Università
Ca'Foscari
Venezia

Master's degree

in

Economics and Finance

Final Thesis

High Frequency Trading and Financial Stability

How algorithmic revolution have reshaped financial markets

Supervisor

Ch. Prof. Monica Billio

Graduand

Aurelio Nocera

Matricolation number

870722

Academic Year

2018 / 2019

Abstract

Financial markets have undergone through an epochal revolution over the last four decades. As always, technology and regulation have played a fundamental role in shaping market landscape. Within this new framework, high frequency traders have emerged as fundamental characters into the new financial scenario.

They employ cutting-edge technology to identify trading opportunities into the markets and profit from them. These players do not work on a human-conceivable time scale: the notions of millisecond, nanosecond and microsecond have indeed become popular in finance only after they have conquered large market shares. Because of their operations, which span from cross-market arbitrage opportunities and event trading to market making activity, High Frequency Traders have significantly influenced market dynamics, accelerating some of them and also introducing completely new ones.

Understanding how High Frequency Traders play their game is then of fundamental importance to understand how financial markets work today.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

Hereby, I would like to express my gratitude to some people who have supervised my student path, have always been by my side or have become part of my life during the last years contributing to my personal growth.

I would like to say thank you to Professor Monica Billio for giving me the opportunity to work with her, supervising my research and for all meaningful conversations and precious advice.

To my parents Antonio and Antonietta and my sister Sara, for having always supported me during my life, having provided the opportunity to study abroad, and having always treated me with their amazing love.

To my venetian family from Ca' Foscari International College, where I spent two wonderful years of my life. In particular, I want to mention Irene, but also Andrea, Federico, Stefania, Lorenzo, Gabriele, Laeticia, Lucia, Clara, for their loving presence, precious and sincere friendships and help during these years.

To my old dear friends Alessandro, Dario, Giorgio, Alberto, Simone, Gianluca, Damiano, Leila, Adriana, Andrea, Caterina and Giacomo who know me from long time and are still by me side.

To Tulane University in New Orleans, where I spent an exchange semester which completely changed my life. To my Hong Kong friends, Andrea, Nicolo', Riccardo and Logan who made me feel to be at home despite distance.

Without you all, nothing of what I have achieved up to today would have been possible. With you all, I hope to achieve even more and become a better person.

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Following the Flash Crash of May 2010, High Frequency Trading (HFT) has suddenly become one of the most debated topics in the financial industry. Up to that moment, however, HFT was barely an unknown topic even for professional traders and financial experts. Nevertheless, HFT did not appear from nowhere and did not grow into a vacuum. Several changes into markets' infrastructure, regulatory interventions and technological advancements have contributed to its birth, growth and diffusion across the market and different asset classes. We are going to explore the most important changes which have made HFT possible and explain their impact on the market structure.

CHAPTER 1: AN EPOCHAL REVOLUTION

1.1 QUANT, ALGORITHMIC AND HIGH FREQUENCY TRADING

During the 90's of 20th century, Finance has undergone through a cultural revolution. Flock of mathematicians, statisticians and physicists were hired on Wall Street and started populating financial firms' trading desks. These analysts, who became known as "*quants*", look at financial markets through quantitative lens borrowed from their original natural sciences' approaches. They developed and implemented complex mathematical models in order to find and exploit quantitative patterns into market data, originating what became known as "**quant trading**". Quant trading can be defined as "*a mathematical model-fueled trading methodology that represents a radical departure from established technical and fundamental trading styles*" (Aldridge, 2009, pag.15). In summary, quantitative trading looks at the market for identifying any profitable price discrepancies, with the aim of ruling them out through the implementation of statistical arbitrage strategies. To succeed and profit from market inefficiencies, quants rely heavily on algorithms which are useful to encode in computer language their trading strategies.

According to Banks (2014, pag.134), **algorithmic trading** is "*any form of trading that makes use of computerized process to determine when, where, and how to execute*

financial transactions in the marketplace". Buy and sell decisions are made in advance by a portfolio manager who has a specific investing or trading horizon. Therefore, algorithms take only care of *optimizing the execution process* of decisions already taken by humans in the first place. Algorithms are frequently designed and programmed to replicate the most common execution strategies of human traders. They submit and monitor these orders and can also adjust and readjust them according to the constantly changing conditions that they observe in the market. For this reason, considerations about routing process, timing, size and price details all play a crucial role in determining success: minimizing price disruption effect. [1]

At the beginning, these lines of codes were developed on proprietary desks of major banks, which specifically created new working division for coding purposes. These products were then sold to banks' buy-side clients base. Eventually, even agency brokers and software developers joined the group of algorithms' developers. Today we have reached a market configuration such that broadly *all class of* professional market participants use algorithmic trading.

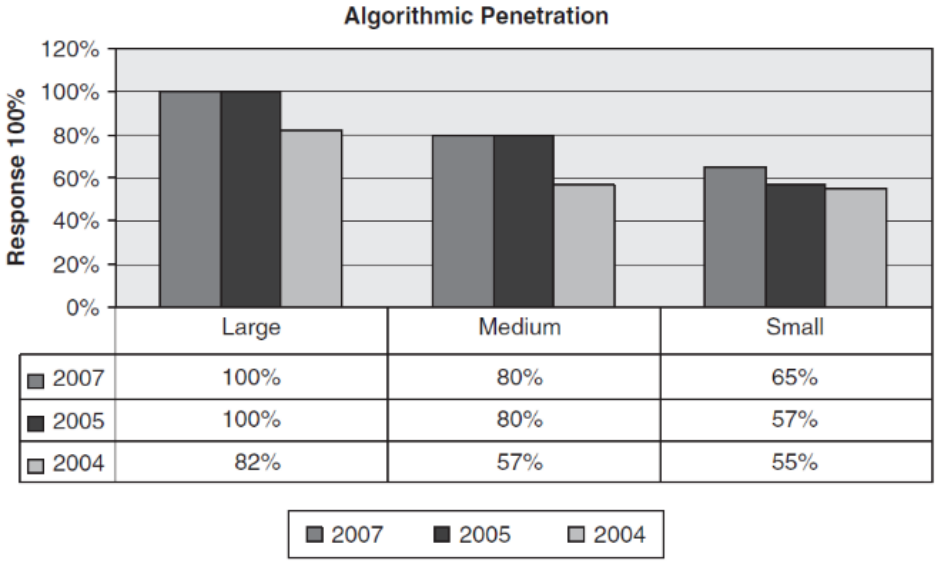

Figure 1- Algorithmic penetration over 2004-2007.

Source: Kim (2007, pag. 52), Original Study: TABB Group, June 2005.

[1] Algorithmic trading strategies can become predictable and display patterns. We will address this topic in Chapter 3 while we will talk about pinging and some predatory strategies implemented by HFT traders.

Algorithms exhibit several advantages: they do not feel emotions, do not hesitate and are perfectly accurate in carrying out their tasks. They never get tired, distracted or overexcited. They do not even show personal behavioral biases[2], do not dispose of free will and then are not tempted to break the rules. They can process more data in less time than a human can do and react faster to signals ensuring lower costs of execution and orders' monitoring. Finally, they can perform all these tasks in an automated way without no human intervention beyond programming phase.

Automation by itself does not have neither positive nor negative effects on the markets. Abergel et al. (2012) explain that algorithms' effect on market quality are likely to depend on the nature of the trading strategies coded by algorithms. According to Hasbrouck and Saar (2010), we can identify two main types of algorithm:

- Agency algorithms (AA)
- Proprietary algorithms (PA)

The buy – side[3] often need to rebalance a portfolio's structure following market events. To pursue this goal while minimizing transaction costs, they use agency algorithms. These codes adhere to the algorithmic trading's description previously provided: they optimize in real-time execution of trading decisions. Given this purpose, they are also often used by brokers who need to execute orders for their clients. The optimal trading strategy depends on market conditions and managers' horizon. The software and algorithms employed for pursuing this goal are build, test and run by Quants.

Proprietary Algorithms are used mainly for two types of activities: (1) electronic market-making and (2) arbitrage or statistical arbitrage trading.[4] These algorithms are used by banks' proprietary trading desks, hedge funds, proprietary trading firms or individual traders. Their success frequently depends on speed, since being able

[2] However, they can reflect the one of the people that have programmed these lines of code.
[3] Buy-side: institutions that trade for investing and asset management. Pension funds, mutual funds and money managers belong to this category.
[4] More on both activities in Chapter 2 – HFT Trading Strategies.

to quickly react to market changes is of paramount importance to capture profit opportunities. In fact, "*usually the first mover gets the best price*" (IBM,2008, pag.3).

Sornette and Von Der Becke's (2011) define **High Frequency Trading** (HFT) as "*the ultra–high- speed version of algorithmic trading"*. Speed is obviously a crucial aspect for High Frequency Trading. As explained by Durbin (2010, pag. v) "*HFT refers to the buying or selling of securities wherein success depends on how quickly you act, where a delay of a few thousandths of a second, or milliseconds, can mean the difference between profit and loss"* [5].

With the support of a summary picture realized by Aldridge (2009), we can formalize the relationship that exists among HFT, Algorithmic trading and traditional long-term investing.

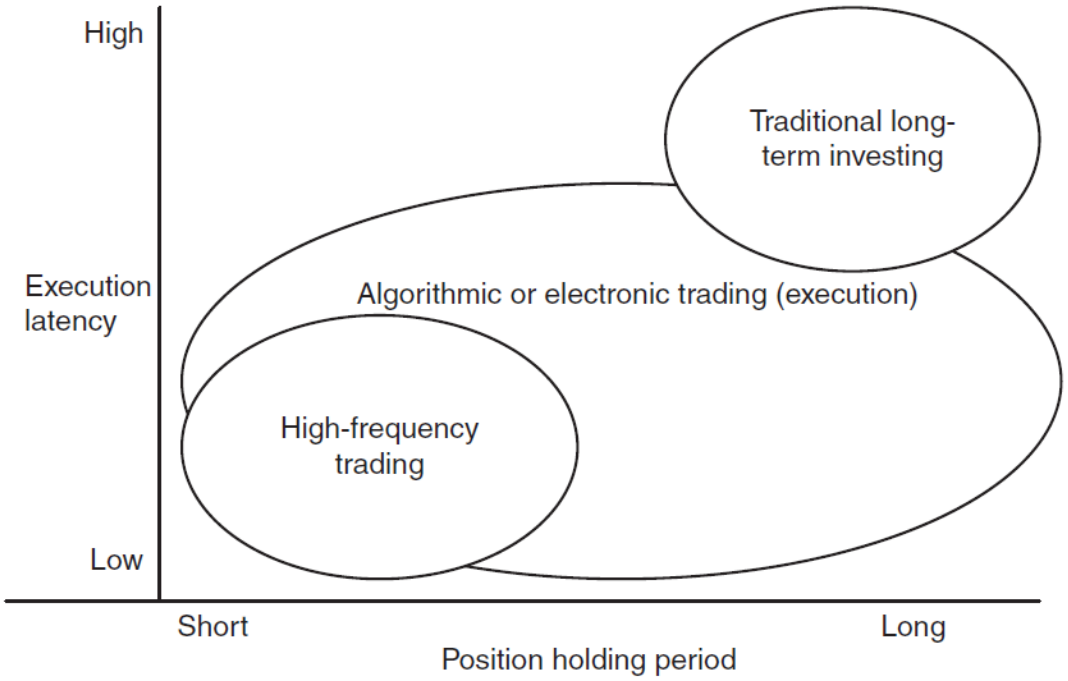

Figure 2 – Relationship among HFT, Algorithmic Trading and Long-term investing. Source: Aldridge (2009, pag 17).

HFT has not always been a natural component of the trading landscape. In the summer of 2009, according to Lewis (2014), the word "*high frequency trading" was*

[5] In the time that a trader presses a button on his keyboard to submit an order, several hundred transactions realized by algorithms which work at high frequency time span can potentially take place.

for most people, even on Wall Street, completely unknown". The lack of understanding of what was going on in the market back there triggered a race to hire people who were capable of disentangling and understanding HFT. According to Aldridge (2009, pag.1), during the worst months of the 2008 crisis, 50 percent of all open positions in finance involved expertise in high-frequency trading. Given this backdrop, it should not impress that High Frequency Trading has quickly became one of the most discussed topics. Not only on the Street, but also in the media. Durbin (2010, pag. v) in his book's introduction approached HFT's topic raising the following question ""*What is high-frequency trading? Great question*! *And it's about time for an answer, because everyone seems to be talking about it – and forming strong opinions about it- and when that happens, it's usually a good think to know just what it is.*" Also Abergel et al (2012, pag.4) observe how algorithmic trading and high frequency trading has suddenly popped up among the most debated arguments over the last decade. In particular, "*a search on articles from newspapers, magazines, academic journals, trade publications, etc, containing the words "algorithmic trading" on EBSCO yields 2502 hits over the period 2005-2011 and only 329 over the period 1999-2004.* "

But before delving deep into the main features of HFT's business to understand its popularity, we instead need to explore first the whole **general financial framework** which has allowed HFT to flourish. As observed by Harris (2003, pag. 89) "*Market structure is extremely important because it determines what people can know and do in a market. […] To trade effectively, you need to know the structure of every market in which you trade."*[6] It is now the time to approach how financial markets have reached the current configuration, where Wall Street is no more where the money is. Following the launch of new exchanges and introduction of innovative trading platforms, markets have become profoundly fragmented. Today speed plays a fundamental role: in electronic market, we need to observe market evolutions on a time scale measured in milliseconds, nanoseconds or even microseconds.[7] Market are built on technological infrastructure which have reshaped traditional stock market outlook, how the order flow is managed, and the

[6] On the same page, the author provides a definition of "market structure": it is the combination of "the trading rules and the trading system used by a market".
[7] A millisecond is $1/10^3$ seconds, a microsecond represents $1/10^6$ of a second, and, finally, a nanoseconds, accounts $1/10^9$ of a second. Blinking an eye requires 400 millisecond.

way data providers offer their services to market operators. Regulation, technological innovations and inner market dynamics have then sparked these profound changes.

In the following paragraphs we will deal with all these topics to draw the most updated picture of financial markets, representing them exactly how they look like after this revolution.

