 00:09:00 BST, GBP/USD traded on Reuters at levels between 1.20 and 1.22, still a wide range (figure 18). Market functioning took longer to recover. Measures of market depth on Reuters remained volatile until nearly 00:20:00 BST, a period over which the CME triggered a number of further trading halts (figure 22)³¹⁹.

Shortly after the first pause in trading on the CME ended, at 00:07:25 BST, the futures price hit the daily lower limit of 122.17³²⁰. At this point, futures trading was floored at this price for a period of two minutes. Some transactions were completed on the CME over this period, despite the spot price continuing to fall on other platforms. But when prices had not risen by the end of the two-minute window, a two-minute trading halt was imposed at 00:09:29 BST. The market reopened at 00:11:29 BST with a new lower limit (just over 120), but conditions remained impaired. Finally, amid the continued heightened price volatility, a second velocity logic event was triggered at 00:11:57 BST, again halting trading for 10 seconds on the futures exchange³²¹.

Stage 3: Over time the market began to recover, although it is difficult to identify a clear shift to the recovery phase. By around 00:20:00 BST, prices in both the futures and spot market had settled around 2% lower against the Dollar than their levels immediately prior to the event, although relatively high trading volumes persisted for a period of three to four hours. Bid-offer spreads on Reuters remained wider than usual, but not at extreme levels, for the rest of the night. Broader spillovers were generally limited. UK governments bonds registered a relatively large move as trading opened on 7 October, but the moves were orderly. And there appeared to be little impact on risky asset prices. GBP/USD closed in London on 7 October 1.4% lower than on the previous day³²².

³¹⁸ *Id*

³¹⁹ *Id*

³²⁰ Futures prices are not directly comparable to spot FX prices, but the two should move closely in line, absent large changes in relative interest rates or the cross-currency basis.

³²¹ *The Sterling 'flash event' of 7 October 2016* (2017), Bank for International Settlements, 8-9

³²² *Id*, 9-10

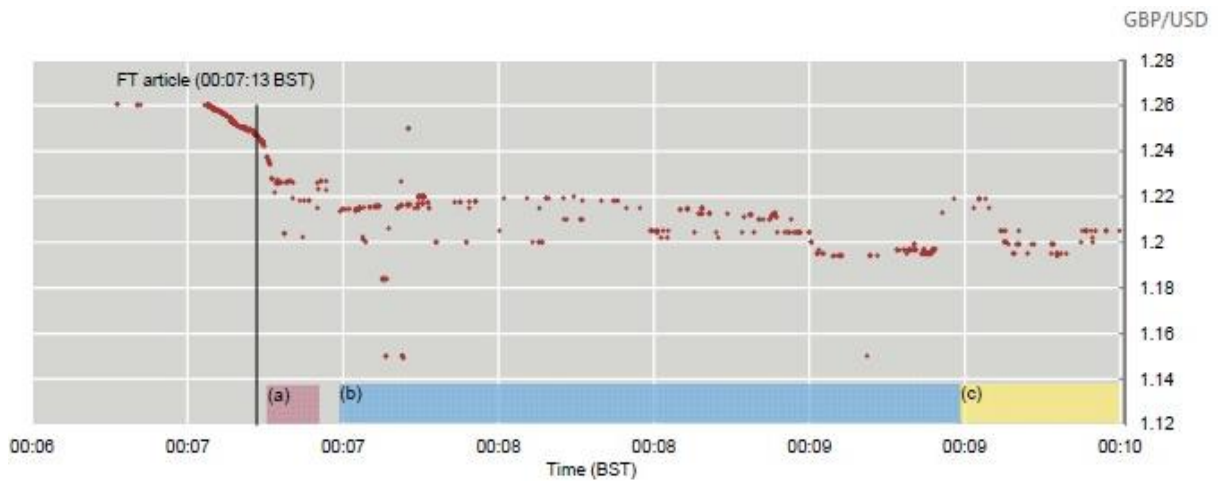


Figure 18: Thomson Reuters Matching GBP/USD trades. Source: Bank of England calculations.

- (a) 00:07:15 BST: sharp price movements over a two-second window trigger a velocity logic event which pauses trading on the CME for 10 seconds.
- (b) 00:07:29 BST: the futures price reaches its lower limit of 122.17 (based on the change on the day). The exchange remains open, but transactions cannot take place below this price on the CME. A two-minute monitoring period begins.
- (c) 00:09:29 BST: as the futures price has not rebounded from the lower limit by the end of the monitoring period, a further two-minute trading halt is triggered on the CME. At 00:11:29 BST, the exchange reopens with a new (lower) price limit.

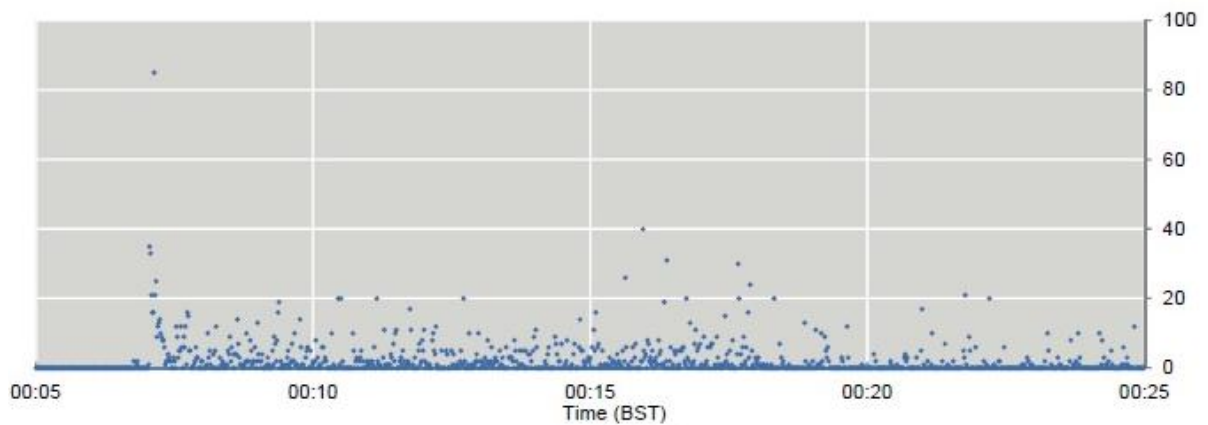


Figure 19: Thomson Reuters Matching GBP/USD trading volumes. Source: Bank of England calculations.



Figure 20: Thomson Reuters Matching GBP/USD sweep-to-fill costs. Source: Bank of England calculations.

Sweep-to-fill costs are calculated as the weighted average spread required to buy or sell a given quantity of £5 million. Gaps in the series represent periods where there is insufficient depth in the order book to complete a transaction of this quantity in the respective direction.

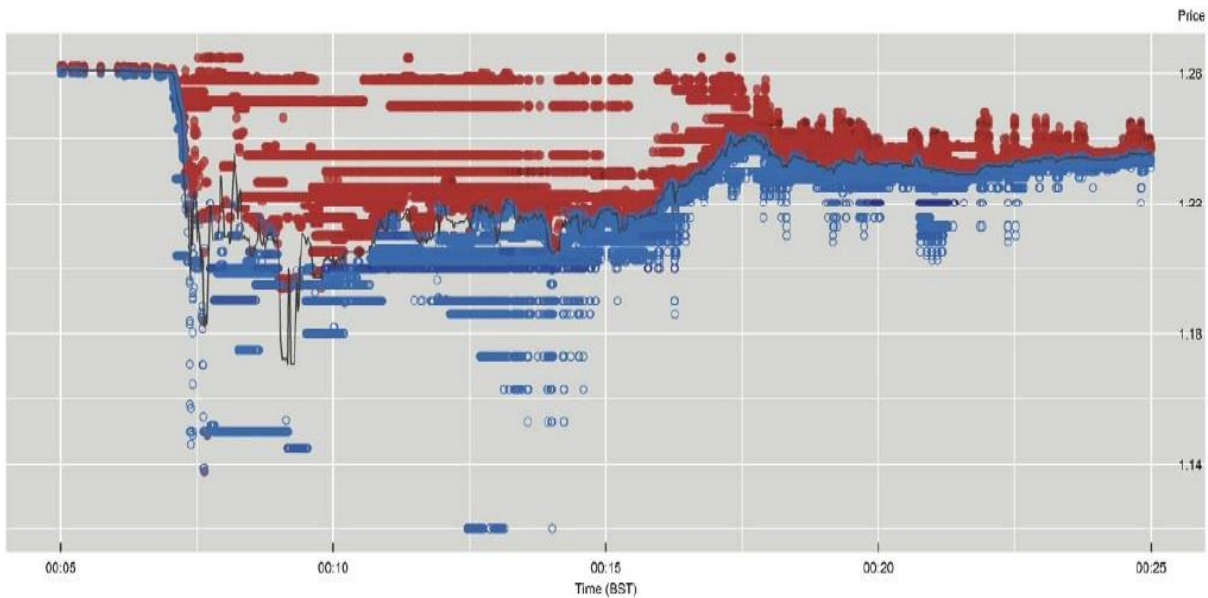


Figure 21: Thomson Reuters Matching GBP/USD order book behavior. Source: Bank of England calculations.

The blue and red circles represent resting bid and offer limit orders, respectively, in the order book. The black line represents the implied mid-price, and the intensity of the blue and red colors signifies the size of the order.

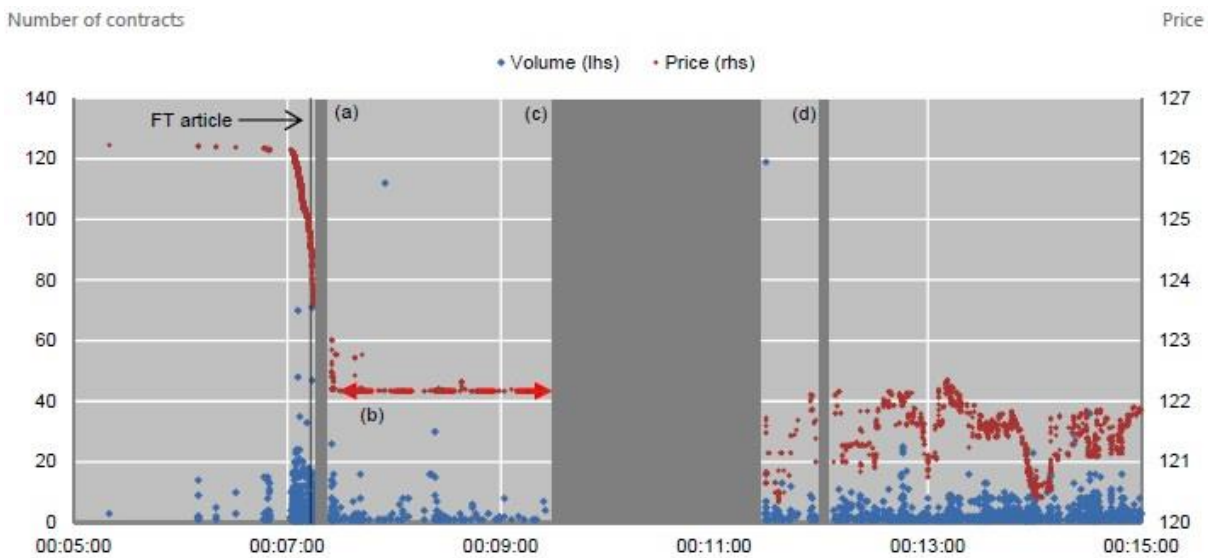


Figure 22: CME GBP/USD futures volume (blue) and traded prices (red). Source: Bank of England calculations.