1.2 INFRASTRUCTURAL AND REGULATORY CHANGES

During the first half of the 20th century, the technological infrastructure of financial markets was barely rudimental. It was an era where market complexity was at its lowest historical level: traders used telegraphs to communicate among each other and to obtain market information, while the order submission process was handled through a pneumatic-tube stage system. Market news circulated slowly making the price discovery mechanism inefficient since information was incorporated into prices only after long time. Then, thanks first to computer developments, financial markets started their modernization in the early 1970s. This advancement set also the stage for future adoption of electronic trading. The foundation of **NASDAQ** in 1971 represented indeed the first step into the whole process of market electronification. Approximately 2500 stocks were then quoted on what "*was initially a simple computer bulletin board*" (Gregoriou, 2015, pag 156). Execution time became quicker and reporting trading activity easier. The speed of information flow benefited from the introduction of the **Intermarket Trading System (ITS)** (1978), which allowed market participants to electronically receive trading data.

The "**Designated Order Turnaround**" (DOT), introduced by in 1976, allowed the NYSE's clients to route orders electronically to the trading floor. With the adoption of the SUPERDOT in 1984, it then became possible to route small orders directly to specialists on the trading floor rather than to a floor trader. When securities markets were operating in a completely manual way, the news and

information flow was complex, slow and inefficient.[8] Following the Crash of 1987, the regulatory authorities responded by changing some rules which were deemed to concede too much freedom to Wall Street players. One rule in particular was blamed to be responsible of having made the crisis escalate quickly. At that time, clients were required by regulators to call their brokers in order to submit an order to the markets. During the Black Monday in 1987, Wall Street traders avoided on purpose to pick up calls during the highest panicking moments so that they would have not be forced to take on clients' orders. The new rules introduced by SEC made easier for computers to do the jobs done previously by people, allowing clients to put their own orders manually through **electronic system.** As explained by Michael Lewis (2014, pag.3) "*the 1987 stock market crash set in motion a process – weak at first, stronger over the years – that has ended with computers entirely replacing the people*."

In 1997, the S.E.C introduced the "**Order Handling Rules**". While the regulatory body intervened to restore confidence into the market, by increasing its transparency and proposing new methods for improving execution, it also "*shut down the private market that brokers and institutions were using to trade with each other*" (Arnuk and Saluzzi, 2012, pag. 68), stopping the corruption and illegal procedures which proliferated among NASDAQ's market makers. Through the "**Limit order handling rule**", market makers received new recommendations regarding how to properly handle an order received inside the spread. The main purposes of these indications were increasing price transparency and discovery mechanism plus reducing the spread by ensuring better execution. The "**Quote Display Rule**" prohibited dual posting between NASDAQ and other trading platforms.

The indirect consequence of this regulatory intervention was the proliferation of **Electronic Communication Networks (ECNs)**. As observed by Zubulake and Lee (2011, pag 17), "*ECNs became the main outlet for unwanted limit orders from market makers*" (on the NASDAQ). ECNs are fully electronic subset of Alternative Trading

[8] See Aldridge (2010, pag 7) for more details.

Structures (ATSs) which offer to their users the right to electronically access the market and submit orders into the network via a computer terminal. These trading venues operate at best market price and handle orders according to a price-time priority basis. Execution is first tried internally; when no internal match is available, the order is then routed to other platforms. Thanks to their operative models, ECNs **reduced the need for intermediation** and severely **lowered transaction costs** and **execution latency**. Moreover, they offer anonymity and real time electronic price discovery. Therefore, they quickly attracted a huge clientele. Instinet, the largest ECN, received mainly institutional traders and market makers orders. Island, instead, obtained order flows of "*the under- served but rapidly growing client segments of hedge funds, proprietary desks, program trading desks and retail flow*" (Zubulake and Lee, 2011, pag 19). ECNs quickly proliferate across the whole market spectrum. Despite they were practically an execution venue, from a regulatory standpoint they were categorized as broker-dealers and not as exchanges.

Therefore, in 1998, the S.E.C stepped into financial landscape to regulate this situation with the "**Regulation Alternative Trading System**". According to the new rules, ECNs were then required to become an exchange or register themselves through self-regulatory organizations. This new regulation solved also a limit of the previous "Fix order handling rule" by increasing pre-trade transparency. In particular, all ECNs which traded 5% or more of the volume in the National Market System (NMS) securities were now required to *publicly* display not only market makers and specialists' quotes, but also the ones from institutional investors and all players which do not classify as market makers. This adjustment was conceived to *further consolidate market* and make it more representative of the real quote situation. Up to this time, ECNs contributed positively to order transparency, increasing market access and they enhanced competition. They were mainly diffused on NASDAQ, where ended up to account for a consistent percentage of its order flow. However, they were still absent from NYSE, which was protected by Rule 390. This rule banned any trading activity off NYSE's trading floor for all the stocks listed on NYSE before April 1979.

To stop this unfair advantage, S.E.C **abolished Rule 390** in May 2000 eliminating its anticompetitive nature. ECNs' diffusion was boosted even further and broker-dealers, which were previously forced to sell part of their order flows to NYSE, increased their internalized order flow subtracting further market share to the NYSE. But regulatory actions of S.E.C did not come to an end in May 2000.

In 2001 in fact S.E.C introduced **decimalization**[9] for quoting process all stocks traded on the American financial markets. The minimum price increment size was then changed from the different available sizes (1/16, 1/8 or 0.0625$) to a single common value, $0.01. S.E.C intervened with this regulatory proposal to pursue essentially 3 goals: reduce transaction costs for individual and small investors; make price easier to understand and, finally, adjust American equity markets to global standards, since other countries already embraced this quoting system from a while.

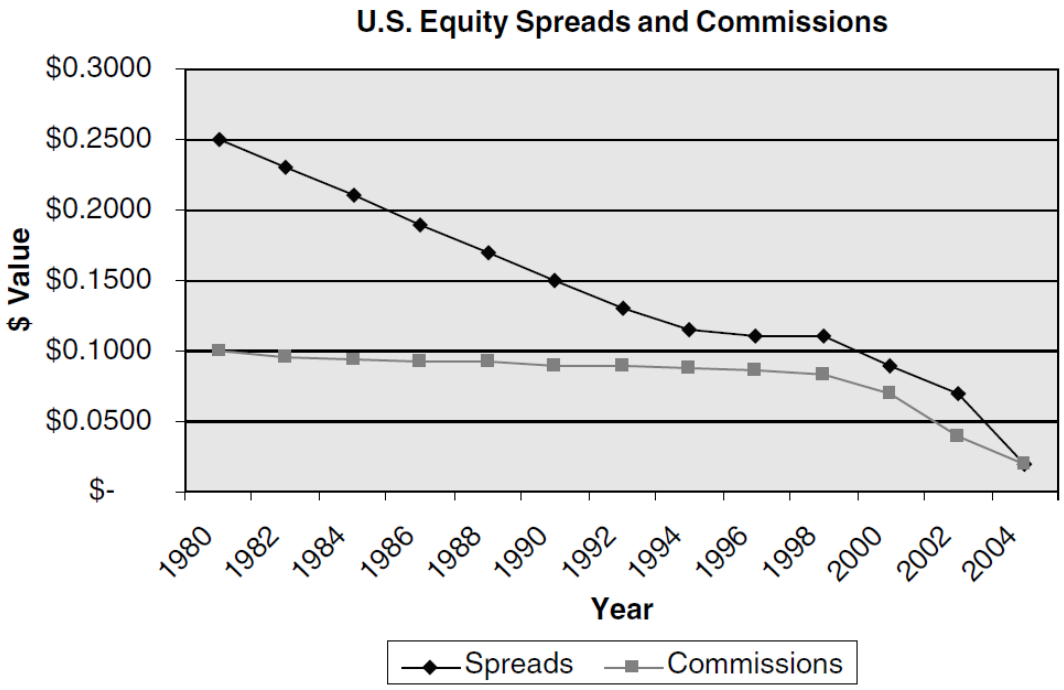

Figure 3- Reduction in spread and commission in US Equity Market. Source: Kim (2007), pag 7. Original source: TABB Group, June 2005.

[9] "Decimalization was proposed by Congress in the Common Cents Pricing Act of 1997, which was later mandated by the Securities and Exchange Commission order 34-42360 in January 2000. Its introduction was executed in three phases to minimize disruption in financial markets" Kim, 2007, pag.6

The picture shows that decimalization effectively contributes to decrease spread, which was already declining thanks to other electronification market advantages. Such reduction in spread simultaneously meant a reduction in profitability for some market operators, established market makers in particular, which in certain occasions went out of the business. The remaining operators instead, to avoid being cannibalized and disrupted by competition, had to adopt new technologies as electronic order management system and routing technology to offer to their clients the most updated services. A significant reduction in spread was not the only consequence regarding the Limit Order Book (LBO) 's features. Because market makers were now able to distribute their bid and ask quotes over a wider range of price levels, they reacted by filling the broad range of available levels in the LBO with their own quotes. Therefore, even depth for each level of the LBO decreased.

In 2005, S.E.C proposed **Regulation NMS** with the aim of promoting competition among exchanges and allowing for greater access to the market. To fully understand this intervention, we need to analyze four of its elements:

1. Order Protection Rule or New Trade Through Rule (Rule 611)
2. Access Rule (Rule 610)
3. Sub Penny Pricing Rule (Rule 612)
4. Market Data Rules and Plans (Rules 601 and 603)

According to the **Order Protection Rule,** exchanges are prohibited to execute an order at a price worse than the best available price, the National Best Bid and Offer (NBBO) [10] . This rule was extended from NYSE, where it was already in place, to the whole equity market and it eliminates the unfair advantage enjoyed by the NYSE up to that moment. This new adjustment was proposed with the aim of increasing protection of displayed prices on the LBO, favor limit orders, encourage market liquidity and depth, and finally foster competition among orders intended to supply liquidity. The **Access Rule** establishes for exchanges and ECNs a fix fee per traded share (at no more than 0.03$) ; moreover, it requires to market centers to develop

[10] Some critics believe that using only the price criteria for defining best execution is a too narrow definition.

and establish procedures to prevent locked or cross markets and also enabled market players to rely on private access linkages supplied by a wide range of connectivity providers. The **Sub Penny Rule** prohibits to market participants to display or accept quotes whose price increment are less than a penny. Finally, **Market Data Rules and Plans** increase access to market data ad establish a new principle for redistributing data revenues, rewarding trading venues which produce the most reliable and useful data according to investors' judgment.

The final results of Reg NMS have been consolidating market regulation into a unified national regulation with no difference between single stock exchanges [11] and pave the way to a full widespread adoption of electronic trading. (Kim, 2007) In the Post Reg NMS market structure, all major exchanges have been forced to launch automatic electronic trading platforms to remain competitive into financial market landscape. **Electronic trading** enables counterparties to be connected through an electronic execution protocol and eliminates the old-style voice brokerage. The most important advantages provided by electronic markets are the following ones:

- More orders can be processed, therefore more volume can be traded.
- Orders are processed at higher speed;
- Faster response to traders' needs is ensured;
- More data can be processed in real time ;
- Reduction of operative and trading costs, as well as fees because buyers and sellers are matched directly without human intermediation;
- Facilitating supervision, as every detail of the trading process is registered electronically
- Moreover, advantages are cumulative because any further progress in information technology enhances all the above benefits.

[11] As observed by Kim (2007, pag 129) "prior to Regulation NMS, the lack of consistent intermarket trading rules for NMS stocks had divided the equity markets into a market for exchange-listed stocks and a market for NASDAQ stock […]. Exchange-listed stocks were subject to the Intermarket Trading System (ITS) rules. […] The result of the ITS rules has been a less than optimal regulatory environment for both exchange-listed and NASDAQ stocks. The ITS trade provisions were from an era of manual markets".

As we will see in the next chapters, electronification of markets has represented a fundamental step for the birth and growth of algorithmic trading and HFT. As observed by Gregoriou (2015, pag 157) "*in a sense, the SEC thought that replacing Wall Street by computer was a great idea. What it probably did not expect is that it also sown the seeds for the development of HFT as we know it today*". After having addressed regulatory changes, we now move to consider how market macro-structure has changed following the forces ignited by competition.

1.3 FRAGMENTATION AND CONSOLIDATION

Competition started to increase in the aftermath of Order Handling Rule, when a consistent number of new and innovative ECNs and ATSs were founded. They attacked the NASDAQ and NYSE duopoly which characterized the market until 1997. From 2002, instead, a wave of consolidation was triggered by ECNs, NASDAQ and NYSE. Archipelago, one of the most famous ECN, merged with REDIBook forming ArcaEx. Instinet acquired Island, which was then trying to establish as the most important ECN in the market. Together they formed INET.

Fearing rising competition and in need of acquiring the technological knowledge that they were lacking for their own operational purposes, both NYSE and NASDAQ started their own acquisition campaign which was fueled by capital raised by going public and revenues obtained by trading fee collection.[12] NASDAQ acquired INET, while NYSE bough ArcaEx and the former duopoly was soon reestablished. The acquisition of ArcaEx allowed NYSE to offer to its clients an electronic platform and the newborn NYSE Group Inc. became the "*first open all electronic stock exchange*" (Kim, 2007, pag 46). Scared of a duopoly's return, institutional investors reacted by backing existing regional exchange and supporting the creation of new execution venues. Large investment bank firms decided instead to develop internal crossing engines to subtract order flow from exchanges. Then, as we already explored in the

[12] Demutualization (the shift from non-profit to for-profit publicly listed business) was identified by the International Organization of Securities Commissions (IOSC) as one important risk since "due to increased pressure to generate investment returns for shareholders, a for-profit exchange may be less likely to take enforcement action against customers or users who are a direct source of income for the exchange".

previous chapter, Reg NMS was introduced with the aim of fostering competition.

Up to 2002, the stocks that didn't trade on the New York Stock Exchange traded on Nasdaq. No stocks traded on both exchanges. In 2015 instead, according to O'Hara (2015) while looking at American financial markets an observer could count 11 lit equity exchanges, 50 alternative opaque trading systems (crossing networks) and hundredths of internalization pools.[13] The times when US financial markets was basically a duopoly has gone. NYSE market share for stock trading regarding NYSE – listed stocks has sensationally plunged in less than eight years from 80% to 20% (Angel et all, 2015).