- (a) 00:07:15 BST: sharp price movements over a two-second window trigger a velocity logic event which pauses trading on the CME for 10 seconds.
- (b) 00:07:29 BST: the futures price reaches its lower limit of 122.17 (based on the change on the day). The exchange remains open, but transactions cannot take place below this price on the CME. A two-minute monitoring period begins.
- (c) 00:09:29 BST: as the futures price has not rebounded from the lower limit by the end of the monitoring period, a further two-minute trading halt is triggered on the CME. At 00:11:29 BST, the exchange reopens with a new (lower) price limit.
- (d) 00:11:57 BST: a second velocity logic event is triggered by sharp movements over a two-second window, again pausing trading for 10 seconds.

7.4. Descriptive Analysis of the Pound Flash Crash

In this section, we perform a descriptive analysis of the Pound Flash Crash using GBP/USD minute-by-minute dataset obtained by *Thomson Reuters Eikon* data-provider. This data includes quotations from 09:00 am CET on Thursday 6 October 2016 to 04:30 pm CET on Friday 7

October 2016, for a total of 1891 observations. To illustrate and describe better the dimension of the Pound tumble, we mainly concentrate on low prices representing the minima values touched by the GBP/USD for every minute. Graphs and computations have been carried out through *Matlab* software.

The first step for the analysis is to calculate the price changes, or returns, in order to perform some descriptive statistics.

We compute the price changes of the GBP/USD quotations minute-by-minute according with the price returns standard formula:

$$R_t = \frac{P_t - P_{t-1}}{P_t} \%$$

where P_t is the exchange rate at time t and P_{t-1} is that at time t-1, the prior minute.

Since the dataset is composed by 1891 observations, the price changes are definitely 1890.

Figure 23 illustrates the returns trend for the GBP/USD quotations under examination. The day preceding the crash, the Pound values were reporting a moderate falling trend with the passing of the hours, but without registering abnormal spikes in volatility measures; conversely, shortly after 1 o'clock CET, the Sterling dramatically plunged down in value by losing about 9 percent in a sole minute in an unexpected way if compared to the previous quasi-flat situation. During the flash event window, the returns reported an explosion of volatility with shocking price changes both on the up and down side, that lasted for several minutes and only after a few hours the situation appeared to stabilize even though, price fluctuations raised again in the afternoon.

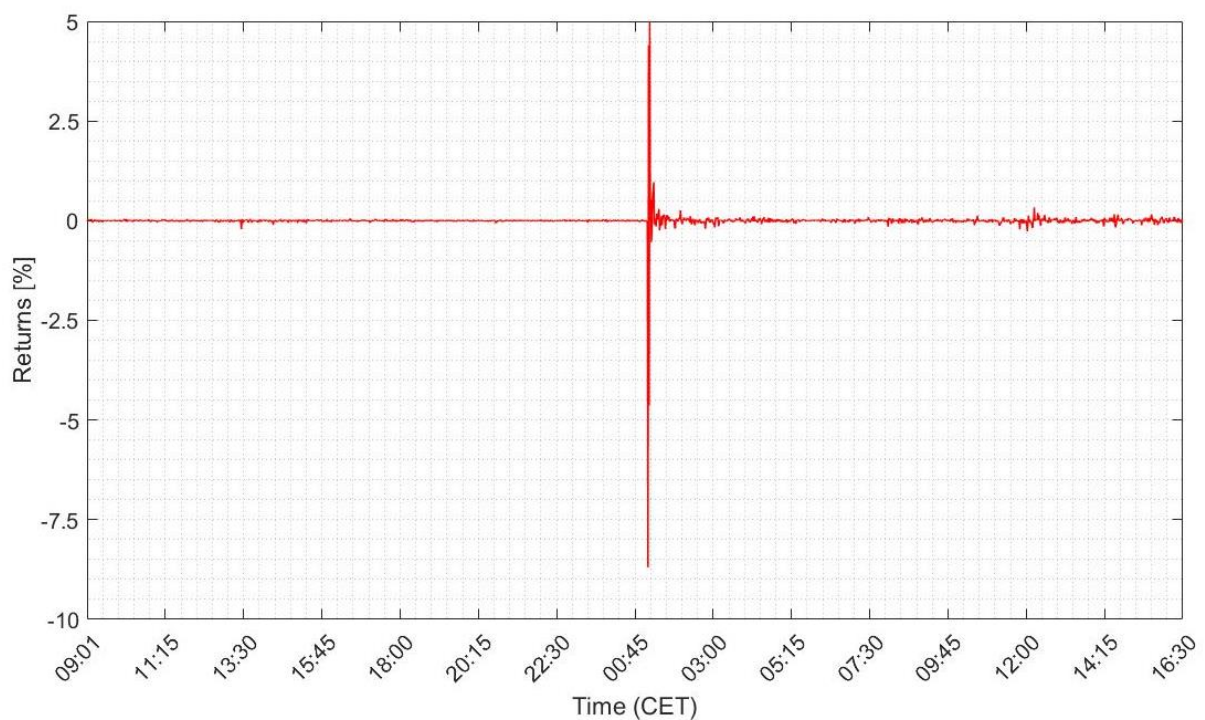


Figure 23: GBP/USD Flash Crash Price Returns. Source: author's elaboration on Thomson Reuters data.

In addition, Table 4 provides an excerpt of the prices and computed returns by using both closing and low prices. Specifically, it contains the flash crash values at 01:08 am (in bold type) and closer observations, respectively for the 5 minutes backwards and afterwards the breakdown. Numerically, the flash crash led to a price contraction of almost 9 percent in just one minute in terms of low quotations, and the three minutes that followed registered an averaged price oscillation of over 4 percent in absolute terms as the flash event triggered a very turmoil phase; while the near pre-crash situation reported a modest 0.0-point variation.

Similarly, closing prices indicate, with a lower magnitude with respect to low prices, similar conclusions: a falling in prices of over 3% at 01:08, and a subsequent high price volatility from here on.

Date	Time (CET)	Close Price	Return %	Low Price	Return %
2016-10-07	01:02	1.2609	-0.0238	1.2608	0.0000
2016-10-07	01:03	1.2608	-0.0079	1.2605	-0.0238
2016-10-07	01:04	1.2610	+0.0159	1.2606	+0.0079
2016-10-07	01:05	1.2605	-0.0397	1.2605	-0.0079
2016-10-07	01:06	1.2608	+0.0238	1.2604	-0.0079
2016-10-07	01:07	1.2602	-0.0476	1.2596	-0.0635
2016-10-07	01:08	1.2190	-3.2693	1.1500	-8.7012
2016-10-07	01:09	1.2126	-0.5250	1.2006	+4.4000
2016-10-07	01:10	1.2039	-0.7175	1.1450	-4.6310
2016-10-07	01:11	1.2108	+0.5731	1.2020	+4.9782
2016-10-07	01:12	1.2059	-0.4047	1.2049	+0.2413
2016-10-07	01:13	1.2131	+0.5971	1.2113	+0.5312

Table 4: Prices and Returns around the Pound Flash Crash. Source: author's elaboration on Thomson Reuters data.

Summary Statistics

Starting from price returns, we can conduct some descriptive statistics useful to summarize features of the intraday Pound Sterling flash crash.

We report some of the most used measures for time-series samples; specifically, measures of central tendency such as Mean, and measures of variability as the Standard Deviation, the Minimum and Maximum values, Skewness and Kurtosis.

Price returns can be averaged over time to obtain a sample lower-frequency return estimate which denotes the returns performance of the considered period. We use arithmetic mean:

$$Mean = E[R] = \frac{1}{T} \sum_{t=1}^T R_t$$

Variation in sequential returns is represented by the volatility measure, and the standard deviation. Standard deviation is simply defined as the square root of the average squared deviation of the returns from its mean:

$$\text{Standard deviation} = SD[R] = \sqrt{\frac{1}{T-1} \sum_{t=1}^T (R_t - E[R])^2}$$

Skewness indicator measures whether the return distribution skews towards either the positive or the negative side of the mean.

$$\text{Skewness} = S[R] = \frac{\frac{1}{T} \sum_{t=1}^T (R_t - E[R])^3}{\left(\sqrt{\frac{1}{T} \sum_{t=1}^T (R_t - E[R])^2} \right)^3}$$

Kurtosis is a measure of fatness of the tails of the return distribution. The fatter the tails, the higher the chance to have an extreme positive or negative return. The standardized normal distribution, for example, has a kurtosis of 3. A distribution with positive excess kurtosis is called leptokurtic.

$$\text{Kurtosis} = K[R] = \frac{\frac{1}{T} \sum_{t=1}^T (R_t - E[R])^4}{\left(\frac{1}{T} \sum_{t=1}^T (R_t - E[R])^2 \right)^2}$$

Table 5 contains the summary statistics applied to low prices quotations, respectively, for the entire sample period, for the pre-flash crash period and for the post-crash. This is useful to give a framework about the changing price returns caused by the flash event both in terms of tendency and variability.

Indicators confirm the trend and the observations made with the GBP/USD price returns. In fact, post-crash period was much more volatile compared to the pre-crash period, with a standard deviation 16 times higher. Accordingly, also the minimum and maximum returns are much higher, with a range of 9.6 in the post-period and of 0.26 in the pre-period.

Furthermore, mean and skewness are positive, confirming that fact that in the hours after the flash event, the Sterling was recovering.

Finally, values of kurtosis much higher with respect of 6 October indicated the high risk of the post-crash phase to obtain extreme observations, far away from the mean value.

Therefore, the GBP/USD minute-by-minute low price returns of 6-7 October 2016 exhibited an opposite behavior before and after the flash crash, inasmuch the event triggered a phase of extreme turbulence.

Sample	Obs	Mean	Std Dev	Min	Max	Range	Skewness	Kurtosis
Entire	1890	-0.00076331	0.27844	-8.7012	+4.9782	13.6794	-13.431	632.75
Pre-crash	967	-0.00094653	0.01708	-0.2128	+0.0475	0.2603	-2.5056	28.83
Post-crash	922	0.00886530	0.27646	-4.6310	+4.9782	9.6092	5.6039	269.19

Table 5: Descriptive Statistics on 6-7 October. Source: author's elaboration on Thomson Reuters data.

Correlation of returns

Another metric useful to describe distributions of returns is autocorrelation, which is a measure of serial dependence between returns and their lagged ones. For example, autocorrelation of order 1 is the correlation of 1-minute price return with 1-minute price return that occurred 1 minute earlier, as well as, the autocorrelation of order 2 is the correlation of 1-minute return with 1-minute return that occurred 2 minutes earlier. The autocorrelation value of order p can be determined as follows:

$$\rho(p) = \frac{\sum_{t=p+1}^T [(R_t - E[R])(R_{t-p} - E[R])]}{(\sum_{t=p+1}^T (R_t - E[R]))^{\frac{1}{2}} (\sum_{t=p+1}^T (R_{t-p} - E[R]))^{\frac{1}{2}}}$$

The autocorrelation function ranges from -1 to 1.

High correlation, say 0.5 and higher, implies a persistent positive relationship between current and lagged observations. Low autocorrelation, say -0.5 and lower, in turn implies a persistent negative relationship.

Autocorrelation allows us to check whether there are any persistent momentum/reversal relationships in the data. For example, it is a stylized fact that a large swing in the price is typically followed by a reversal, as the case of the Sterling flash crash.

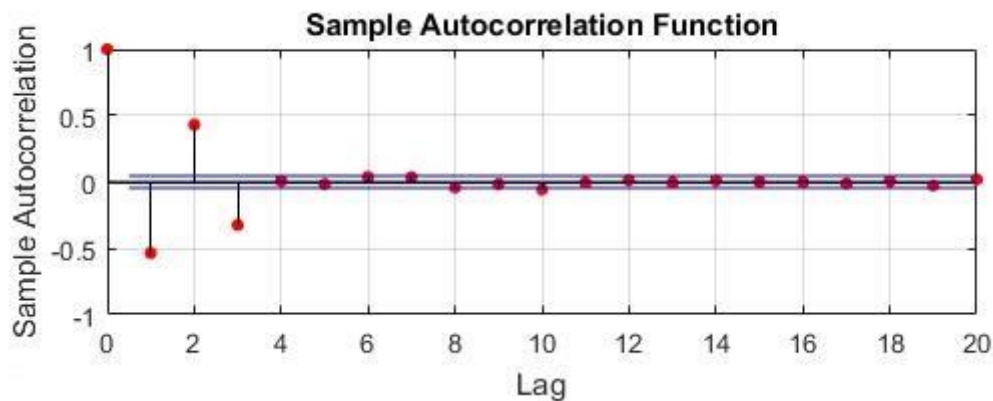


Figure 24: ACF of Returns on 6-7 October. Source: author's elaboration on Thomson Reuters data.

Figure 24 shows the sample autocorrelation for the low prices of 6-7 October. As you can note, the current return is negatively correlated with respect to the first 1-minute lag return, by supporting the idea that of a reversal behavior of the returns during the flash crash event, that is in presence of a negative performance at a specified minute, it will be highly probable to have a positive return in the following minute. The second lag is conversely positively correlated, while the third is negative again. This further strengthens the reversal relationships in returns observed during that period. The conditional mean is also confirmed by the Ljung-Box test, which strongly rejects the hypothesis of no-serial correlation at the relevant lags.

After controlling the presence of serial correlation, we also test whether the financial series of the GBP/USD presents volatility clustering and persistence, i.e. large price changes tend to be followed by large price changes, while small price changes tend to be followed by small price changes, as preceding considerations appeared to identify.

In order to test the presence of heteroskedasticity, we perform the Engle's ARCH Test which assesses the presence of autocorrelation on squared residuals of the price returns of GBP/USD. The test strongly rejects the hypothesis of no ARCH effect, then the returns are conditionally heteroskedastic.

7.5. Analysis of Liquidity during the Pound Flash Crash

In this section, we report an analysis of intraday liquidity of the Sterling Flash Crash carried out by the Bank of England staff.

Liquidity parameters help up to figure out better the market conditions during the 7 October compared to those on a normal day.

First, liquidity conditions typically vary over the course of the trading day. Though the FX market is open for 24 hours a day during the working week, the majority of trading in GBP/USD takes place between 7 am and 5 pm UK time, with volumes highest when both the London and NY markets are open. Accordingly, this is when measures of liquidity appear to be at their strongest.

Supporting this, Figure 25 shows a number of key measures of activity and liquidity averaged for each hour of the day, based on data from 3-6 October, the four days before the flash event. Despite occasionally large imbalances between buying and selling order flow and traded volume (left-hand panel) during 7-9 am and 1-5 pm BST, i.e. when the London and New York are open, this is when measures of liquidity (right-hand panel) are also the highest. Conversely, both trading volumes and liquidity measures are observed to drop off outside these core periods, particularly during the early hours of Asian trading, around midnight BST³²³.

³²³ *The Sterling 'flash event' of 7 October 2016* (2017), Bank for International Settlements, 4-5

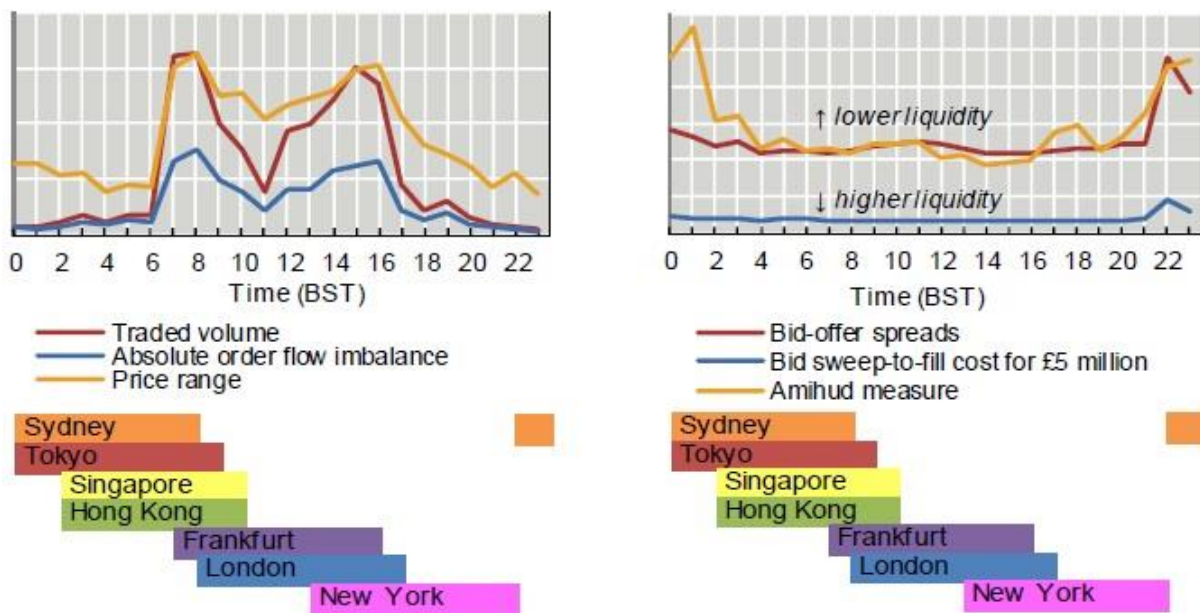


Figure 25: Intraday GBP/USD activity on Thomson Reuters Matching (3-6 October). Source: Bank of England calculations.

Sweep-to-fill costs are calculated as the weighted average spread required to buy or sell a given quantity of Sterling versus the Dollar.

Amihud measure calculates the price impact of individual trades as the ratio of the associated price move (measured since the last trade) to the traded volume.

A statistical analysis by Bank of England staff provides evidence to suggest that liquidity, as measured for example by bid-ask spreads, is more sensitive to changes in realized price volatility and traded volumes at night, compared with during the day. The left-hand panel in figure 26 shows the sensitivity of bid-ask spreads to realized volatility, as measured by the price range per minute, estimated on data from 10:00 pm BST on 2 October to midnight on 6 October. The data are divided into two subsamples covering “daytime” (01:00 am-10:00 pm BST) and “night-time” (10:00 pm-01:00 am BST). An increase in intra-minute volatility of 10 pips during the day is associated with a 0.3 pips higher bid-offer spread, but during the night this rises to 2.0 pips, even though the confidence interval is wider³²⁴.

The 7 October flash crash took place during the night, when liquidity conditions appear to be more sensitive to price volatility. This may have played a role in amplifying the effect of the initial shock. The right-hand panel in figure 26 compares the actual move in bid-offer spreads (red line) with the path predicted by the stylized model estimated using “daytime” data (yellow line) and “night-time” data (blue line). This comparison suggests that the deterioration in liquidity observed during the first minute of the event, from 00:07 to 00:08 BST, was broadly

³²⁴ The Sterling ‘flash event’ of 7 October 2016 (2017), Bank for International Settlements, Appendix B

consistent with what might be expected at that time of night given the large price move. But after that, the observed increase in bid-offer spreads is greater than what the model predicts³²⁵. Movements in bid-ask spreads over the flash event window are likely in practice to have reflected a more complex interplay between market participants' trading strategies, observable prices and order flows. Nevertheless, taking the model's estimates at face value, it is reasonable to conjecture that the initial shock might have had a significantly smaller impact had it occurred during the day, given the large difference between the blue and yellow lines³²⁶.

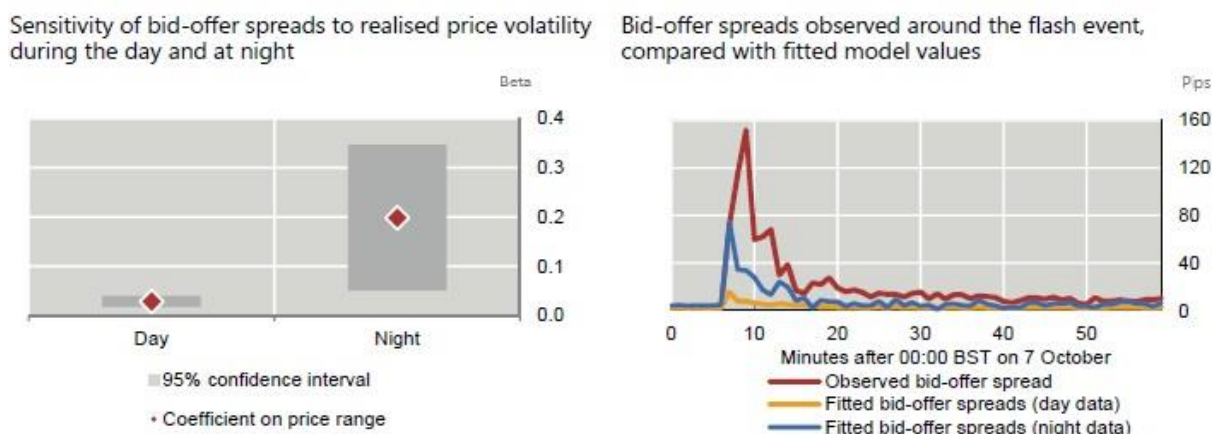


Figure 26: Estimating the drivers of market liquidity. Source: Bank of England calculations

Market liquidity is calculated using 3-6 October 2016 data from Thomson Reuters Matching and the following equation:
 $bid - ask\ spread_i = \alpha + \beta * price\ range_i + \gamma * traded\ volume_i + \mu$

Actually, on the night of 7 October, trading activity was significantly higher than usual. Volumes on Thomson Reuters Matching and CME were several hundred times their average overnight rate. Bid-offer spreads were significantly wider than their usual overnight average. And other measures of liquidity, such as sweep-to-fill costs³²⁷ and the Amihud³²⁸ measure, were also much higher than their normal overnight levels, indicative of lower liquidity. See figure 27 for a comparison of these and other liquidity metrics on Thomson Reuters with respect to the average levels prevailing in the 3-6 October, while figure 28 shows the Amihud measure for CME futures exchange in comparison with the previous week and the following weeks³²⁹.