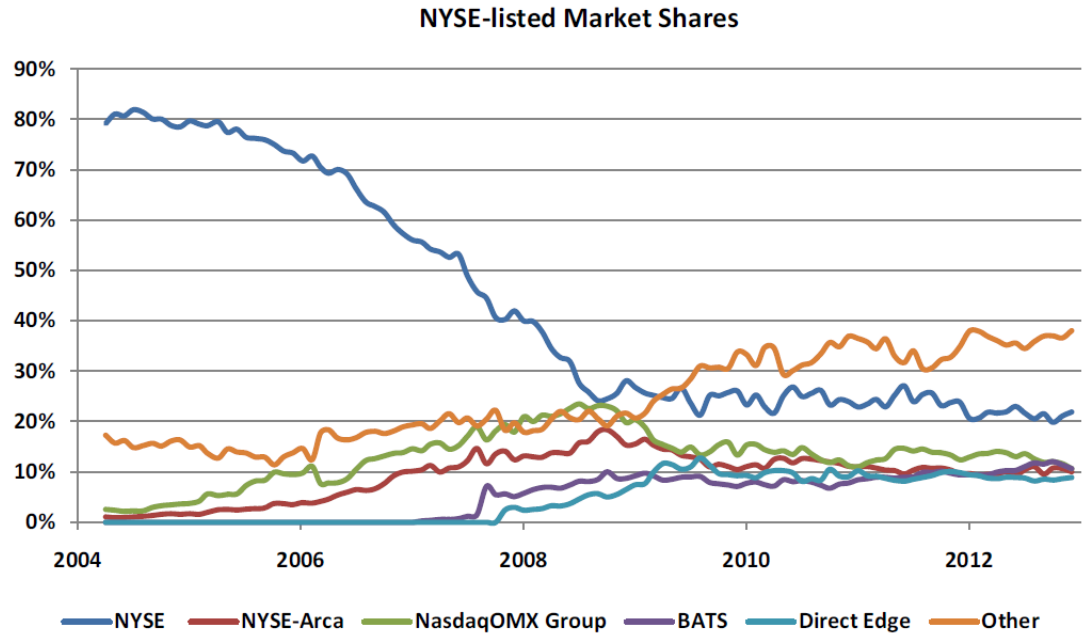

Figure 4 - NYSE's falling market share for stock trading activity related to NYSE- listed stock. Source: Angel et al (2015).

A study from Harvard Business School can help us to understand why this current stage represents a natural ending state of an evolution process where market relies on technological infrastructure. Indeed, as explained by Malone et all (1989) "*companies which try to corner the market with their supply are wiped out by the*

[13] More detailed regarding the meaning of "lit" exchanges, "opaque trading systems" and "internalization pools" are provided in the next pages.

"evolution from single – source electronic sales channel toward "electronic markets" that include many suppliers offerings". The authors foresaw that this type of market, where computers support the construction of supply and demand, are meant to experience a profound disruption into marketing and distribution patterns because of technology. The final market configuration is such that we have a clear distinction between winners and losers. The first ones are the companies which are capable of embracing technology into their business model or which use technology wisely. The latter are the ones which do not adjust to the changes or which try to "*lock in customers through obsolete arrangements*" (Malone et al, 1989).

Exchanges indeed compete on the following fields:

- Explicit costs connected with trading fees.[14]
- Liquidity
- Execution speed related both to market access and orders' execution.
- Data, related to order flow, trading activity, limit order book's status.

Trading venues has not only grown in number, but also in complexity. The traditional public stock exchanges, as NASDAQ and NYSE, but also alternative class of privately owned exchanges has always been **fully transparent** markets, which means that they offer:

- Post trade transparency, with a complete and accurate release of information about the trading process
- Pre trade transparency, with full-detail regarding both sides of the limit order book.
- Market data, which are accessible to all market participants in real time and at low costs.

[14] For each trade, the two counterparties are required to pay a fee to the platform to execute their orders.

Thanks to technological developments, market advancements and investors' needs, nowadays in the market we can also observe another kind of trading venue.

1.4 DARK POOLS

This new venue execution belongs to the off- exchanges category and its main features are opacity and restricted access. On the Street, these platforms are called **Dark Pools.** Private entities owned or operated by brokers or banks, they can also represent an opaque segment of a public lit exchange. Their opacity is related with:

- Anonymity of trading activity and absence of operative rules publicly available;
- No access to resting and executed orders on the book (pre-trade opaqueness);[15]
- Delay into information flow regarding trading process (post-trade opaqueness)

A dark pool can then be conceived as "*an accumulation of orders to buy or sell assets, but whose existence is not publicly known or advertised*" (Banks, 2014, pag 3).

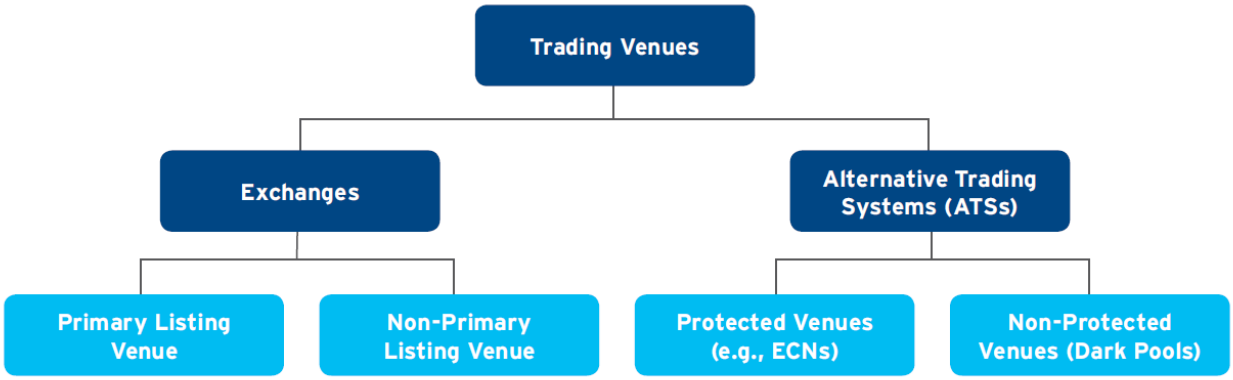

Figure 5 - Trading venues topology. Source: "Market Fragmentation: Does it Really matter?" , Citigroup Transaction Services (2012)

[15] Liquidity is not displayed in dark pools. For this reason, it is known as "dark liquidity" in contrast with the one provided by transparent trading venues, which is known as "lit" liquidity.

Dark pools' clients must tolerate some trust risks to operate through them. But trust is not the only risk. Execution is not guaranteed when an order is sent to a dark pool. Moreover, other market players can extract order flows information out of dark pool by exploiting its own structure with suitable techniques. [16] Nevertheless, dark pools are also attractive for several reasons, both for clients and operators. Clients use dark pools because:

- they offer discretion; market players can reduce the price impact of their orders or lower the probability of being adversely selected.
- offer frequently lower explicit costs in terms of fees
- better price, because dark pools must abide to price improvement rule.

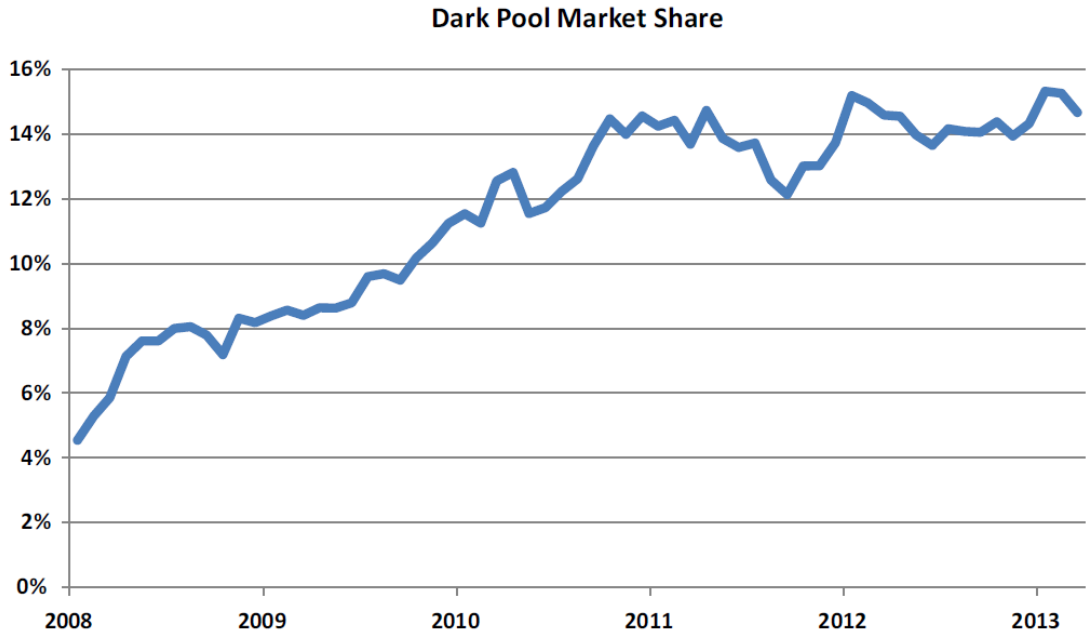

Figure 6 - Dark Pools increasing market share. Source: Angel et all (2015).

Dark pools indeed have to execute their trades at a price within the spread of the lit market: both the buyers and the seller get a better price than the one that they would get by trading aggressively at National Best Bid Offer (NBBO). This explains why

[16] More information about "informational leakage" exploiting dark pools inner dynamics will be provided in the next pages.

dark pool pricing technique is known as **derivative pricing**: the prices available in these opaque venues are pegged to the ones observed on lit markets. As explained by Lehalle and Laruelle (2018, pag 21), "*the rationale of such a rule is that from a theoretical point of view, deals at imported prices do not participate in the price formation process*". There is a wide range of existing dark pools. The most famous ones look like **crossing networks** (CNs) or **Internalization Pools** (SI). CNs are public dark pools which cannot contain proprietary order flow, while internalization pools allow their holders to submit their own orders inside the trading venue.

To fully grasp the reason behind rising importance of dark pools, we need to put their role into perspective of High Frequency Traders. Dark pools were indeed originally created to provide to institutional traders a way to escape predatory strategies of HFT traders. HFT have gradually improved their ability to detect big orders into lit market by exploiting trading patterns' predictability and knowledge of market microstructure. Feeling to be as prey for HFT traders, institutional investors demanded a place to trade safely, where they can hide their orders away from HFT traders. Dark pools were then founded exactly to satisfy this need offering anonymity during the trading process and opaqueness. But the relationship between HFT and Dark Pool has become by far more complicated with the latest market developments, because now the situation has in fact come full circle. After becoming a successful business, with rising market share and more traded volume, dark pools started to explore any possible route to grow even further. The easiest way they found to enhance their volume was to open the doors to HFT and let them enter into the dark to trade.

1.5 LATENCY

After electronic markets diffusion, order execution speed is not limited anymore by human capabilities of handling, processing and executing orders. These new trading platforms has introduced a "**matching engine**", a network of computers, to connect market buyers and sellers without human interaction. The only constraint today is indeed represented by how fast an electronic signal can travel between the

beginning and the ending point of the transactions. For example, between the data center in Chicago that houses the Chicago Mercantile Exchange and a data center beside the Nasdaq's stock exchange in Carteret, New Jersey. Such limit is simply determined by the law of physics: the speed of light.

Latency is defined as the amount of time it takes for an investor to:

- receive a signal from the trading venue
- process the information
- react by submitting a new order (a new quote or a cancellation)
- receive feedback from the market

In electronic markets, latency depends on:

- computing power of trader's IT system (hardware and software)
- trading platform IT systems
- low- latency services offered by trading platforms to traders

The physical location of trader's computers servers next to exchange's own servers, which then provides high speed access, is a service known as **co-location**. Co-location has certainly become the most famous type of low-latency service. By renting space inside their facilities, exchange increase their revenues. On the other hand, traders are interested in obtaining the spot which is closest to the matching engine because it allows them to observe market movements first (*information* speed) and therefore respond to them quicker and adjust their orders. Speed differences indeed are costly. When a trader does not get quick access to the market and his orders are then executed with delay (*matching speed*)[17], he could not be able to catch up profitable trading opportunities. Speed execution is determined both by distance, which is minimized through co-location, and investments in technology infrastructure. Case sizes, materials, type of fiber used, digging process. They all

[17] Informational speed and matching speed are concepts introduced by Foucault and Moinas (2018). As observed by the authors, informational speed and matching speed "*are conceptually different, but in practice, they are bundled and difficult to disentangle*".

matter. They all make the differences between being first or not. Translated in monetary term: profit or loss.

Thanks to technology, the gap between the maximum theoretical trading speed and the real trading speed has significantly decreased over time, barely hitting zero. The demand of this new service, speed, has triggered a " **race for speed**" between different players who tried to fully take advantage of modern technological development. Telecommunication companies were the first ones which started to provide infrastructure to Wall Street players. Several projects have been realized regardless natural landscape structure, roadway infrastructure, urban utility. But what is meant by "*fast*" has changed rapidly in the financial markets. Up to a point, Wall Street' demands reached in fact a paradoxically situation which is perfectly pictured by Michael Lewis in his book. During the construction process of a telecommunication line which would have then be sold to Wall Street firm, the main project's investors ask to its construction engineer to cross diagonally rather than proceeding by ziz-zag because otherwise he would have cost him "*a hundred nanoseconds*"(Lewis, 2015, pag.13)[18]. From 2007, this "arms race for speed" has never ended. In June 2018, Nasdaq invested into a project to synchronize a giant network of computers on a nanosecond timeframe. Balaji Prabhakar, a Stanford University electrical engineer who has been hired by the exchange for designing the algorithm, [19] declared to the New York Times that "*the financial industry has easily become the most obsessed with time*". (Markoff, 2018). The opening sentence of Markoff's article was even more captivating. Referring to a technology developed by computer scientists from Google and Stanford University, which will offer the possibility to track time down to 100 billionths of a second, the author observes that "*this could be just what Wall Street is looking for*".

[18] A nanosecond is a billionth of a second.
[19] This algorithm will help the stock exchange to "accurately order the millions of stock trades that are placed on their computer systems every second" (Markoff,2018).

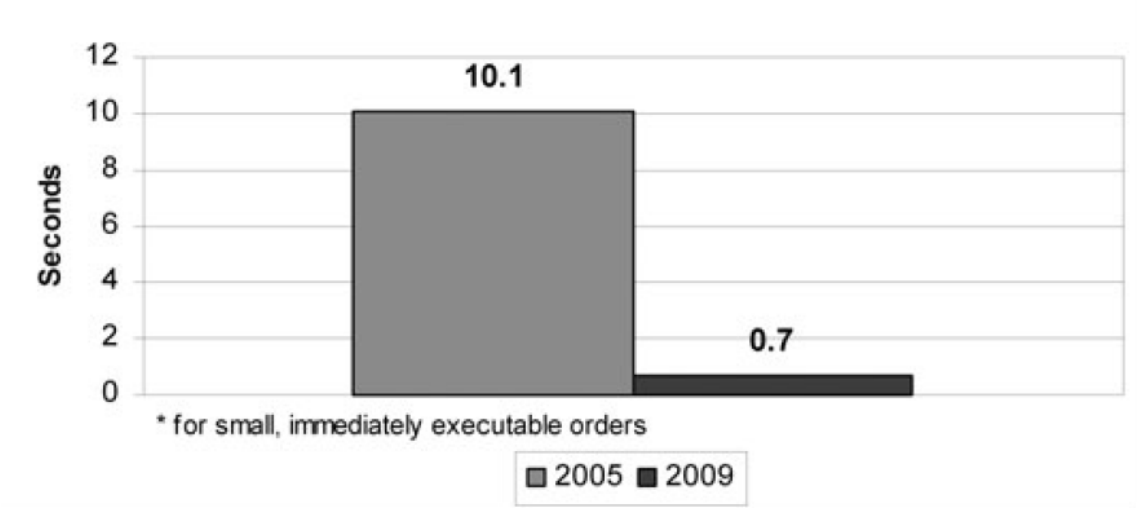

Figure 8 – Average trade size for NYSE listed stock; S.E.C 's calculation.
Source: Durbin (2010), pag ix. As commented by Durbin, "Humans are fast, but not that fast".

Getting access to the fastest line has clearly become crucial to traders. It allows them to perform the most profitable arbitrage strategy because gaining money is not a consequence of having the smartest strategy, but getting access to the fastest connection. But access to the market without proper data is not useful. Therefore, exchanges also offer direct access to data feeds through the so-called **Direct Market Access (DMA)** low-latency service**.**

To sum up, the diffusion of electronic trading and dark pools, market which have become progressively more fragmented and a arms race for speed which continuously reduce latency time have clearly reshaped financial markets structure. The old -style picture with trading floors full of yelling human traders is not reliable anymore. HFT traders play a fundamental role into this new market scenario and this is exactly the topic that we are going to address in the next chapter.

CHAPTER 2 – HFT: MARKET ROLE AND TRADING STRATEGIES

HFT has took financial markets by storm. Everything "*from the way traders trade, to the way markets are structured, to the way liquidity and price discovery arise – all are now different in the high frequency world*" (O'Hara, 2015). Today, given both current microstructure and financial markets landscape, there is no chance to understand the markets without understanding how high frequency trading operates.

2.1 HFT BUSINESS

Up to now, there is not yet a well-accepted definition of High Frequency Trading. However, in 2010, Securities and Exchange Commission (SEC) described in one report (SEC,2010, pag 45, note 39) the 5 main features of HFT traders:

(1) The use of extraordinarily high-speed and sophisticated computer programs for generating, routing, and executing orders
(2) 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 (i.e high turnover of capital)
(4) The submission of numerous orders that are cancelled shortly after submission
(5) Ending the trading day in as close to a flat position as possible

HFT is indeed a low margin, high volume, profitable business. With respect to low-frequency trading business, as the one carried out by institutional investors or hedge funds, the daily number of trades is incredibly higher, the usual holding period is tremendously lower, the average traded size is enormously smaller and finally the average gain per traded position is conspicuously tinier. The decision to not carry on inventory overnight is fundamental for risk management purposes. In fact, in today global markets, trading operates virtually on a 24-hour cycles for several asset classes. Volatility therefore can exacerbate price movements during other international exchanges' trading hours, without giving to a trader the possibility to hedge its positions due to different working hours. Moreover, overnight positions imply the so-called carry costs required to

keep a position opened. A daily flat position then reduces both risks and costs.

On February 2009, Aite Group estimated that "*HFT accounted for over 60% of trading volume coming through the financial exchanges*" (Aldridge, 2009, pag 1). On November 2011, a Tabb Group's representative declared to Bloomberg that according to his firm's estimates "*HFT is responsible for over 77% of transactions in the UK market*". According to Nanex instead, "*HFT currently generates approximately 35% and 70% of K and U.S equity trades, respectively*". But quantitative estimates are volatile, time contingent and they vary across markets, trading venues and time period. What is likewise important is perception: among Wall Street players, regulators and also on the media side. Marko Kolanovic, from JP Morgan, speaking about the relationship between machines and humans in the finance industry, wrote "*First, we note that for short term trading, such as high frequency trading market making, humans already plays a very small role*" (Kolanovic, 2017, pag 8). Markoff (2018) talking about equity market, declared that "*stock trading is now dominated by computers, that make buying and selling decisions and execute them with blazing speed*". Finally, there is a widespread consciousness among regulators that high frequency traders play a dominant role in certain market activity.[20]

HFT has been originally implemented by small-proprietary funds far away from the traditional Wall Street players, in terms of business model but also with respect to geographical considerations. Citadel, Tradebot, KGC, Virtu Financial, Hudson River Trading, and Tower Research are the most famous names of HFT firms. Many of them are located outside Lower Manhattan, as in New Jersey, Missouri and Illinois. These new players enjoyed several advantages, as less stringent regulatory requirements but also more sophisticated technological skills. Thanks to these characteristics, they originally ripped off significant trading revenues from consolidated Wall Street players. [21] Then, as technology has spread all over and these practices have become more popular, they have also been adopted by other players, as hedge funds and proprietary trading desks of bulge bracket firms.

[20] We will address this topic in Chapter 3.

[21] Remember discussion in Chapter 1. Because of technology advancements, some old-style established market players as old-school market makers went out of business because they could not keep up with competition. The one from HFT players.

HFT firms have been among the few players who registered positive returns during 2008 financial crisis. This industry has then reached its profit record in 2009, at approximately $5 billion, growing quite faster from 2005. Then, profit has fallen because of higher competition and reasonable consolidation[22]. Because of business model features, HFT firms in fact frequently compete with other peers for trading volume. A decline in volatility, following 2008 financial crisis, is certainly another reason why their profitability has declined.

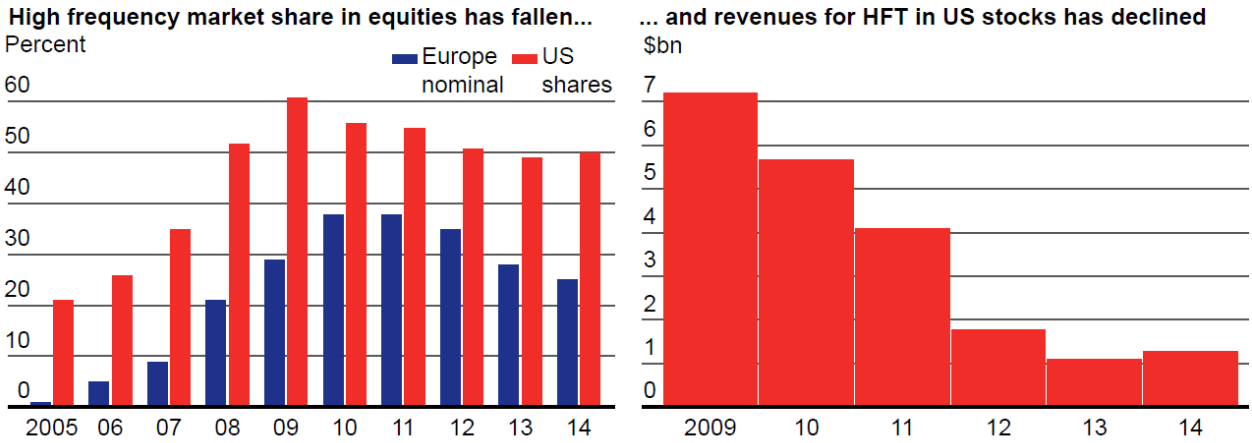

Figure 9- HFT market share (in EU and US) and revenues (US).
Source: Gregoriou, 2015, pag. 134. Original source: TABB Group.

Because of their low correlations with traditional long-buy and hold strategies, HFT strategies represent a valuable diversification tool for portfolio management. (Aldridge, 2010, pag 2). Moreover, because of operative features, HFT represents a trading approach which could potentially ensure stable profit despite actual market conditions and regimes. In fact, HFT methodologies have been gradually implemented also by large institutional investors, hedge funds and proprietary desks of bulge bracket firms. Nowadays HFT is widespread, across all asset classes and around the world. It started into U.S equity markets, before expanding rapidly thanks to progressive electronification of stock exchanges and then the one of also

[22] The aforementioned KGC is the result of the merger between Knight Capital Group and GETCO.

other asset classes' trading venues. As observed by Ait-Sahalia and Saglam (2013, pag.2) "*HFT may have plateaued in some of the markets where it was first introduced, but it is still expanding globally in new markets*".

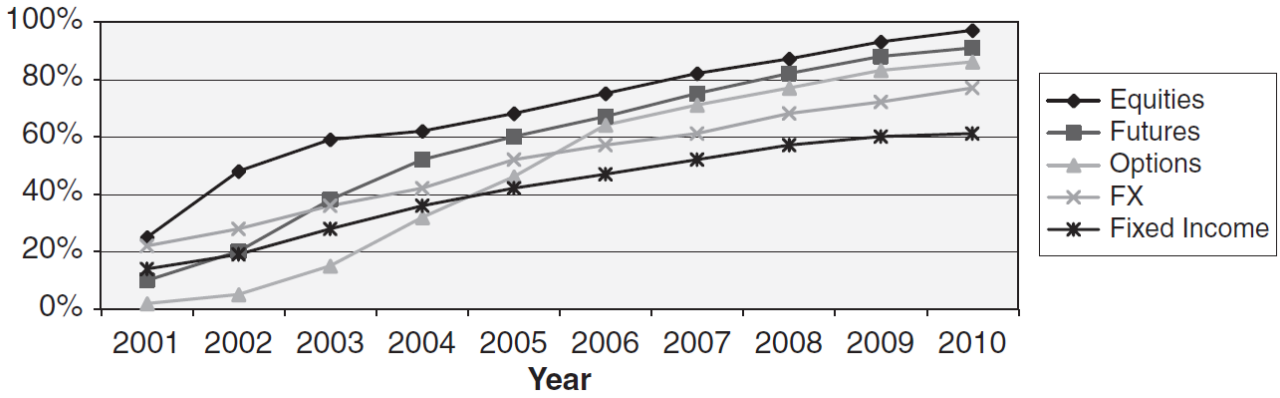

Figure 10 - Adoption of electronic trading capabilities by asset class.
Source: Aldridge (2009), pag. 10; Original study: Aite Group.

As observed by Menkveld (2016, pag .2) "*there is a symbiotic relationship between new electronic venues and HFTs"*. High Frequency Traders need the services provided by these modern trading platforms, such electronic automated trading platform, low fees and as co-location services to be effectively profitable. All these requirements are of fundamental importance, because they determine a positive environment for HFT's diffusion. The absence of any of these conditions can impair HFT activity. As observed by Wang and Zheng (2015, pag. 5), adoption of HFT in the Chinese future market has been delayed until 2005 because of high trading costs and low liquidity, despite an automatic electronic platform was already available. On the other hand, the trading activity of these High Frequency Traders reduces transaction costs for operating through these infrastructures and also increase volume, determining an increase in revenues for these exchanges.

Combining features of HFT's operativity, as high capital turnover and average traded size, together with our knowledge about which market they traded as first, the equity market, we can investigate some of the market dynamics that have emerged in the previous decade focusing our attention on the NYSE stock exchange example. The average number of daily trades has increased by more than 10 times

in less than 4 years. Simultaneously, the average trade size has more than halved. HFT has then provoked a profound impact on the consolidated market dynamics by completely revolutionizing them.

Figure 11 – Average daily trades for NYSE listed stock; S.E.C 's calculation.
Source: Durbin (2010), pag vi.

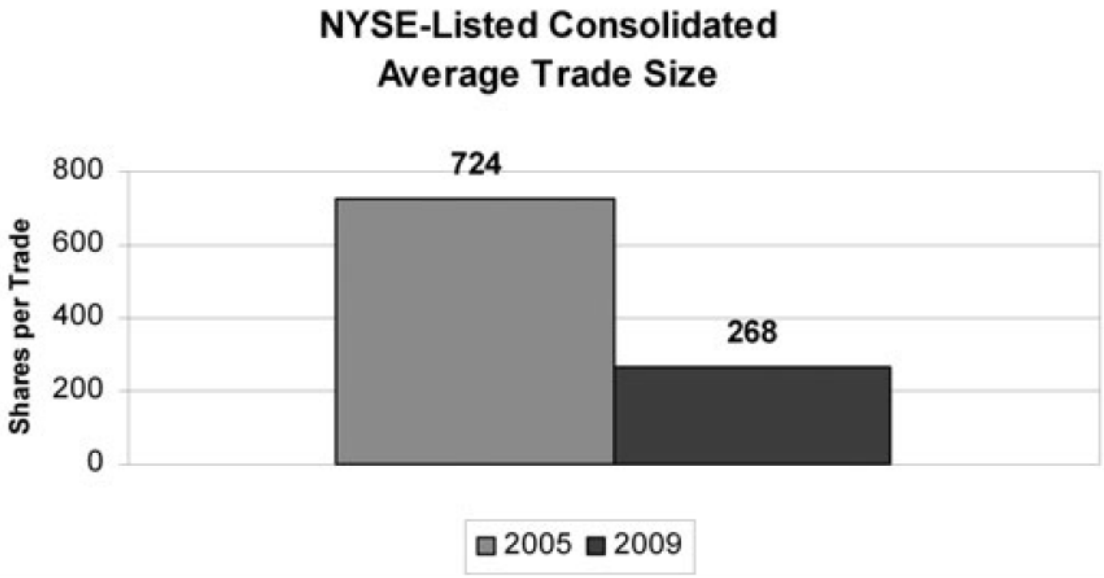

Figure 12 – Average trade size for NYSE listed stock; S.E.C 's calculation.
Source: Durbin (2010), pag vi.