³²⁵ *Id*

³²⁶ *Id*

³²⁷ Sweep-to-fill costs are calculated as the weighted average spread (from the implied mid-price) required to buy or sell a given quantity (£5 million) of Sterling versus the Dollar.

³²⁸ Amihud measure calculates the price impact of individual trades as the ratio of the associated price move (measured since the last trade) to the traded volume.

³²⁹ *The Sterling 'flash event' of 7 October 2016* (2017), Bank for International Settlements, *Appendix B*

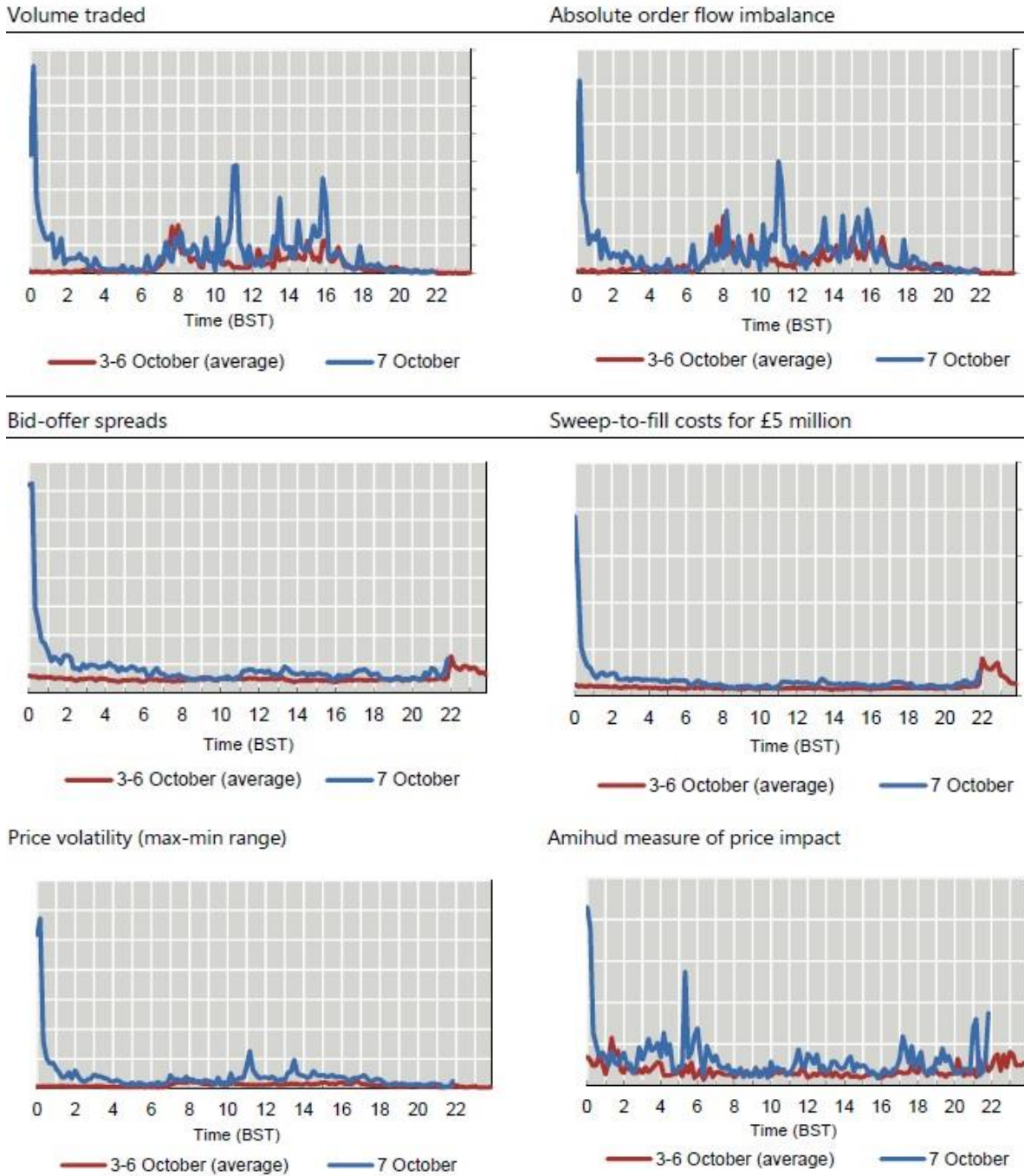


Figure 27: Thomson Reuters Matching trading activity and liquidity metrics. Source: Bank of England calculations

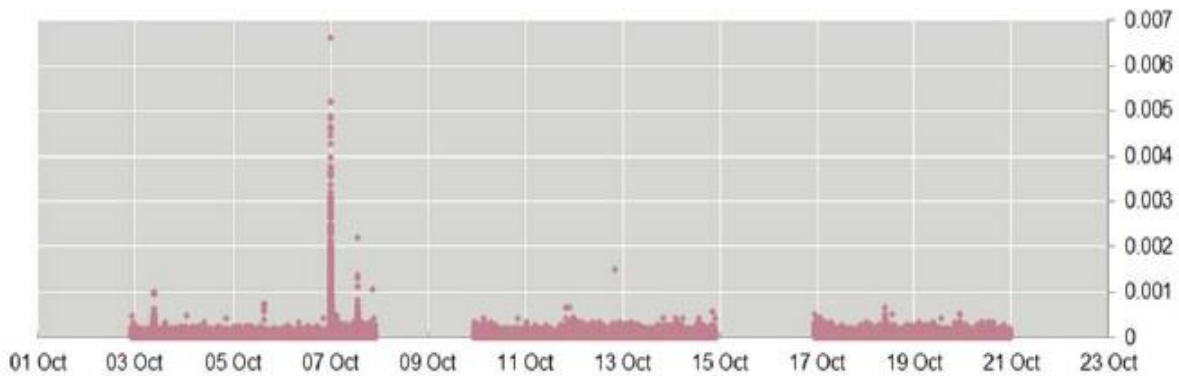


Figure 28: Liquidity in CME GBP/USD futures. Source: Bank of England calculations.

7.6. Triggers, Vulnerabilities and Amplifiers of the Pound Flash Crash

A number of idiosyncratic factors are likely to have contributed to and amplified the moves of 7 October, rather than to a single clean driver.

Firstly, it is important to reiterate that this event occurred during a typically illiquid period of the trading day for Sterling, i.e. outside the currency's core time zone, further impaired by regional bank holidays, including in China. Indeed, during this period, there is typically a shallower order book and a heightened sensitivity to increased volumes and/or volatility, which suggests that the market is more prone to register large price moves. The presence, in such time zone, of staff less experienced in trading sterling, with lower risk limits and risk appetite, and with less expertise in the suitability of particular algorithms for the prevailing market conditions, appears to have further amplified the movement³³⁰.

Additionally, market contacts have indicated several further candidate triggers of the flash crash.

Early commentary pointed to the publication of the Financial Times article on Hollande's considerations about Brexit, where he claimed that the UK would have to pay the price for have choosing a "hard" Brexit, as the driver of initial Sterling depreciation. However, it is likely that this release has only added marginal weight to the move as it did not contain new information because the comments from former French President were made at a widely attended event earlier that evening and had features in similar form on a variety of well-known news websites prior to the FT article.

Hence, other factors are more likely to have acted as trigger and amplifier of the flash crash. Potential candidates are for example, a so-called "fat-finger"³³¹ trade, a deliberate attempt to move the price lower during a typically illiquid period, or Asian retail trading in Sterling. Nevertheless, in each case market participants were unable to offer definitive evidence to substantiate these hypotheses, also for the lack of hard data.

As well as the time of day, contacts have identified a confluence of potential preexisting vulnerabilities going into the event that appear to have amplified both the price movement and the deterioration in market functioning. Chief among these are dealers' options-related hedging flows and client orders, including stop-loss orders. Both represent a source of mechanistic demand for liquidity in response to big changes in the level of the exchange rate³³².

³³⁰ *The Sterling 'flash event' of 7 October 2016* (2017), Bank for International Settlements, 11-13

³³¹ A fat finger error is a keyboard input error whereby an order to buy or sell is placed of far greater size than intended, for the wrong security, at the wrong price, or with any number of other input errors.

³³² *The Sterling 'flash event' of 7 October 2016* (2017), Bank for International Settlements, 11-13

Dealers usually seek to hedge options positions by buying or selling the exchange rate pair (the underlying) in order to maintain a neutral position with respect to small further price movements, but for large price movements this can involve buying or selling large quantities of it. And this hedging may be undertaken in an automated manner, without respect to prevailing liquidity conditions or the potential market impact³³³.

Similarly, retail traders can attach stop instruction to their orders to quickly sell or buy the exchange rate at certain price level in order to limit potential losses from large market fluctuations. Relatedly, retail FX clients can sometimes also be automatically stopped out by their brokers, mechanistically and without discretion, in the event of market losses causing them to no longer meet their margin requirements. Again, this has the potential to lead to trades being executed irrespective of the prevailing liquidity conditions and market impact³³⁴.

In particular, the positioning at a similar price threshold of pre-determined instructions for selling Sterling in order to hedge options positions, and the execution of stop-loss orders as the currency depreciated have been an important driver for the sharp plunge in GBP/USD.

UK supervisory data gathered from 12 of the most active dealers in the Sterling spot FX and options markets suggest that options hedging and, to a lesser extent, stop-loss orders may help to explain the significant order flow imbalance observed during the flash event. Figure 29 displays the potential cumulative impact of these mechanistic flows. However, these data do not capture retail brokers and therefore may underestimate the role played by stop-loss orders left by such traders or automatic stop-outs³³⁵.