Despite our focus on equity market, high frequency trading strategies can be applied to any securities which is **sufficiently liquid** and **exhibits enough volatility** to be profitably traded. Moreover, the product need to be traded on an electronic platform, to ensure fast execution and no human intermediation. Currencies and equities therefore represent the ideal products to satisfy High Frequency Traders' appetite. However, a lot of HFT funds are product agnostic. Other instead choose to specialize in specific securities. The perception of having a technological and modeling advantage is often fundamental for selecting the range of products to be traded. Specialization can be driven not only by specific trading skills or technological capabilities, but also by geographic reasons.[23]

Despite the selected product and the chosen regions, the factor that determines the success of a trading strategy is the **modeling process framework**: each strategy in fact must be carefully designed according to the market microstructure where the algorithm will be applied. As observed by Wang and Zheng (2015,pag.7), "*a profitable algorithm in one market may not work in another market*". Designing and operating an HFT trading system is a 3 steps process which requires to perform the following operations:

1. Data analysis and model building
2. Trading system design
3. Implementation and capital allocation

HFT benefits from immediate access to market data: order book information, most updated economic and financial news, political changes and tech breakthroughs. They all define current market conditions and then feed the knowledge appetite of HFT traders. Exchanges, brokers, independent news financial firms offer all these valuable services and updates and compete fiercely in this field.
As raw data, which can potentially affect security's price because of their informative content, all these pieces of information are analyzed through the lens of

[23] There are some HFT companies which exploit their proximity to certain trading venues to specialize in some specific asset classes. For example, many Chicago HFT firms exploit their location close to the Chicago Mercantile Exchange to implement trading strategies for derivatives and commodities.

technical analysis, fundamental valuation and market microstructure models. These models can be internally developed by trading firm's research departments or sometimes be inspired by academia discovers. Once information is digested, the observations derived from this burdensome and highly quantitative analysis form the building blocks of "*probability – driven econometric inferences*" (Aldridge, 2009, pag.14) which should help to forecast future price movements. Models are generally run through Matlab, R and Python software. [24] Before investing real money, the reliability of the produced trading signals is backtested against historical data. A two-year time span is the average period with respect to which trading strategies are tested. This temporal choice should allow to face several and different market situations. Despite this procedure surely reinforces strategy's robustness, historical back-testing have two severe limitations: historical returns are never a guarantee of future ones; simulations cannot replicate orders' impact on the markets and then reflect which would have been the real profitability of the tested strategy.

Once trading signals are obtained from these quantitative models, they need to be turn into executable market orders. To submit such orders to the market, a trading system must be designed, built and organized. Trading firms face several choices: hardware, software, message protocol are the most important pieces of a trading infrastructure that need to be selected. A firm needs also to understand whether it wants to rely on external furniture or build its own equipment in-house. Market solutions are cheaper and easier to use than highly tailored solutions. On the other hand, they cannot perfectly suit all the trading needs a firm should reasonably have. High Frequency Firms, whose profitability is determined by their capability of minimizing latency, often choose to design their own trading infrastructure by themselves from scratch. Whatever the choice is, trading infrastructure's success is determined by its leanness, structural flexibility, capability of being constantly updated and speed.

Hardware are mainly related with data storage systems. Unstructured data are

[24] More technical detailed on the process can be find in "Algorithmic Trading: Winning Strategies and their Rationale", Ernie Chan, 2013.

downloaded from social media like Facebook and Twitter. Market data are provided by various vendors such as Reuters and Bloomberg, but also from exchanges and brokerage firms. Essentially, market participants deal with thousands and thousands of data terabyte. High Frequency Firms work on tiny time interval, which measures a fraction of a second. This trading frequency severely increases the data volume to be managed. **Software** instead deals essentially with order execution process. Algorithms play the lion share of activity with regards to this task. The number of algorithms available today is enormous; this make the final choice more difficult, but it also offers a better suitability to satisfy a wide range of needs. **Message protocol** allows traders to connect with trading platforms and obtain orders execution. Following market fragmentations, trading systems are generally prepared using "Financial Information eXchange" (FIX) protocol.[25] This "*special sequence of codes optimized for exchange of financial trading data*" (Aldridge,2009, pag 31) ensures that codes are "**platform independent**" and then can be understood easily regardless the destination determined by the routing process.

Setting up such kind of trading system takes on average 18 months and it is a quite expensive process. With respect to traditional financial institutions, which face fixed cost from inception until go-to-market phase, the bulk of monetary and temporal costs for HFTs belong to the first phase. The second and the third phases are instead cheaper and faster because of lower complexity. These features are reflected in the trading's cost curves below.

[25] "The FIX protocol is a standardized way for various participants in the trading process to communicate information." Source: "Inside the black box, a simple guide to quantitative and high-frequency trading", Rishi K Narang, Second Edition, 2013.

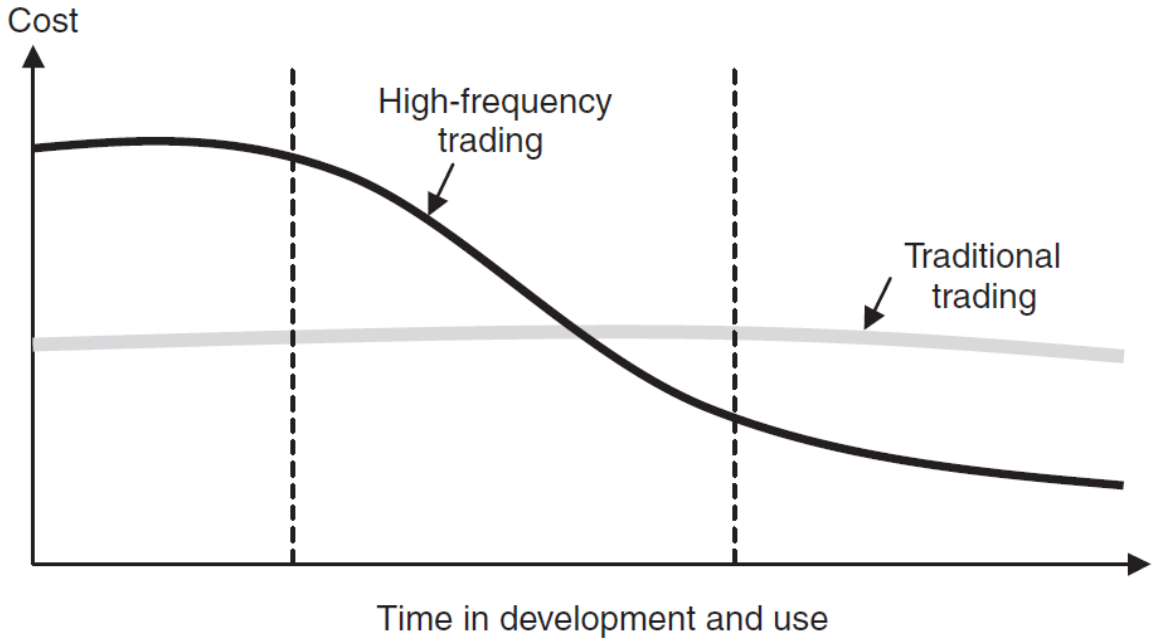

Figure 13 - High Frequency Trading systems, costs and time.
Source: "High Frequency Trading, A practical guide to algorithmic strategies and trading systems", Aldridge, 2010.

2. 2 HFT' STRATEGIES

As we already seen, the main features of high frequency trading activity are a conspicuous number of trades, high-turnover of capital following changes into market conditions and a low average gain for each trade. From a strategy implementation perspective however, there is no clear-cut business model for high frequency traders: trading opportunities range from micro-second price moves to several minute or hours long strategies. Some of the strategies are nothing else than old-school trading methods applied at a higher frequency, thanks to advancement of technology and computer management capabilities. Then we can also identify a subset of strategies that are unique to HFT given the technology on which they rely on. Other, instead, rely on the specific capability of HFT of quickly taking advantage of information flows, both regarding economic and financial conditions and order book status. Any HFT strategy requires to be implementable through an algorithmic trading programs run by a computer. Before addressing the most common trading strategies of HFT firms, listed in the table below, (Aldridge, 2014), it is important to

give some preliminary information regarding features, risks and main elements of these operations.

In a HFT framework, price movements are analyzed at a microscopic time span: fraction of a seconds, as milliseconds, microseconds or even nanosecond. The maximum theoretical gain which can be obtained at every frequency is determined by the sum of all time-interval movements at a given frequency. It is reasonable to assume the lower the observation period, the smaller could be the price movement. However, at the end of the day, when we sum up all the time-interval profits, "*the gain potential in the high-frequency space is nothing short of remarkable (as is the maximum potential loss).*" (Aldridge, 2009, pag 76).

Despite trading frequency and investment style, what High Frequency Traders look for is *persistence of the underlying tradable phenomena*. **Persistency** is connected with price movements which do not exhibit randomness and therefore can be predicted with reasonable certainty. Several tests can be run on return time series to identify these prices' features. Persistency also means that a specific feature is recurrent. This combination is exactly what HFT traders desire. Once they spot such kind of occurrence, they use their trading tools to benefits from that presence.

HFT operates with leverage and they aim to capture high Sharpe Ratio. Leverage allows the firm to cover costs: transaction costs, given that HFT is a high-volume business; operative costs, because as we saw designing and implementing a HFT trading system is a complicated and expensive procedure; research process, because of market monitoring requirements. High Sharpe ratio instead reduces the risk of a catastrophic loss and simultaneously means consistently profitable operation.

The classification proposed follows a functional approach as suggested by Gregoriou (2015). Since HFT is a very heterogeneous concept, it is indeed not easy to identify "*a clear mapping between institutions and strategies*" (Gregoriou, 2015, pag 159) which instead could have justified an institutional approach.

Table 1 – Classification of High Frequency Strategies

Strategy	Description	Typical Holding Period
Automated Liquidity Provision	Quantitative algorithms for optimal pricing and execution of market-making positions	< 1 minute
Market Microstructure Trading	Identifying trading party order flow through reverse engineering of observed quotes	< 10 minutes
Event Trading	Short-term trading on macro events	< 1 hour
Deviations arbitrage	Statistical arbitrage of deviations from equilibrium: triangle trades, basis trades, and the like	< 1 day

Source: Aldridge (2009)

Arbitrage strategies

Academics refer to arbitrage opportunities as "the possibility of a risk-free profit at zero cost". These opportunities arise when instruments that are structurally correlated behave differently, breaking the "one price" market law. These situations are triggered by trading imbalances in one market venue, which eventually create

these price oscillations which are immediately killed by the prompt interventions of arbitrageurs. Despite this procedure is often critiqued as being "unfair" or "speculative", it fundamentally helps the market in the price discovery process and in maintaining fair prices.

In order to trade profitably an arbitrage opportunity, a trader should carefully study the market to spot it and then design a proper execution strategy. This is a complex process which can be summarized in 5 steps:

1. Find a **stable** relationship based on prices.

   Within a statistical framework, we define "stable" a relationship which proves to hold with 90% confidence, which represents the lowest acceptable confidence threshold. The assumption behind arbitrage opportunity is that a stable relationship, if violated[26], would **mean revert** towards its long term equilibrium level, which is established with the help of meticulous statistical studies.

   Such identification can be achieved through data mining, scanning tons and tons of market data. Once a sufficiently clear association is pointed out, it needs to be interpreted and then explained with logical arguments. Relationships can be indeed random or spurious; in those cases, they would have little predictive and staying power. That's' the reason why traders should test these findings against economic theory. Otherwise, they could also rely on academic researches which have proved existence of solid and long-lasting relationships which populate the market. [27]

   A solid knowledge of probability theory and statistics is crucial to successfully operate in this battleground for two reasons: the first is related

[26] Such violation is expressed by the distance of the current level of our variable from its historical mean. This distance is generally quantified by the numbers of standard deviations.
[27] We can refer to these arbitrage opportunities based on violation of economic theory as fundamental arbitrage.

with the study of market data, which requires to identify the aforementioned stable relationship. For trading profitably the most stable and simple relationships, technological advancements are more important than modeling skills because being the fastest movers is what allow to capture these opportunities. However, for the less stable relationships which exhibit more variance and could turn to be indeed even more profitable, a thorough mathematical understanding and modelling skills is the combination that enable market players to profit from these inefficiencies. The second reason regards market structure. There is a time gap between exchange platform sending out the current bid-ask information to trading desks and traders' orders reception by the trading platform. During these infinitesimally small-time instants, prices can change because of new orders or information. Since arbitrage is a business which aim to capture even the smallest spreads, there are some operations which are implemented to capture the tiniest differential. It is then important to work with enough margin of safety to avoid that adverse market movements during these time intervals could eliminate gains transforming instead traders' moves into losses. [28]

2. Define a "**pseudoinstrument**", a combination[29] of market securities which represent our trading opportunities.

3. Calculate market data related with this pseudo instrument.

As explained by Gregoriou (2015, pag 161), "*knowledge of the order flow allows one to play optimally the arbitrage in size*". Trading prices, bid ask spreads, bid ask volumes, trading sizes and multiple trading are the most important variables. This is a fundamental step because it allows trader to analyze market framework as if they were trading a single instrument and all

[28] Because of its complexity, arbitrage has always been a hot debated topic.
[29] Actually there is no restriction about the number of assets which could be involved in arbitrage. According to Wang and Zheng, "statistical arbitrage can be also applied to trading repeatedly only one asset with the same algorithm" (Wang and Zheng, 2015, pag 38). For more information about how to implement these trading strategies, refer to Chapter 7 of "High Frequency Trading and Probability Theory", Wang and Zheng, 2015.

the fields have a similar meaning.

4. Define a proper trading multiple for each leg (traded securities) of the arbitrage. Leg depends on the trading size of the underlying instrument.

5. Run a code to calculate trading strategy parameters. [30]

The preliminary forms of arbitrage opportunities were related to rule out price discrepancies between exchange platforms which traded the same security in different countries. With markets developments, globalization, technological advancements, arbitrage has become a more complex business. These developments have significantly increased the number of arbitrage trading strategies that can be explored into the markets.

Calendar spread is about trading the same product with different delivery months. The underlying assumption is that expected futures prices for different delivery months would share the same or very similar price movements. A well-known example of this strategy is the spot-futures spread, which requires to open a position in the spot market and another one in the futures market. Cross-market arbitrage is another trading opportunity. It requires to identify a product which trade on different trading platforms and open positions to arbitrage away any price differences. Cross product arbitrage relies on internal relationship between several products to trade away price differences.

Despite the selected strategy, arbitrage is a business which needs to be implemented systematically in order to deliver profitable results. A single trading operation could indeed turn into a losses and carefully designed risk management actions must be disposed. Arbitrage in fact turns out to be a successful business when we consider its average result. To define a successful algorithm arbitrage strategy, a trader needs to identify an underlying "*ergodic stationary process with a positive mean, so that the*

[30] For more practical information related to encoding a strategy in a proper algorithm, refer to "High Frequency Trading and Probability Theory", Wang and Zheng, 2015 , pag 45-46.

strong ergodic theorem assures the accumulated profitability increasing in a stable manner" (Wang and Zheng, 2015, pag 3).

Arbitrage involves several risks. A not comprehensive list of them is presented here:

- Market risk: securities' price can change in adverse direction with respect to the one desired.
- Opportunity and time risk: design a proper trading window is crucial for capturing profit and avoid losses.
- Contract size: in certain occasion, products cannot be traded with desired quantity.
- Missing risk: not all the desired positions can be opened and closed exactly as suggested by a designed trading strategy. A miss happens when after having opened or closed a position in an instrument, a trader cannot open or close the one in the other related instrument. It is the most significant risk of an arbitrage strategy, which can also lead to the biggest failure in some occasions.[31]

Moreover, statistical arbitrage strategies can be influenced by adverse market conditions or idiosyncratic performance of securities. Transaction costs or wide bid ask spread can wipe out any potential profit. Finally, it is of paramount importance to observe that due to operative timing, these kind of strategies could require to keep a position opened for more than one day, waiting for mean reversion. This means that differently from HFT market makers, HFT arbitrageurs tend to close the day with high inventories (Lhabitant, Gregoriou, 2015).

Given the nature of the arbitrage business, specifically what ensures a trading edge and its risks, high frequency trading is perfectly suitable for this activity. In fact, arbitrate strategies are among the most popular strategies employed by high frequency traders. As we already seen, these players have quick and stable access to the most updated market data. Moreover, they operate using highly sophisticated

[31] The solution is trying to minimize the miss rate. We have two solutions: first, it is closing the first position which lead to lose bid ask spread and commission; second, open a less than optimum position in another instrument, but this could not lead to a profit.

and advanced technological infrastructure. Since mathematical modeling, data storage and trading speed are fundamental to profitably exploit arbitrage opportunities, it is easy to understand why HFT rely heavily on these activities to bring profit to their trading desks.

Event arbitrage strategies

Event arbitrage is about trading on information flows which moves the markets. We are talking about piece of information which are reasonably considered as likely to significant affect a security's price. News can be economic or industry specific. Information which are already expected by market players are incorporated into prices thanks to expectations. Therefore, in order to have a significant impact on the price, an information must contain an "unexpected" feature. Event arbitrage is so important to the market because according to the *efficient market theory*, a market is **efficient** when information is immediately incorporated into price. This means that we can expect to find some traders in the markets who are eager to catch these trading opportunities as soon as they arise.

These traders are often High Frequency Traders. What makes HFTs traders suitable to perform this role is a set of their operative characteristics: news can be identified immediately only using tick by tick data, which are accessible because of their subscription to premium services offered by brokerage firms and exchanges; news can be traded profitably only when it is possible to rely on fast-speed technology infrastructure and highly sophisticated quantitative models; HFT can promptly take advantage of cross-asset relationships, since quite often an information does not affect the price of one security only, but a more wider range.

Beyond what makes HFT suitable for event arbitrage, to profitably trade news flows a market player must have a profound and comprehensive knowledge of how the economics machine works. Knowing how market could react to certain events, understanding which could be the most and less disruptive announcements[32], being

[32] Eichenbaum and Evans (1993) and Grilli and Roubini (1993) noticed how central banks' news decisions could be the less disruptive ones, since they are bound by a mandate in

informed about market operations are all skills which help a trader to forecast how markets will digest information and trade profitably on the basis of these expectations.

The trading process for arbitrage event opportunities is a four steps procedure which requires to:

1. Identify dates and times of past events in historical data for each event type.

   For each event, we need to specify an "*event window*", which is the time period which elapses from the news announcement until the end of the trading opportunities. To perform a significant and valuable analysis, we need to understand which could be the variables affected by the news. Once they have been identified, a trader downloads and storages these data, before working on it. They can be sampled according to different trading frequencies in order to study trading opportunities on different time scale. In this operative phase, recording accuracy and database' dimension (number of variables and length of the observation period) are crucial to provide reliable studies.

   To get access to significant market data, traders can employ a **strategy spider** "*to scan the internet quickly and filter valuable information*" (Wang & Zheng, 2015, pag 54). Obviously, a thorough and reasonable research cannot be conducted on the whole internet. It is therefore fundamental to identify a restricted group of reliable sources, such as official sites for economic reports, official sites for government announcement, financial reports of listed companies, established new sites and also some quick insider sources, such as Twitter financial profile.

2. Assess historical price changes at the desired trading frequencies in occasion of similar events.

---

their operations.

Quantitative analyst come into playing field for this game. They perform *sign tests* to assess trends' existence. In the absence of event, price changes should be equally positive or negative; when instead an event occurs, the price chance should be **persistently** positive or negative according to how market digests information. When the answer to the test is statistically significant, a suitable event arbitrage strategy could be designed to profit from these observations. This kind of test should be performed at different trading frequencies, because markets can react to different events with different attitudes: the length of the adjustment period and the persistency of reactions are indeed very volatile and unstable.

3. Estimated expected price responses based on historical price behavior surrounding past similar events.

   The sign test offers useful support to perform this task. It provides a valuable hint for preliminary economic forecast because it gives information about market's reaction direction. However, further refinements can be pursued thanks to modern quantitative technique. In particular, quantitative economic analyst can perform **event studies,** which are quantitative estimates of announcements' impacts, for producing more specific point forecast estimates.

4. Drawing conclusions for different trading frequencies and design suitable trading strategies.

   A trader should first define a proper trading window. This generally begins just before the event and ends shortly afterward; it can also be defined in advance whether date event's announcement is known in advance. Moreover, solid economic and financial knowledge are crucial during this phase. A deep understanding of how financial markets works could help to frame cross-securities correlation movements. These observations could help both to reduce risk and multiply profitable trading opportunities. For

example, we known that equities and stocks options are intimately connected. Therefore, an option market maker must then keep a very close ear on stock market news.

Financial markets nowadays are buried by tons and tons of information each day. A not fully comprehensive summary of news which can reasonably be considered to have potential to affect the markets is presented here below.

Table 2 - Economic and market announcements.

Type of event	Variables to observe
Corporate news	Quarterly and annual earnings release Mergers and acquisition announcements New product launches Stock splits Dividend policy changes
Industry news	Industry regulation Tariffs Economic conditions related to a particular industry
Macroeconomic news	Interest rate announcement by major central banks Economic indicators determined from government collected data Regional economic performance

Market making

Making the market is the activity of providing liquidity to market players who demand it to trade. **Liquidity** is the ability to buy or sell whenever you want

(quickly), whatever you want (high volume), at low costs, and at a price close to the security's consensus value (small price concession).
Market makers (MMs) are the intermediaries[33] who provide this valuable service. Without their presence, there would not been anyone embedded into the market design which could ensure to an investor that he will be able to trade immediately in the market. MMs indeed solve the structural problem that time of arrival of buy orders and sell ones are not always matched.

We can identify two types of MMs:

- Contractual market makers (CMMs)
- Non contractual market makers (NCMMs)

Contractual market makers are also known as order flow internalizers. They often have economic and contractual obligations with exchange or trading platforms where they operate. In particular, they are obliged to meet some price, volume and depth requirements. These rules can change according to security, geography or regulatory framework. CMMs engaged in valuable economic relationships with brokerage firms: they pay to get into business with these players and guarantee their clients' order execution; from them, they receive privileged access to clients' order flow. [34] Despite this legal and economic relationship, CMMs are not always obliged to make both sides of the market. On the other hand, **NCMMs** are often free from obligations and can therefore provide liquidity with more flexibility.

Market making is a costly business which involves *several costs and risks*. **Costs** are related with order processing, market monitoring, data access and managing clients and providers' relationships. **Inventory risk and information asymmetry** represents the two main risks. An efficient management of inventory allows a MMs to avoid excessive processing costs, incur into losses during market trends and ensure he will be able to transact against other players. Since MMs often rely on

[33] To understand their role, K Narang (2013, pag 266) provides a useful analogy from daily life. MMs can be indeed compared to distributors of goods.
[34] As we will see later in Chapter 3, privileged access to order flow has been a hot topic for its operative implications.

passive orders for their activity, they risk being adversely selected by informed traders. Being fast in submitting, executing and cancelling order is of paramount importance to not incur into severe losses. [35]This is true both for CMMs, which internalizes most of the order flow (sometimes even through Dark Pools), and NCMMs, which operate primarily on lit exchanges interacting with professional investors, which are likely to be informed traders.

Those costs and risks are internalized by MMs through constantly adjusting **bid-ask spread** to new levels: movements in a MM's quoted bid-ask spread reflect the risks which he is bearing with his inventories. These risks and costs obviously influence optimal trading period for MM activity. Optimal market making conditions depends indeed on knowing how to model *temporary market imbalances*, which are the result of differences between how an individual players trades and how a dealer internalize his order flow. To understand where those imbalances lie in, we need to study the **Limit Order Book (LBO) 's dynamics**. The assumption is that LBO's shape contains precious information that can be extrapolated through market microstructure models, based on ticker-tape data, whose final aim is to study the price formation process. HFT heavily relies on market making activity to capture consistent profits. According to Aldridge (2009, pag. 127), trading on market microstructure is in fact *"the holy grail of HFT traders*".

The LBO' shape is determined by:

- its *breadth* (length), which summarized all the quoted prices for a given security (despite some of them are not visible to market operators)[36];
- *depth* (height), which indicates the amount of security offered or required for a given price level;
- asymmetry level, which provides information about current market driving

[35] From these prerequisites, it is clear how market making is not a suitable activity for a human with no help of technology.
[36] Only a limited numbers of price levels (ex: the best 5 bids and asks for US equity market) are visible. This does not mean that the ones observed are the only bids and ask quoted. A trader can see all the other quotes as soon as the market will move towards these other directions.

force.

Once this information has been identified[37], they are incorporated into market models[38] to exploit any profitable opportunity.[39] All these quantities are influenced by the probability distribution of market orders' arrival. Therefore, a trader should pay attention to:

- frequency of new bid quotes
- frequency of new ask quotes
- latest changes in frequency of new bid quotes
- latest changes in frequency of new ask quotes
- relative risk aversion of the trader
- traders' reservation prices

All these variables are influenced by the kind of traders who are acting in the market. According to Harris's classification (Harris, 1998) , we can identify three different kind of actors:

1. **informed traders**, who possess material information about impending market move.
   They are likely to be HFT money managers, institutional investors or other proprietary traders. Because they would like to profit from their informational advantage, they are likely to use limit orders close to market consensus or market order.

2. **value oriented traders**, who play the market according to the estimate provided by their proprietary fundamental valuation models. They use limit orders which are generally far away from the current market prices.

[37] Study of these parameters can delivery important hints on how the market is about to evolve. For example, depth can help to forecast future volatility of an asset. The more volatile the valuation of a traded asset, the sparse would be their orders and the lower would be the quantity of asset available for each level of the LBO.
[38] We will see an example of this application in "2.3 Machine Learning and HFT".
[39] More detailed information on dynamic optimizing techniques from HFT can be found in Ait- Sahalia and Saglam (2013).

3. **uninformed traders**, who aim to profit from providing liquidity and by following short-term price momentum. They post passive orders to provide liquidity and fill in the remaining side of the broad trading order spectrum.

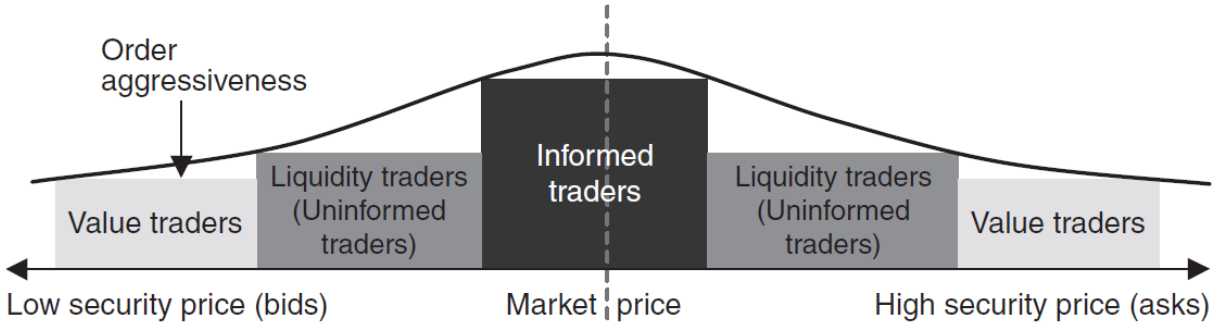

Figure 14 - Graphical representation of order aggressiveness and trader type distributions in the limit order book. Source: Aldridge (2009, pag 147).