Not all of this volume is likely to have been transacted given the level of dysfunction and the fact that options hedgers in particular may retain a degree of discretion. But, taken at face value, it represents a significant proportion of the total value traded during the event window and would constitute sizable volume even during the more active London trading hours³³⁶.

The initial price move and sharp increase in selling flow on 7 October aligns closely with a breach of the 1.26 level in GBP/USD, potentially indicative of a role for mechanistic selling needs around this level. And it is striking to note that the largest net selling pressure observed in these data coincides with a fall in GBP/USD through 1.24, that is the point at which there was a significant pickup in price gapping and broader dysfunction³³⁷.

³³³ *Id*

³³⁴ *Id*

³³⁵ *Id*

³³⁶ *Id*

³³⁷ *Id*

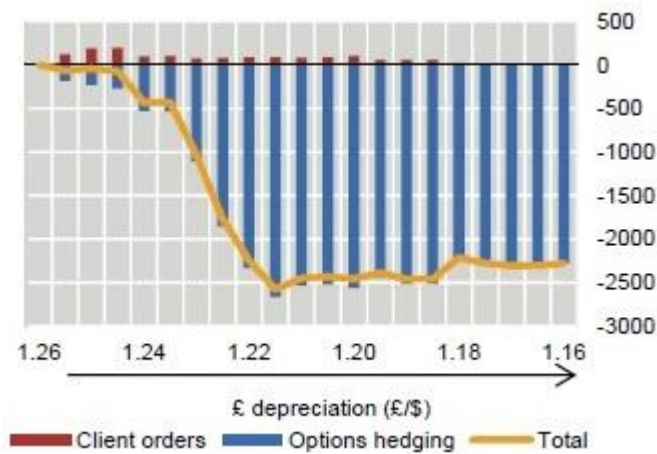


Figure 29: Cumulative net potential selling from options hedging and client orders. Source: Bank of England calculations.

Another important amplifier of the Sterling flash crash may have been the liquidity withdrawal of market-makers, who unlike equities market-makers have no obligation to provide liquidity in some circumstances.

Typically, liquidity providers become naturally more cautious in pricing risk during bouts of volatility and they wide prices before ceasing quoting altogether when certain thresholds are breached, such as P&L limits, or when wide, stale or off-market pricing inputs are identified.

Market makers exhibited a variety of responses to the increasing market dysfunction on 7 October³³⁸.

Some remained active throughout, with automated pricing widening for a time in recognition of the increased volatility, while others widened initially and then halted as protection mechanisms kicked in. Some firms withdrew first from voice trading, others from e-trading. Such halts lasted anywhere between two and thirty minutes. Some recommenced trading automatically when conditions stabilized, but many required management override. For some it was their algorithmic trading that restarted first (although often with human intervention to allow the restart), for others their voice activity. Those withdrawing liquidity cannot be readily categorized by type of institution, but various major dealers, principal trading firms and firms representing a retail client base confirmed that they withdrew liquidity provision from the market during the event³³⁹.

Generally, the withdrawal of liquidity in response to these factors was rapid, though, given the time of the day, not necessarily unexpected. Said that, the complete erosion of resting orders to buy Sterling for very short periods was highly unusual. And the fact that the futures exchange was halted for a large proportion of the event may have further amplified the withdrawal of

³³⁸ *The Sterling 'flash event' of 7 October 2016* (2017), Bank for International Settlements, 13-14

³³⁹ *Id*

liquidity in the spot market, given the extent to which automated market-makers rely on the CME as an additional pricing source and hedging tool³⁴⁰.

7.7. The Sterling Context prior the Flash Crash

Beyond idiosyncratic factors that hit the Sterling during the 7 October, it is important to provide context on broader moves preceding the flash event in order to understand in which background the sudden price collapse happened.

As a matter of fact, the British Sterling saw the 2016 as a year extremely turbulent with a persistent decline of its value with respect to other global currencies. Specifically, the Pound has been under pressure starting from the notorious June 23rd Brexit referendum where British people decided to leave the European Union, and from here on, the developments of international trade relationships and domestic political leadership became uncertain. This led to a substantial loss of confidence in the solidity of Sterling, by changing it from a relatively simple, cyclical currency to a very political currency.

Under this uncertainty scenario, the Friday June 24th, the Sterling quotation against the Dollar registered an extraordinary downfall of over 8 percent, passing from \$ 1.4878 to 1.3678, that is a loss of 1200 price interest points in a single day and touching an intraday minimum of \$ 1.32 (figure 30). This is also the time where volatility has reached extremely high values both in terms of implied and historical measures, with spikes well above the usual rates (figure 31 and 32).

In the following months, Pound oscillated just over \$ 1.30 until the end of September 2016 and variability of prices returned to be more moderate.

However, the week preceding the flash crash was characterized by another negative climate about UK stability.

Indeed, on 2nd October 2016, British Prime Minister Theresa May announced plans to trigger EU article 50³⁴¹ starting from Spring 2017, by confirming the intention of the United Kingdom's withdrawing from the European Union. This implied that, during the week prior to 7 October, Sterling had again trended lower against Dollar, with the GBP/USD value that decreased from a level of about \$ 1.30 at the beginning of the week to just over \$ 1.26 at the 6 October London close, a weekly deflationary pressure of about 3 percent but daily moves in Sterling during that week had been relatively orderly and measures of implied and historical volatility had ticked up only slightly from their post-EU referendum lows. Nonetheless, at that

³⁴⁰ *Id*

³⁴¹ Article 50 sets rules for the withdrawing of a European member, including economic negotiations and future relationship with the Union.

time, this is likely to have further increased the traders' loss of confidence in the currency because from the day of referendum up to the 6 October, Sterling had lost about 15% against Dollar.

Table 6 reports some summary statistics about turmoil periods preceding the flash crash.

Summarizing, the sharp intraday volatility hike observed on 7 October happened in an overall UK currency instability scenario as a result of the Brexit referendum, including news and decisions with regard to. This fact may have influenced a lot on behavior, trading activity, and responses that market participants adopted during that day, for instance in terms of risk hedging and stop-loss levels and in general, of countermeasures against the initial price decline.

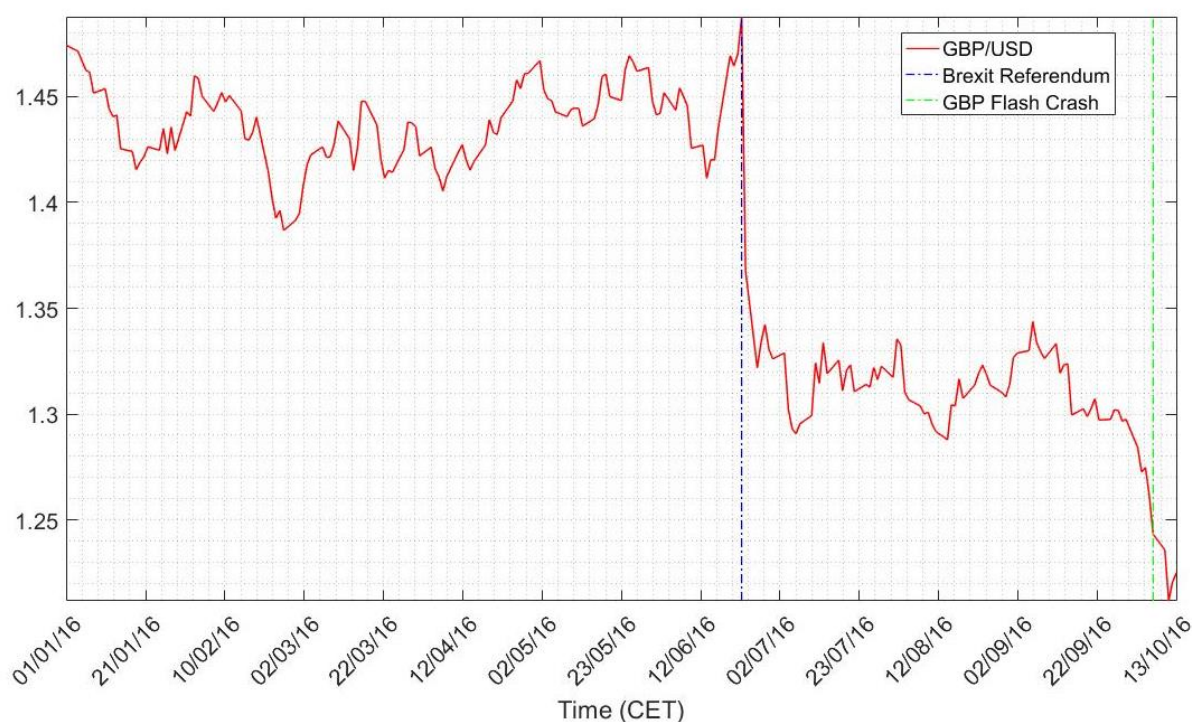


Figure 30: GBP/USD Daily Closing Prices before Flash Crash. Source: author's elaboration on Thomson Reuters data³⁴².

Sample	Obs	Mean	Std Dev	Min	Max	Range	Skewness	Kurtosis
1 Jan – 6 Oct	199	-0.07403	0.92051	-8.0656	2.25440	10.3200	-3.4013	30.794
24 Jun – 6 Oct	75	-0.21208	1.23530	-8.0656	1.91640	9.9820	-3.6045	23.415
3 Oct – 6 Oct	4	-0.70165	0.58713	-1.0511	0.17287	1.2240	1.1078	2.294

Table 6: Descriptive Statistics before Flash Crash. Source: author's elaboration on Thomson Reuters data.

³⁴² The dataset is obtained by *Thomson Reuters Eikon* data-provider and includes daily closing prices from 1st January 2016 through 13th October 2017 (excepting weekends) for a total of 205 observations.