To transform limit order book information in valuable trading signals, MMs can rely on two categories of models presented by Lyons (2001):

- Inventory models
- Information models

These models aim to explain how order flows' information produces price changes in two different cases: in the absence of news, for inventory models; when news hit the market, for information models.

Thanks to their market monitoring capabilities, execution speed, highly quantitative modeling skills and data storage and manage capabilities, HFT is then perfectly tailored for market making. Moreover, HFTs firms frequently engage into "*inter-market arbitrage*" (Easley et al, 2013, pag xix) , which consists into offering this service simultaneously on different markets and asset classes.

2.3 HFT AND MACHINE LEARNING APPLICATION

In the eighties, the data available for financial practitioners were simply a collection of daily closing, opening, minimum and high prices and volume traded. Data analysis tools were limited, expensive and complicated: among the most popular machines employed in the financial industry, there was Alpha DEC, a gigantic object whose cost was $20 million US dollar and which could take more than one day to perform a Montecarlo simulation on a single financial security (Aldridge and Krawciw, 2017). Those times are now gone and the whole financial industry today relies heavily on data to perform scenario analysis, forecast future price movements and allocate capital, understand macroeconomic framework, take lending decisions and so on. In summary, to **find alpha and hedge its risks** with new methodologies, using terabytes of data looking even at the wildest variables.

Data landscape has significantly changed, *on a qualitative and quantitative perspective*. Data are available in real time and at the highest frequency which has ever been possible, making real-time data availability a commoditized service. Sources of data are also different: financial analyst do not rely anymore only on market data, as stock prices and fundamental related variables as earnings, but also on "alternative" variables. As explained by Kolanovic (2017), data generated by individuals (ex: search trends), data generated by business processes (ex: commercial transaction) and data generated by sensors (ex: satellite image data) are fully scrutinized in the financial universe today. Therefore, the amount of data used by practitioners has significantly grown: Aldridge and Krawciw (2017) estimated that data related only to one day of trading from a single exchange can require 10 GB of storage space. Then, when we take into account also the aforementioned alternative sources, *dimensionality* reaches unprecedented historical level. Cloud spaces have then become of fundamental importance for analysts to storage data.

Store data is the first step of the increasingly complex data analysis process. A lot of data comes into *unstructured* form and then they need a first level of analysis and several adjustments before they can be effectively used for trading purpose. The

tick-by-tick quotes of a Limit Order Book and web contents represent the perfect example of unstructured data used for financial applications in the world of High Frequency Trading. These large datasets are analyzed with modern machine learning techniques.

Machine Learning (ML) is a subfield of computer science which uses sophisticated mathematical models based on statistical principle to produce forecasts. The singular aspect of machine learning is that algorithms learn by themselves how to proceed and implement process, once they are programmed to implement this task. Machine learning techniques span from *supervised and unsupervised* learning methods to *Reinforcement* and *Deep Learning*. **Supervised learning** methods use two datasets with the aim of finding and explaining the underlying relationship. **Unsupervised methods** are instead focus on the structure of the dataset and their scope is to identify its building blocks. These two represent method of classical machine learning, which are considered as advanced statistical methods. **Reinforcement Learning** (Sutton, 1998) instead is "*a learning method where the learner is not told which actions to take, but instead must discover which actions yield the most reward by trying them while he maximizes a numerical reward signal*". **Deep Learning** is "*a method to analyze data by passing it through multiple layers of non-linear processing units – neurons*" (Kolanovic and Krishnamacahari, 2017, pag.19).

We are going to focus our attention on the last two methods by exploring some of their applications related to HFT. In particular, we are going to consider how reinforcement learning can be applied to predict price movements from order book state and how it could help to optimize trade execution, by analyzing suggestion proposed by Kearns and Nevmyvaka (2013). Then, thanks to Ganesh and Rakheja (2018) study we will discuss deep neural networks applications.

For their applications, Kearns and Nevmyvaka (2013) rely on market microstructure data obtained directly from exchange data feed: orders placed, orders executed, and orders cancelled. Through a reverse engineering process, all these data contribute to reconstruct the shape of the LBO. In reinforcement learning applications, they define the **state** with respect to which an agent implements his

actions. Further refinements[40] can also be implemented, to obtain a more informative and better described action state. Successful reinforcement learning applications indeed depend on how well-defined the action state is. Kearns and Nevmyvaka (2013) highlight one drawback of Reinforcement Learning applications to HFT with respect to other predictive well-rooted models in finance. There is indeed no sufficient historical evidence regarding which variables should be included into state's definition of this kind of applications. This is due to youth of both market microstructure and HFT knowledge domains, but also to proprietary nature of research on this topic which is carried out internally by financial trading firms. For these reasons, they started by building a very simple models and then they also considered additional features and the research advancements produced by these.

Conditional on selected state space defined, several strategies (i.e collection of intertemporal actions) are tested with the aim of identifying the optimal learning method. As observed by the authors, the advantage provided by machine learning techniques consist in detecting on a **a per-stock basis** optimal procedure rather than finding innovative strategies which have never been implemented before. Moreover, the authors also pointed out two results of paramount importance: **market microstructure predictability last only for short-term period** and this time window is often not enough to capture profitable opportunities because of **transaction cost' levels**, which then ends up reducing trading convenience. Therefore, what is needed is a better understanding of state space's description, because only through an improved knowledge of undergoing dynamics it will be possible to identify and exploit really profitable opportunities.

Ganesh and Rakeja (2018) also focus their attention on predicting future price movements behavior and design a trading system which trade on these predictions, realized on the basis of the current LBO shape. They also rely on historical trading data (i.e previous day complete tick by tick data) but instead implement a deep learning model to pursue their goal. A further refinement, which overcome the

[40] In this case, the authors tested also bid-ask spread, bid-ask volume imbalance, signed transaction volume, immediate market order cost

trading costs' problem already illustrated, is that they focus "*on predicting only those price movements which are substantial enough to cross the bid-ask spread*" (Ganesh and Rakeja, 2018, pag 2). **Deep learning** model use neural network models, which are inspired by the working of human brain. In particular, we can observe a collection of nodes which form a network of multiple levels, where each node represents a neuron, designed as a non-linear processing unit. Neurons are indeed trained through a learning dataset, whose final result is computing model weights which play a fundamental role for final predictions. The authors employ a Multi-Layer Perceptron (MLP) model[41]. The main innovative feature of this study is represented by the usage of online data related to the ongoing trading process in the market, which then enables a continue refinement of model's weights on the basis of current dynamics, improving significantly the final predictions.

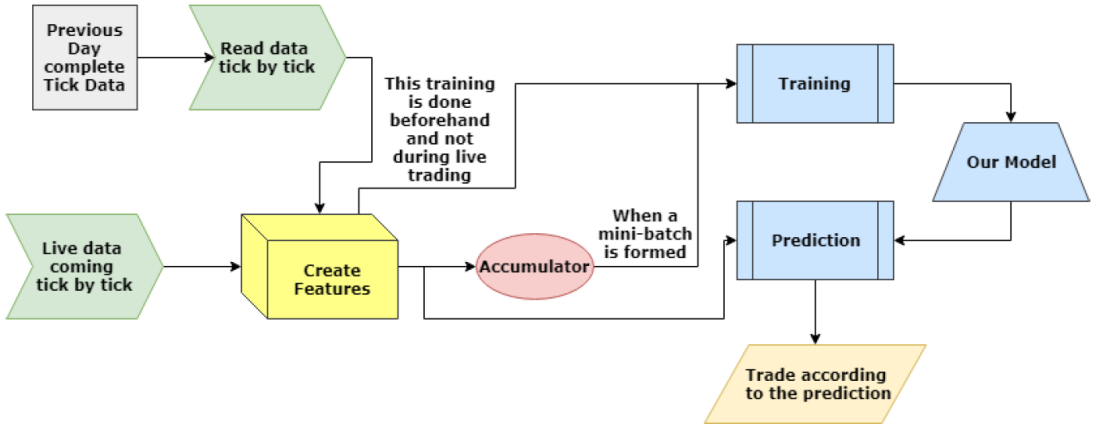

Figure 15 - Trading system model designed employing deep learning model. Source: Ganesh and Rakheja, 2018, pag 3.

The discussion of these 3 models enable us to point out the main features of working with high frequency trading data in a machine learning framework. This task is challenging because of granularity of data, which are produced continuously by the trading process on a fraction of a second- scale; high frequency trading is not yet a well-understood topic, therefore there are not enough evidences

[41] A model such that "input signal passes through each node of the network only once and they are also known as "feed-forward" network" (Kolanovic and Krishnamachari, 2017, pag.19)

regarding which variables are the most significant; transaction costs frequently overcome the discovered potential profitable opportunities. Despite all these challenges, machine learning models seem to exhibit valuable insights into HFT's applications and therefore clearly deserve attention by people interested in knowing more about HFT and discovering profitable opportunities hidden in the markets' most obscure folds.

CHAPTER 3 – DEBATE ON HIGH FREQUENCY TRADING

After having discussed the main features of High Frequency Trading and the most common trading strategies implemented by this type of traders, we need to understand the impact that this new trading methodology has determined on the market. High Frequency Trading today is widely diffused across most of asset classes and the whole market infrastructure has evolved to accommodate these traders' needs. Several services offered by market players and technology providers have been specifically tailored according to the speed desired by High Frequency Traders. Nowadays High Frequency Trading plays a fundamental role into the markets and this process cannot be reverse anymore.

During the previous years, several news and scandals inflamed the HFT's debate. During 2008, the year of the worst financial crisis in the world history, HFT firms were among the few players in the market who recorded positive results. Some performances were even quite impressive, with double digits returns, despite the whole collapse in the U.S stock market. Then it was the time of flash orders scandals. Finally, on May 6, 2010, the Flash Crash made the markets, financial experts, technologists and regulators panicking and people started thinking that a new financial crisis was coming again.

Despite these events had shaded lights on grey area of HFT, a fruitful debate requires a rational and agnostic comparison between benefits and disadvantages. Because of the whole general framework, which includes technological, regulatory and operative adjustment, it is not easy to isolate the role that HFT has played in a market which has gone through a lot of different processes. But by reasoning on its specific operative features, by relying on some empirical results and insider observations, we can start to assess what has meant, is meaning and will mean HFT for financial markets.

HFT has profoundly impacted market quality, contributing to its improvement. By providing more liquidity, HFT have reduced transaction costs and increased market volume. At the same time, they also contributed to change the nature of liquidity

itself, which has become less stable, contingent on temporary trading imbalance and vulnerable to intense quote cancelling. Because arbitrage opportunities are erased quicker than they have always been and market microstructure information as well is impounded into price, price discovery has benefited from HFT's activity. Volatility, on average, does not show any significant improvement following the rising market share conquered by these traders. But because of endogeneity problem, results about this variable need to be treated with cautious. Moreover, illiquidity events as Flash Crash, send a warning sign against relying too much on average result. Finally, HFT contributed to crowd out from the market old-style traditional market makers and less technological sophisticated players who did not embrace algorithmic revolution in time.

3.1 HFT BENEFITS AND CONTROVERSIES

The main benefits of High Frequency Trading are:

- Reduction of transaction costs, increase in liquidity and market volume
- Increase of market efficiency, through a better price discovery mechanism
- Less vulnerability to human panic and reduction in human intermediation

To stimulate better liquidity provisions, two regulatory interventions from SEC that we already discussed had played a fundamental role. **Decimalization** (2001) has indeed reduced the minimum tick size increment to 0.01$, fostering competition for price improvement and allowing for better quotations from Market Makers. **Reg NMS 2005** has been designed with the aim to increase competition among different exchanges and incentivize access to the market. Then, a structural change in the trading model introduced by exchanges, such as the "**maker – taker**" rule[42], has surely contributed to incentivize market making activity increasing market volume. The quantitative competition - more market players operating in the same segment – has increased **market breadth**. More levels of the Limit Order Book (LBO) are now filled with orders from market makers. Qualitative competition –

[42] Liquidity takers are charged with a fee for each of their transaction, while liquidity providers are rewarded with a rebate for their service.

market operators bargaining for lowering prices - has instead **reduced the bid-ask spread**. LBO's consecutive levels are now separated by smaller and smaller distances. Moreover, as observed by Meknveld (2016), HFT reduce the amount of time it takes to spread to be reduced to lower level when it has been previously widen by provisional trading imbalances. Technological race for improving market connectivity and reducing latency time have also further boosted competition **reducing transaction costs**. Within this framework, where technology is crucial for success and spreads are so tight, HFT players enjoy a powerful advantage[43] in providing liquidity and they have indeed gradually conquered more market share. The end results of these profound changes have been **declining transaction costs**, **more market volume and lower bid-ask spread**. Both has been warmly and positively welcomed by investors: retail and institutional ones.

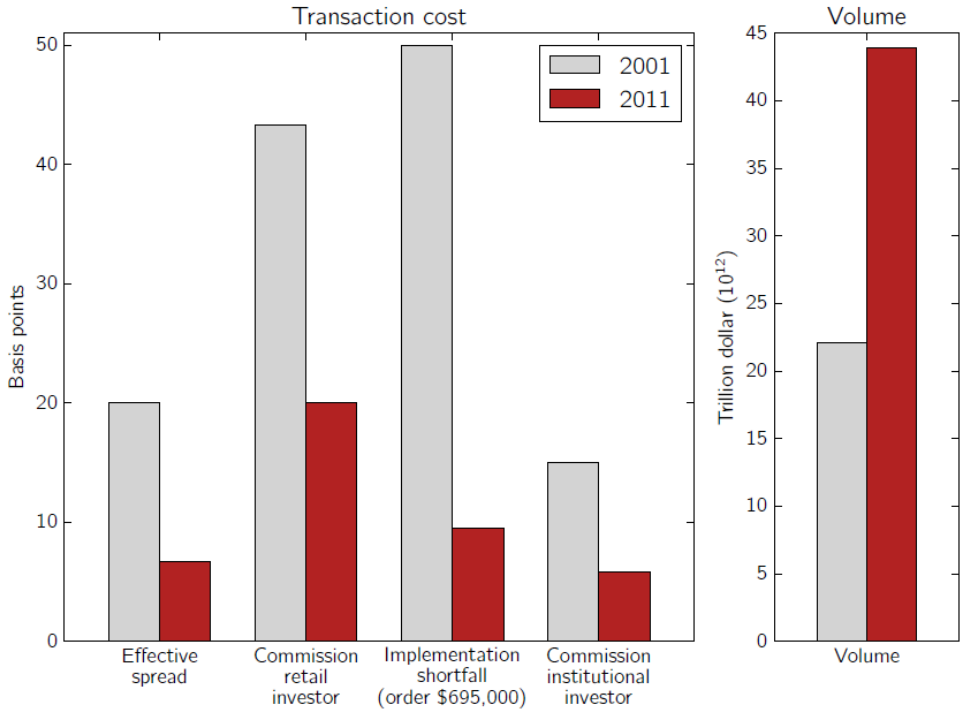

Figure 16- Reduction in transaction costs and increase in volume in US equity market. Source: Menkveld (2016), pag. 4

[43] Because of superior market monitoring capabilities, immediate access to market which enable them to capture even the smallest spread, lower legal obligations.