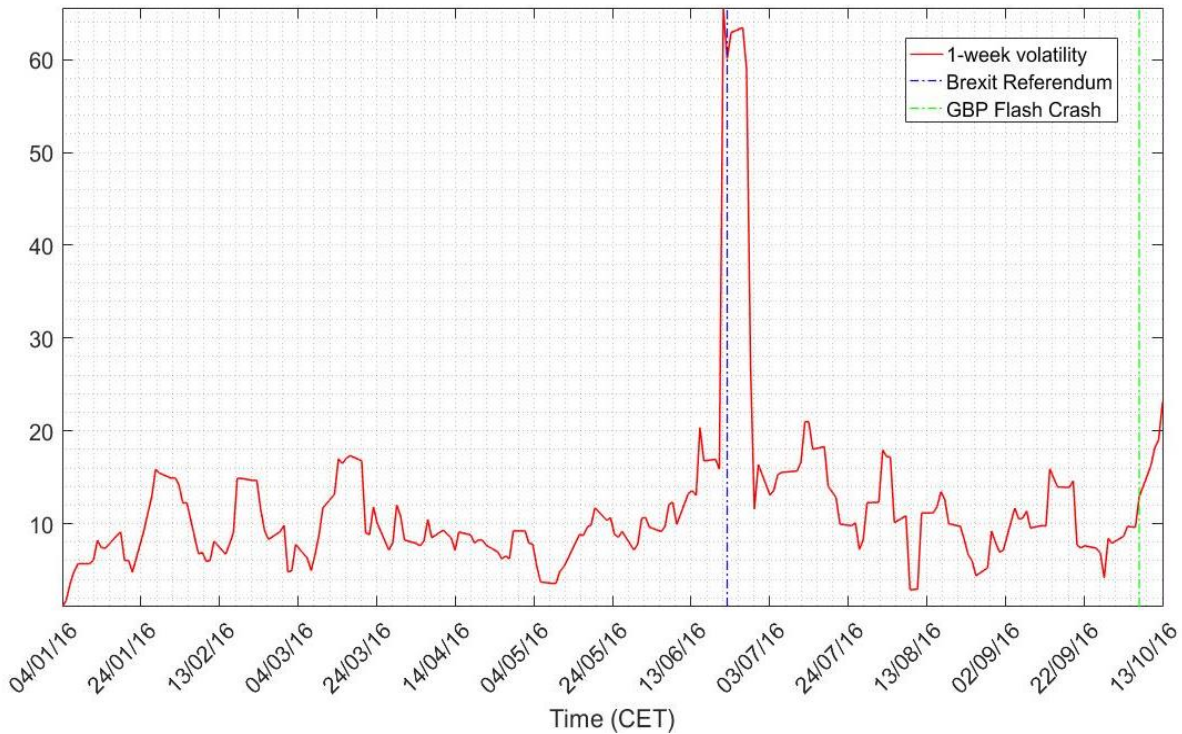


Figure 31: 1-week moving historical volatility before Flash Crash. Source: author’s elaboration on Thomson Reuters data.

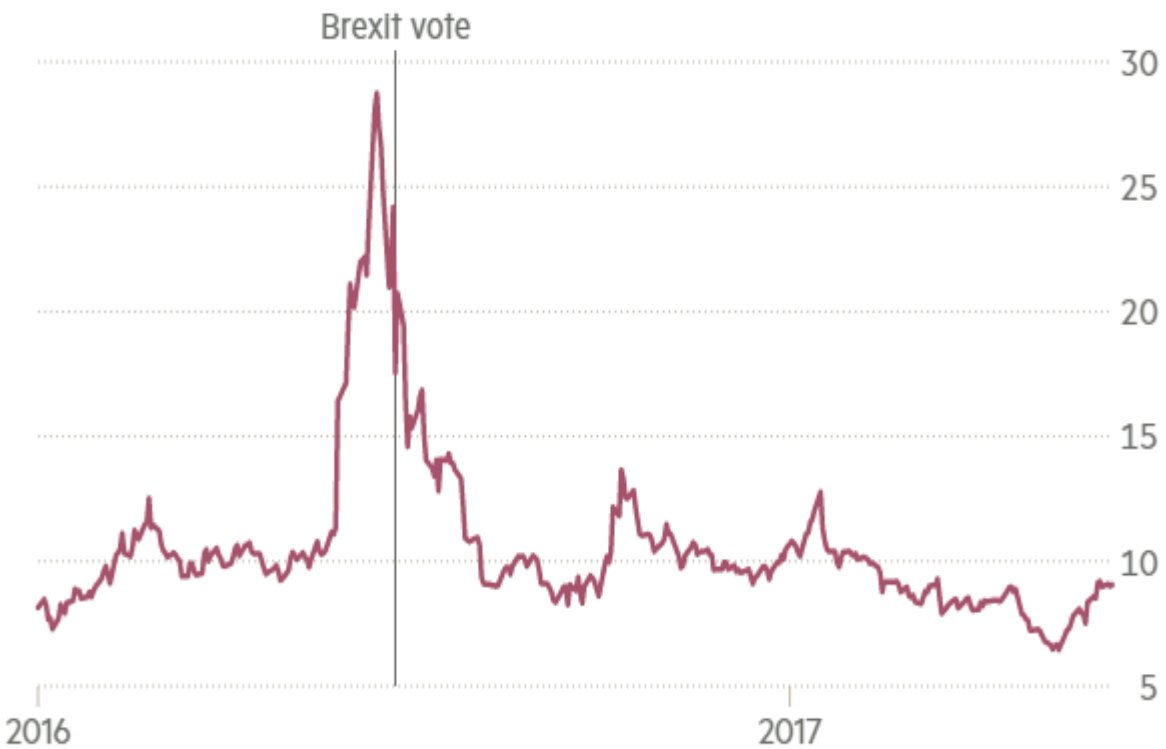


Figure 32: 1-month implied volatility. Source: Financial Times on Bloomberg data.

7.8. The Lasting Impact of the Pound Flash Crash

After reporting the potential triggers and the broad context in which the flash crash occurred, the fundamental target is to attempt identifying the potential market consequences of the event

on the future Sterling trading; in particular, we assess whether the volatility outburst observed on 7 October left some trace on the following weeks and months in terms of higher price turbulence and in general of higher transaction costs for market participants and whether other similar intraday price behaviors occurred.

Broadly, there were little observable evidences of a lasting impact on market functioning and pricing behavior because of the Sterling flash event.

Wide bid-offer spreads on 7 October appeared to quickly recover to around normal levels, with no evidence of impaired market functioning when trading recommenced on the evening of Sunday 9 October in Asian markets, and measures of the price impact of trading activity showed no signs of any persistent impact on liquidity in the futures market in the weeks that followed (figure 27)³⁴³.

But in the weeks following 7 October it was notable that a number of retail trading platforms increased their margin requirements for clients trading sterling FX, which may encourage prudent risk management but at the cost of higher transaction costs³⁴⁴.

About that, we observe a slightly increase of implied volatility measures in the close period following the event; 1-week option-implied volatility in Sterling FX pairs remained elevated for a number of days after 7 October before retracing, but it is difficult to determine whether or not this reflected factors other than the flash event (figure 33).

Implied volatility is a forward-looking measure that helps to gauge the sentiment and the expectation about the price changes of the exchange rate in the future; therefore, it can be defined as an indicator of expected exchange rate uncertainty.



Figure 33: 1-week option-implied volatility around 7 October. Source: Bank of England calculations.

³⁴³ *The Sterling 'flash event' of 7 October 2016* (2017), Bank for International Settlements, 10

³⁴⁴ *The Sterling 'flash event' of 7 October 2016* (2017), Bank for International Settlements, 16

Similarly to implied volatility measures computed by Bank of England, we perform the 1-week moving historical volatility on daily closing prices for the period ranging from 1st September 2016 through 15th May 2017 in order to check whether historical data has trended similarly.

The weekly moving historical volatility is computed for each day as the standard deviation over a sliding window of one week (5 days) across neighboring values and the window is centered about the element in the current position. Figure 34 points out that during the week following the crash, the standard deviation slightly increased with respect to previous values but this persistence lasted only for few days and however, in an orderly fashion if compared of that observed in the post-referendum window (figure 31).

Table 7 confirms a more turbulence in daily price returns during the week 10-14 October with respect to 3-6 October but this trend had no persistence in the following weeks.

Furthermore, in the following months there were no observations of similar intraday abrupt price fluctuations, reinforcing the hypothesis of an isolated event occurred during a particular trading context.

To be thorough, on the first month of 2017, we noted high volatility values but these are likely to be the effect of Theresa May's announcement about 12 negotiating objectives in the Brexit procedure, and specifically as a consequence on a potential UK withdrawal from the European single market, that has put further pressure on the British currency.

Therefore, we can conclude that the Pound flash crash probably did not leave particular extreme consequences on the price changes of the Sterling quotations, except for a more prudence on GBP trades and a slightly higher volatility in the subsequent week.

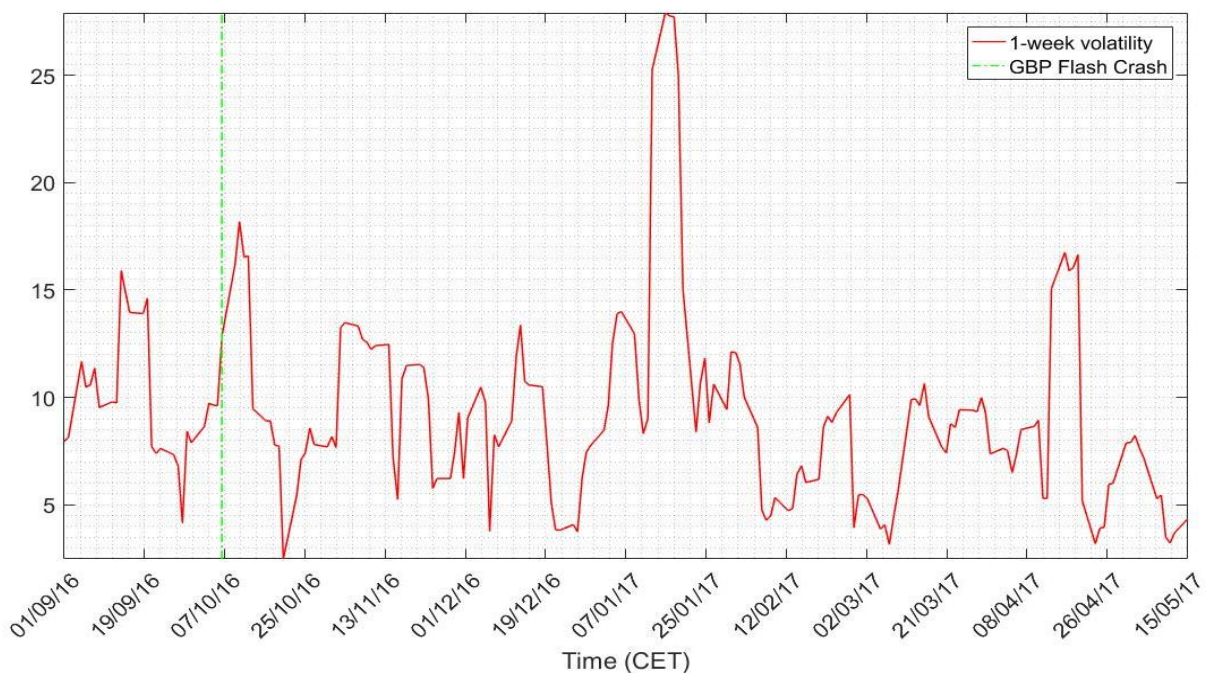


Figure 34: 1-week moving historical volatility after Flash Crash. Source: author's elaboration on Thomson Reuters data.

Sample	Obs	Mean	Std Dev	Min	Max	Range	Skewness	Kurtosis
3 Oct – 6 Oct	4	-0.70165	0.58713	-1.0511	0.17287	1.2240	1.1078	2.2936
10 Oct – 14 Oct	5	-0.39619	1.04140	-1.9502	0.73451	2.6847	-0.4628	2.0749
17 Oct – 21 Oct	5	+0.07961	0.48953	-0.2524	0.94409	1.1965	1.4043	3.1275
24 Oct – 28 Oct	5	-0.08593	0.46700	-0.6941	0.49237	1.1865	-0.1301	1.7213

Table 7: Descriptive Statistics before and after Flash Crash. Source: author's elaboration on Thomson Reuters data.

7.9. Conclusions about Pound Flash Crash

The Pound Flash Crash of 7 October, where the Sterling suddenly depreciated by around 9% versus the Dollar in early Asian trading before quickly retracing, is not a new phenomenon for financial markets; rather, it represents an additional empirical observation in what appears to be a series of mini-flash crash events occurring in a broad range of electronic markets, including equities, government bonds and FX, since the advent of high-frequency trading.

The dynamics of the Sterling event can be divided in three distinct phases. First, the early phase of the move, during which Pound decreased rapidly in response to significant selling flow but in an orderly fashion and with broad participation on key venues. Second, a period of few minutes of extreme dysfunction where Pound fell dramatically because of the evaporation of market liquidity and narrower participation, pointing to a greater role for the actions of individual market participants as a driver of the sharp moves. And finally, the gradual recovery in prices and market liquidity over the hours that followed.

Based on available evidence, this event appears to have been the product of a confluence of factors. Whatever the cause of the initial move, the market was likely to be vulnerable at that time of day to sharp moves and an associated withdrawal of liquidity. Among these, we recognize the occurrence of the event during a typically illiquid period of the trading day, that is outside the currency's core time zone with presence of staff less experienced in trading Sterling, the significant mechanistic demand to sell Sterling to hedge positions, the trigger of stop-loss orders as the currency traded through key levels, and the fact that trading halts were present in futures but not in spot markets. Additionally, a somewhat sterling-negative media report released shortly after the move began is only likely to have added marginal weight to the move.

Typically, mini-flash crash events have proven to be short-lived and without dramatic consequences for the financial stability, but if occurrences of heightened intraday volatility were to increase in frequency, or if disruption persisted for longer in future episodes (at the extreme this could involve a flash event that did not self-correct i.e. one which moved a market

to a new, non-fundamental, equilibrium), confidence in these markets could be undermined, potentially impacting financial stability³⁴⁵. In the FX market, this could manifest in a number of ways. Market-makers may demand additional compensation for liquidity provision, impairing market liquidity via wider bid-offer spreads and/or higher margin requirements. Or end-investors may show an increased reluctance to hedge due to concerns that their hedges will be crystallized not by the persistent changes in rates but rather by short lived market phenomena³⁴⁶. Were these dynamics to play out, there could be knock-on consequences also for the real economy since exchange rates are one of the most traded assets in the world both for speculative purposes but also for simple payments of goods and services.

Nevertheless, the flash event of 7 October provided limited evidence of persistent effects and market functioning appeared to have recovered more quickly with respect to similar episodes observed in recent years. In fact, volatility measures, both historical and implied ones, remained elevated only for a number of days; bid-ask spreads, appeared to quickly retrace to around normal levels when trading resumed, and in general measures of the price impact showed no signs of a persistent effect on liquidity in the weeks that followed; however, it was notable that a number of trading platforms increased their risk requirements with associated higher transaction costs for those who wanted to trade in Sterling in the weeks following 7 October.

Last but not least, there remain unresolved issues about what really happened during the specific time window of one minute where Sterling trading registered the peak of market quality deterioration and specifically, what was the behavior of market participants who continued to trade during that time.

As reported, during the second phase of the flash event, it is likely that the trading activity of individual participants have had a significant impact on market functioning and prices traded, given the lack of depth, and since the only traders that can trade large quantities in such a flash periods are definitely high-frequency traders, this may suggest that their activity is likely to have conditioned the prices behavior at that time.

Indeed, Sterling FX participants, regulators, and public attention are strongly concerned to the high frequency trading activity during that specific circumstance. About that, they ask: what was the role of algorithmic and high frequency trading on 7 October? Did they trigger out the Pound flash crash occurrence? Did their trading play a fundamental role on the sharp decline in prices? Did they leave the markets during the volatility burst? Did they adopt manipulative strategies that contributed the flash crash? Did they make or lose money at that time?

³⁴⁵ *The Sterling 'flash event' of 7 October 2016* (2017), Bank for International Settlements, 16

³⁴⁶ *Id*

These questions will probably remain unanswered if not in all at least in part, because in order to provide this evidence, it should be necessary to have ultra-high frequency data including all trades transacted during that period, an identifier of the traders for each trade and the possibility to distinguish high-frequency traders according to their specific trading characteristics.

Nevertheless, it can be consolidated that surely a high frequency trading component was fundamental on 7 October because the Sterling event showed some similarities with other flash crashes happened in recent years where high frequency trading demonstrated to have had a determinant role in the sharp price movements, either in the triggering or in the reinforcing of downward pressure.

CONCLUSIONS

High frequency trading is a very controversial trading practice that dominates the modern financial markets. It refers to proprietary traders that use complex algorithms and specific infrastructure to generate a very large number of trades at an extraordinary speed, with very short time-frames for establishing and liquidating positions, and high message intraday rates, including orders, quotes, or cancellations.

The use of co-location services and direct data feeds allows to minimize as much as possible the latency for generating, routing, and executing orders, which represents a significant competitive advantage with respect to other traders.

The quick development of HFT has implicated a substantial change on how trades are carried out with a heated discussion whether such phenomenon is beneficial or detrimental for market metrics, and investors.

Academics tried to study and analyze its influence on capital markets but the results show discordant conclusions. This is due by the fact that HFT represents a trading technique that potentially may be applied in a wide range of strategies, each of them have markedly different effects on market quality and investors, which entails a considerable difficulty on evaluating the phenomenon as a whole.

Nevertheless, pertinent literature, overall, claims that HFT can be considered as a significant provider of liquidity for the markets, in the form of increasing trading volume, tightening bid-ask spreads, and greater trading immediacy for participants. However, liquidity offered by HFTs is also dubbed as “phantom” liquidity as it might disappear at any moment since HFTs can modify and cancel their orders before execution in case of changing trading conditions.

Furthermore, HFT quotes play a large role in the price formation process by supporting the discovery and incorporation of information into the securities among trading venues, in particular in a highly fragmented market scenario where they are offered in different platforms at different prices. However, such process is assessed too fast and aggressive with imposition of heavy costs for slower traders that might decide to abandon the markets.

The behavior of HFTs strongly depend by the specific situations of the markets, in fact, they seem to act differently depending on market conditions are normal or turbulent. Indeed, during stressed periods in terms of price fluctuations and scarce liquidity, HFT activity has proven to play a significant role in exacerbating the turmoil of markets. This is demonstrated by observation of episodes of heightened intraday volatility, called flash crashes, where prices collapsed down and subsequent recovered in a time window of few minutes. The most notorious is that of May 6, 2010 that involved the major US securities indices, and where was provided

documentary evidence that HFTs led to a hike of volatility and an erosion of existing liquidity through an incessant trading activity.

Because of this event and followed by other mini-flash crash episodes and manipulative conducts, HFT was increasingly and heavily criticized for its potential risks and disruptive effects on market integrity, investors confidence and financial stability. This led over the years a stricter regulation about its activity, with more duties regarding information and specific requirements.

Accordingly, in this work, we have decided to analyze a recent episode of mini-flash crash occurred on Sterling trading, given that such events are considered and perceived as the result of the diffusion of HFT techniques on the capital markets.

Specifically, on 7 October 2016, the British Sterling depreciated by around 9% versus the Dollar in less than one minute, before quickly retracing much of the plunge.

Like many similar flash events, the Sterling moves seem to have no clear trigger, just unusually large selling flows at a typically quiet period of the trading day, where, theoretically, HFT activity might have acted as provoker of the initial price movement, but also for the lack of hard data and relevant information, the role of HFTs in this circumstance cannot be proven but neither ruled out a priori. We'll probably never know why the flash crash actually triggered out, if it was a fat finger, or some execution algorithms but there is no doubt that there was an electronic component to it considering the impact and velocity of the event. A hypothesis is that some algorithms have heavily reacted to a Sterling negative article by activating a cascading selling flow.

The British Pound, however, has been volatile since the Brexit referendum and during the flash day the Sterling market exhibited relevant idiosyncratic vulnerabilities and technical factors that contributed and amplified the market dysfunction during that time.

Despite the relevance of the price movements, this phenomenon did not have significant consequences for the markets and participants; there were few spillovers and no systemic financial institutions incurred material financial losses in this instance. The volatility remained high for short time periods following the event, and liquidity quickly returned to the market. This might point to market participants having learnt lessons from past episodes, and regulators having adopted a series of well-focused policies to contrast such events.

Nevertheless, these episodes, given their public attention, have the capability to undermine the investor confidence with significant consequences for the real economy.

FIGURES AND TABLES

Figure 1: Smart Order Routing - basic principle	18
Figure 2: HFT versus AT and traditional long-term investing	22
Figure 3: First Generation Algorithms	38
Figure 4: High Frequency Trading Strategies	40
Figure 5: Revenues sources for HFT liquidity providing strategies	42
Figure 6: Momentum Ignition Example	47
Figure 7: Latency differential	50
Figure 8: Flash Trading Example	51
Figure 9: Aspects of market liquidity	56
Figure 10: Cancellation/Execution Ratio	72
Figure 11: Illustration of Spoofing	73
Figure 12: Quote Stuffing situation	74
Figure 13: Equity Indices and Equity Index Futures, May 6, 2010	78
Figure 14: AP's False tweet	83
Figure 15: AP Tweet Flash Crash on April 23, 2013	83
Figure 16: GBP/USD Flash Crash	99
Figure 17: FX spot volume by electronic trading platform	101
Figure 18: Thomson Reuters Matching GBP/USD trades	105
Figure 19: Thomson Reuters Matching GBP/USD trading volumes	105
Figure 20: Thomson Reuters Matching GBP/USD sweep-to-fill costs	105
Figure 21: Thomson Reuters Matching GBP/USD order book behavior	106
Figure 22: CME GBP/USD futures volume and traded prices	106
Figure 23: GBP/USD Flash Crash Price Returns	107
Figure 24: ACF of Returns on 6-7 October	110
Figure 25: Intraday GBP/USD activity on Thomson Reuters Matching (3-6 October)	112
Figure 26: Estimating the drivers of market liquidity	113
Figure 27: Thomson Reuters Matching trading activity and liquidity metrics	114
Figure 28: Liquidity in CME GBP/USD futures	114
Figure 29: Cumulative net potential selling from options hedging and client orders	117
Figure 30: GBP/USD Daily Closing Prices before Flash Crash	119
Figure 31: 1-week moving historical volatility before Flash Crash	120
Figure 32: 1-month implied volatility	120
Figure 33: 1-week option-implied volatility around 7 October	121

Figure 34: 1-week moving historical volatility after Flash Crash	122
Table 1: Common characteristics of AT and HFT	23
Table 2: Exclusive characteristics of AT	25
Table 3: Specific characteristics of HFT	27
Table 4: Prices and Returns around the Pound Flash Crash	108
Table 5: Descriptive statistics on 6-7 October	109
Table 6: Descriptive Statistics before Flash Crash	119
Table 7: Descriptive Statistics before and after Flash Crash	123

REFERENCES

- AFM (2010), *High Frequency Trading: The Application of Advanced Trading Technology in the European Marketplace*, Consultation Paper
- Agarwal A. (2012), *High Frequency Trading: Evolution and the Future*, Capgemini
- Aldridge I. (2010), *High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems*, John Wiley & Sons, Inc.
- Almgren R. (2009), *Quantitative Challenges in Algorithmic Execution*, Working Paper
- Angel J., Harris L., Spatt C. S. (2010), *Equity Trading in the 21st Century*, Marshall Research Paper Series
- Arnuk S., Saluzzi J. (2008), *Toxic Equity Trading Order Flow on Wall Street: The Real Force Behind the Explosion in Volume and Volatility*, Themis Trading LLC White Paper
- Arnuk S., Saluzzi J. (2009), *Latency Arbitrage: The Real Power Behind Predatory High Frequency Trading*, Themis Trading LLC White Paper
- Bershova N., Rakhlin D. (2012), *High-Frequency Trading and Long-Term Investors: A View from Buy-Side*, Working Paper
- Biais B., Foucault T. (2014), *HFT and Market Quality*, Bankers, Markets & Investors n° 128
- Biais B., Woolley P. (2012), *The Flip Slide: High Frequency Trading*, Financial World
- BIS (2017), *The Sterling 'Flash Event' of 7 October 2016*, Markets Committee
- Black F. (1971), *Toward a Fully Automated Stock Exchange Part 1*, Financial Analysts Journal
- Bouveret A., Guillaumie C. (2014), *High-Frequency Trading Activity in EU Equity Markets*, ESMA Economic Report
- Brogaard J. (2010), *High Frequency Trading and its Impact on Market Quality*, Northwestern University
- Brogaard J., Hendershott T., Riordan R. (2013), *High Frequency Trading and Price Discovery*, ECB Working Paper Series No 1602
- Brown E. (2010), *Computerized Front-Running*, Counter Punch
- Caivano V. (2015), *The Impact of High-Frequency Trading on Volatility: Evidence from the Italian Market*, Quaderni di Finanza n° 80, Consob
- Caivano V., Ciccarelli S., Di Stefano G. (2012), *Il Trading ad Alta Frequenza: Caratteristiche, Effetti, Questioni di Policy*, CONSOB Discussion Papers No. 5

Cartea A., Penalva J. (2011), *Where is the Value in High Frequency Trading?*, Working Paper

CFTC and SEC (2010), *Findings Regarding the Market Events of May 6, 2010*, Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues

CFTC and SEC (2011), *Recommendations Regarding Regulatory Responses to the Market Events of May 6, 2010*, Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues

Chaboud A., Chiquoine B., Hjalmarsson E., Vega C. (2009), *Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market*, Board of Governors of the Federal Reserve System International Finance Discussion Papers n° 980

Cvitanic J., Kirilenko A. (2010), *High Frequency Traders and Asset Prices*, Working Paper

Duhigg C. (2006), *Artificial Intelligence Applied Heavily to Picking Stocks*, New York Times

Duhigg C. (2009), *Stock Traders Find Speed Pays in Milliseconds*, New York Times

ESMA (2015), *Regulatory Technical and Implementing Standards*, ESMA/2015/1464

European Parliament and Council of the European Union (2004), *Markets in Financial Instruments Directive*, Directive 2004/39/CE

European Parliament and Council of the European Union (2014), *Markets in Financial Instruments Directive II*, Directive 2014/65/EU

Foucault T., Biais B., Moinas S. (2011), *Equilibrium High Frequency Trading*, Working Paper

Friederich S., Payne R. (2011), *Computer Based Trading, Liquidity and Trading Costs*, Government Office for Science

Furbush D. (2010), *Program Trading*, Concise Encyclopedia of Economics

Golub A., Keane J., Poon S. H. (2012), *High Frequency Trading and Mini Flash Crashes*, Working Paper

Gomber P., Arndt B., Lutat T., Uhle T. (2011), *High-Frequency Trading*, Goethe Universitat Frankfurt Am Main

Greco A. (2011), *Giro di Vite Consob sul Trading ad Alta Frequenza*, la Repubblica, 4 November

Gresse C. (2012), *Market Fragmentation in Europe: Assessment and Prospects for Market Quality*, Driver Review DR19, Government Office for Science

Hagstromer B., Norden L. (2013), *The Diversity of High Frequency Traders*, Journal of Financial Markets, 16(4)

- Harris J. H., Schultz P. H. (1998), *The Trading Profits of SOES Bandits*, Journal of Financial Economics Vol. 50 No. 2 Oct.
- Harris L. (2003), *Trading and Exchanges: Market microstructure for practitioners*, Oxford University Press
- Hasbrouck J., Saar G. (2010), *Low-Latency Trading*, Working Paper
- Hendershott T., Jones C. M., Menkveld A. J. (2011), *Does Algorithmic Trading Improve Liquidity?*, The Journal of Finance Vol. Lxvi, No. 1 February 2011
- Hendershott T., Riordan R. (2011), *Algorithmic Trading and Information*, Working Paper
- Hoffmann P. (2011), *A Dynamic Limit Order Market with Fast and Slow Traders*, Working Paper
- Iati R. (2009), *The Real Story of Trading Software Espionage*, The TABB Group
- IOSCO (2011), *Regulatory Issues Raised by the Impact of Technological Changes on Market Integrity and Efficiency*, Technical Committee of the IOSCO
- Jacob Leal S., Napoletano M. (2016), *Market Stability vs. Market Resilience: Regulatory Policies Experiments in an Agent Based Model with Low- and High-Frequency Trading*, Working Paper
- Jain P., Jain P., McNish T. H. (2016), *Does High Frequency Trading Increase Systemic Risk?*, Working Paper
- Jarnecic E., Snape M. (2010), *An Analysis of Trades by High Frequency Participants on the London Stock Exchange*, Working Paper
- Johnson B. (2010), *Algorithmic Trading and DMA: An Introduction to Direct Access Trading Strategies*, 4Myeloma Press, London
- Kirilenko A., Samadi M., Kyle A. S., Tuzun T. (2011), *The Flash Crash: The Impact of High Frequency Trading on an Electronic Market*, Working Paper
- Kumar P., Goldstein M., Graves F., Borucki L. (2011), *Trading at the Speed of Light: The Impact of High-Frequency Trading on Market Performance, Regulatory Oversight, and Securities Litigation*, The Battle Group Issue 02
- Lewis M. (2014), *Flash Boys*, W. W. Norton & Company
- Liebenberg L. (2002), *The Electronic Financial Markets of the Future, and Survival Strategies of the Broker-Dealers*, Palgrave MacMillan

Markham J. W., Harty D. J. (2008), *For Whom the Bell Tolls: The Demise of Exchange Trading Floors and the Growth of ECNs*, 33 Iowa J. Corp. L.

McGowan M. J. (2010), *The Rise of Computerized High Frequency Trading: Use and Controversy*, Duke Law & Technology Review No. 16

Miller R. S., Shorter G. (2016), *High Frequency Trading: Overview of Recent Developments*, Congressional Research Service

Moosa I. (2015), *The Regulation of High-Frequency Trading: A Pragmatic View*, Journal of Banking Regulation

Moyer L., Lambert E. (2009), *Wall Street's New Masters*, Forbes Sept. 21

Patterson S. (2013), *Dark Pools: The Rise of A. I. Trading Machines and the Looming Threat to Wall Street*, RH Business Books

Prewitt M. (2012), *High-Frequency Trading: Should Regulators Do More*, Michigan Telecommunications and Technology Law Review 19

Puorro A. (2013), *High Frequency Trading: una panoramica*, Questioni di Economia e Finanza n° 198, Banca d'Italia

Ruhl T. R., Stein M. (2014), *The Impact of Financial Transaction Taxes: Evidence from Italy*, Economics Bulletin Volume 34, Issue 1

Sandler A. J. (2011), *The Invisible Power of Machines: Revisiting the Proposed Flash Order Ban in the Wake of The Flash Crash*, Duke Law & Technology Review No. 003

SEC (1999), *Regulation ATS*, Release No. 34-40760

SEC (2005), *Regulation NMS*, Release No. 34-51808

SEC (2010), *Concept Release on Equity Market Structure*, Release No. 34-61358

SEC (2010), *SEC Approves New Rules Prohibiting Market Maker Stub Quotes*, Press Release

SEC (2011), *SEC to Publish for Public Comment Updated Market-Wide Circuit Breaker Proposals to Address Extraordinary Market Volatility*, Press Release

SEC (2012), *SEC Approves Proposals to Address Extraordinary Volatility in Individual Stocks and Broader Stock Market*, Press Release

SEC (2013), *Regulation SCI*,

SEC (2014), *Equity Market Structure Literature Review. Part II: High Frequency Trading*, Consultation Report

Shorter G., Miller R. S. (2014), *High-Frequency Trading: Background, Concerns, and Regulatory Developments*, Congressional Research Service

U.S. Department of the Treasury-Board of Governors of the Federal Reserve System-Federal Reserve Bank of New York-SEC-CFTC, *The U.S. Treasury Market on October 15, 2014*

Wah E., Wellman M. P. (2013), *Latency Arbitrage, Market Fragmentation, and Efficiency: A Two-Market Model*, University of Michigan

Zhang S. (2013), *Need for Speed: An Empirical Analysis of Hard and Soft Information in a High Frequency World*, Working Paper

Zhang X. F. (2010), *High Frequency Trading, Stock Volatility and Price Discovery*, Working Paper

Websites

Bloomberg: www.bloomberg.com

Chicago Mercantile Exchange: www.cmegroup.com

Electronic Broking Services: www.ebs.com

European Securities and Market Authority: www.esma.europa.eu

European Union Law: www.eur-lex.europa.eu

Financial Times: www.ft.com

International Organization of Securities Commissions: www.iosco.org

Nanex: www.nanex.net

Nasdaq: www.nasdaqtrader.com

New York Stock Exchange: www.nyse.com

Tabb Forum: www.tabbforum.com

The Wall Street Journal: www.wsj.com

Thomson Reuters: www.thomsonreuters.com

U.S. Securities and Exchange Commission: www.sec.gov

Zero Hedge: www.zerohedge.com