The time period under analysis allows to compare a pre-HFT's diffusion market configuration (2001) with one where HFT plays instead a significant role (2011). Other changes have also taken place, as decimalization and better algorithmic execution, which have certainly contributed to a reduction in transaction costs, but an increase in volume is rather mainly due to HFT's operations. [44]

Durbin (2010) includes **price consistency** among benefits provided by HFT activity. Despite fragmentation, which has increased the opportunity of price discrepancies, thanks to both HFT arbitrageurs traders' operations and NBBO trade through rule, a market operator or a broker does not need to bother anymore with checking the best venue for execution. According to regulatory provisions, it is ensured that each order is executed at the best national price available across all the different exchanges (which is, exactly, the NBBO). On the other hand, arbitrageurs and pair trading predictors ensure that prices of highly correlated securities move in sync. Search costs connected with trading have then declined, as observed by Angel et al (2015). Moreover, Foucault and Menkveld (2008) point out that in this new configuration the cost of trading is actually transferred to venue platforms and liquidity providers who bargain for providing the best services at lowest price to traders. With respect to the case study ES-SPY under analysis, we can observe that in 2005 arbitrage opportunities were absorbed on a longer period, with several opportunities which lasted for more than 50 milliseconds. In 2011, the bulk of distribution was instead concentrated on a time period inferior to 20 milliseconds. Budish et al (2015) observed that HFT **do not eliminate arbitrage opportunities**, but they instead only reduce the amount of time that they are available. This severe reduction is indeed explained by looking at growing importance of HFT market maker into this business and their ferocious competition into this activity.

[44] Volume can be considered as a proxy of liquidity: an increase in volume therefore could lead us to assert that in certain occasions liquidity has benefited as well. However, when we will talk about "phantom liquidity" we will also explain how in certain occasions this relationship is less representative of real market liquidity conditions.

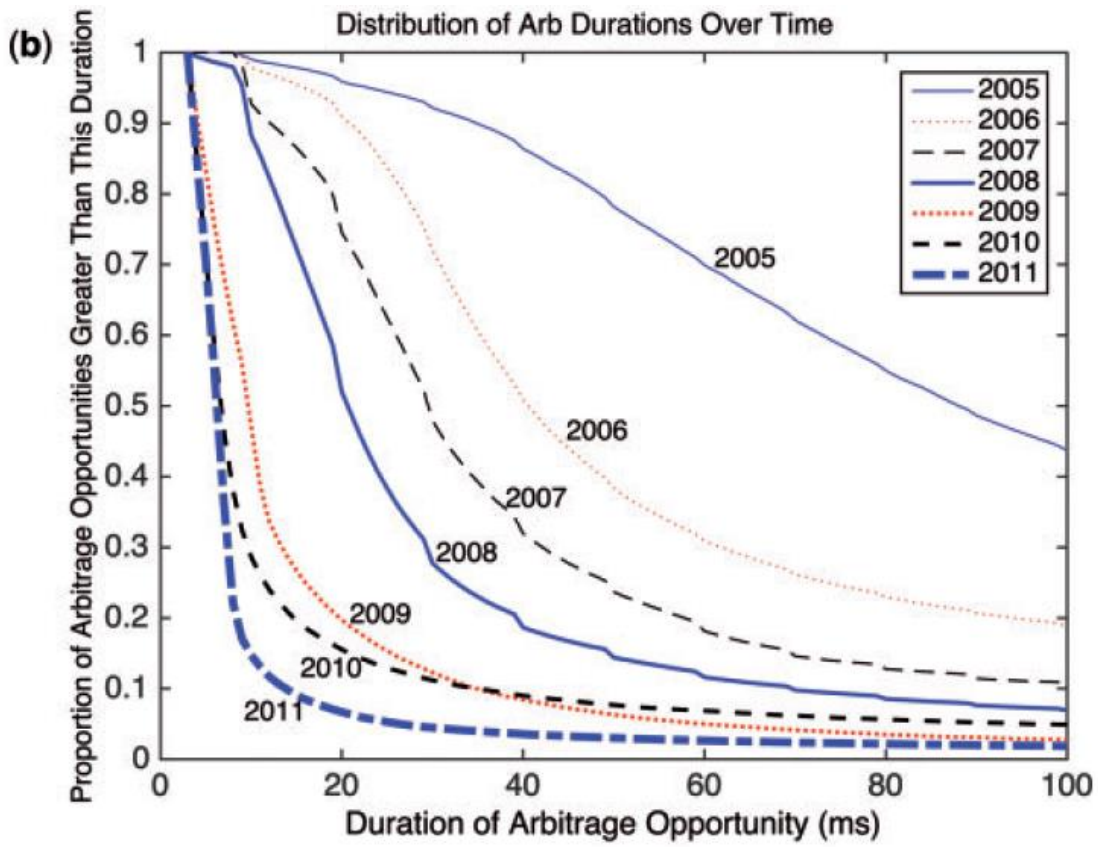

Figure 17 - Distribution of arbitrage durations over time, 2005-2011.
Case: ES & SPY , 2005 -2011. Source: Budish, Cramton, Shim, 2015, pag 24.

HFT's activity has also contributed to improve the **price discovery mechanism**. Thanks to trading speed, most updated technology and sophisticated trading strategies, these traders update market prices as soon as a new piece of valuable information hit the market. As observed by Easley et al (2013, pag 208), "*market now reflect information more quickly than in the past, resulting in prices more accurately reflecting underlying asset values*". As we discussed in Chapter 2, we are not talking only about macroeconomic or fundamental information, but also the ones related with order flow. As explained by Ait-Sahalia and Saglam (2014), thanks to their monitoring capability, HFT traders can forecast from current LBO's state future market dynamics and anticipate them incorporating this information in their quotes. Because they trade expecting an informational advantage coming from their analysis, they would trade more aggressively providing more liquidity. Through this mechanism, market prices are adjusted to new levels more quickly than in the past.

Finally, **computer do not panic** as humans. High Frequency trading does not depend on emotion: this means that it cannot create panicking situation from the scratch.[45] HFT contributed to reduce transaction costs even through reducing need for human intermediation into several practices, market making in particular. Within this field in fact, because their superior speed, monitoring skills and modeling skills, they contribute to **crowd out from the market** less efficient and low sophisticated traditional human market makers (Abergel et al, 2012).

Concerns regarding High Frequency Trading span from elements related to market quality to aspects related with the so-called "dark side" of High Frequency Trading. In particular, we need to discuss the following arguments:

- Market efficiency
- Volatility
- Rogue trading
- Predatory strategies

Thanks to their fast connection and sophisticated quantitative models, HFT traders can profitably trade on the latest news available which can influence the markets. Some news can be related to stocks' fundamental: launch of a new product, which could significantly increase revenues; a breakthrough change in a consolidated dividend policy, and so on. Some other news can affect the industry as a whole rather than a single specific stock, as announcement regarding regulatory interventions or instead a sudden change in a commodity's price. Despite the news' source, Zhang (2010) observed that HFT can provoke overreaction to fundamental news, "**making the market too efficient**" (a phenomenon known also as "*fundamental overshooting*"). In particular, according to this study, the trading process of HFT traders move the market in the right direction but with an excessive magnitude, requiring subsequent further corrections. Sornette and Von Der Becke (2011, pag 7) provide useful insights to understand this process: the authors observe that actually market prices diverge from fundamentals because of **equilibria imbalances**

[45] We will indeed see the role played by HFT during Flash Crash.

between fundamental and momentum traders' activity. When the latter abound, their trading exacerbate price swings. As soon as the news is released, because of their technological advantage, HFT react to it and move the trading process and the market in the direction to digest it. Other High Frequency Traders, observing the market movements, try to join the process to gain profit as well and amplify market movements trading *on the momentum*. Then also institutional investors react to information and trade for adjusting their own portfolio. Their reaction, because of slower technology, frequently happen when prices have already moved. Therefore, it is a combination of different market reactions from heterogeneous players who operate on different time scales, which actually shape the whole process.

The price drop in ULTA, the cosmetic retailer, is a useful example to fully understand the mechanics behind this process (Gregoriou,2015). Moreover, it also offers us the opportunity to discuss an **unfair advantage** from which High Frequency Traders benefited on that occasion, which surely deserve to be scrutinized by regulators.
On December 5, 2013, ULTA missed analysts' earnings estimates. This information became accessible to HFT firms 50 milliseconds after the announcement, which took place 150 milliseconds after 4:00 p.m. Business Wire, a financial information provider, sold *privileged access* to this information to HFT firms. It indeed took other 92 milliseconds before the same information was then accessible to the general public, when Bloomberg released it through its channels. It is useful to specify that Bloomberg bought this information from the same providers, BusinessWire. Despite this procedure is not illegal according to the current standards, since every market operator is allowed to buy access to information, when investors and traders do not enjoy the same fair access to information, competition is altered.

Following the diffusion of algorithmic trading, which as we saw has enabled a better order management process, large orders from institutional investors have become less likely. Despite this, it is still possible to identify their presence into the market. Cutting edge technology and sophisticated data analysis systems enable traders to spot these orders and profit from their presence. Through **pinging,** a practice implemented by submitting a huge number of small orders (on average 100-200 shares), HFT traders try to identify hidden large size orders. The underlying

assumption is that when the order is immediately matched, there is a huge probability that on the other side of the LBO there should be a large order. This practice is frequently used even in dark pools, where it is even riskier given absence of transparency (pre-trade and post-trade). However, this knowledge can be used to trade against the institutional investors with the aim of moving the market in a direction which is averse to these players and profit from their demand. There is a lot of debate about whether pinging is illegal or not. From a regulatory point of view, what matters is that pinging is difficult to detect and prove.

On a liquidity perspective instead, we need to discuss a phenomenon that has been defined by analyst as **phantom Liquidity** (Blocher et al,2019)**.**[46] This expression identifies a situation when the observed market liquidity conditions are not representative of the real ones that a trader face as soon as he tries to trade in the market. It is then frequently associated with market movements which take place as soon as the trader inserts the quote to open or close its position. Such perception can be detected in at least 4 situations, where we can reasonably believe that HFT players contribute to determine this market imbalance. In particular:

1. Once an institutional investor's order is detected, market players react by withdrawing liquidity through cancelling their previous orders to update their quotes to the new market framework.

2. Dark pool's growing influence in liquidity dynamics, making lit liquidity not fully representative of the real demand and supply for trade a given security;

3. Lack of market depth at / near inside quote, which makes the bid-ask spread unrepresentative of the current real Limit Order Book situation.

4. Hidden orders or iceberg orders, which become visible only once the market has move in their direction.

[46] Alternative definitions are also "vanishing liquidity" or "illusory liquidity".

Because of their operative features, HFT trading firms can quickly cancel and update their quotes to new levels. Despite Narang (2013) and other analysts observe how this real time updating process could be actually positive for price discovery mechanism, when observed liquidity is not representative of real costs to bear for trading this situation could **disincentivize willingness of market operators to take risks** (Lewis, 2014). The growing market share of dark pools and increasing reliance on iceberg and hidden orders suggest that some market operators have reacted to phantom liquidity by leveraging on these tools for trying to obtain better execution (Banks, 2014). Therefore, despite HFT has increased market liquidity, when its activity for a given security become predominant with respect to the one of other kind of traders, this could provoke such **detrimental effect on the quality of this liquidity**. As observed into Nanex report (2013), investors are now trying to overcome this problem by relying on market access provided by expensive technology or brokers with quick connection.

One of the harshest critiques moved to HFT is that these players too frequently **trade for the trading's sake,** without providing any valuable service. This idea moves from some simple observations related with market structure features and trading activity patterns. In particular, critics point at the role of HFT traders as market makers. Moving inventories quickly is crucial for hedging adverse selection risk and avoid managing too many positions simultaneously. As soon as these operations are conducted by providing liquidity to investors, a high turnover of capital seems to be justified. However, when HFT *MM trade only for earning rebate* offered by exchanges, exploiting or igniting a trade for their own self-purposes, they **do not provide any valuable service to other players and waste market resources.**

To properly debate the relationship between High Frequency Trading and Volatility, we need first to take into account the related "**endogeneity issue"**. As explained in Chapter 2, HFT traders love volatility and they need a certain degree of price variability to trade in and out of a security in a profitable way (Aldridge, 2009). It is then easy to observe HFT and high volatility together, because these players benefit from this market condition. It is then difficult to disentangle a clear relationship between them, since it could happen that HFT enters into the market when they are

already volatile, rather than contributing to determine volatility itself. On the other hand, as we already observed, HFT firms mainly compete among their peers. Their business model, according to which they operate on fraction of a second, force them to interact mainly with similar players. From their interaction, occasionally it would then seem that "*that market price moves much more rapidly and farther than it otherwise would*" (Durbin,2010, pag.178).

These two premises are of fundamental importance to address this complex topic. Up to now, results are indeed controversial. Some empirical studies have asserted that High Frequency Trading has proved to have **a positive impact reducing volatility**. By working on data from NASDAQ, and comparing actual price with the ones obtained under the assumption of no HFT's trading activity, Brogaard (2010) found that only one out of 120 stocks would have not benefited in volatility reduction from HFT activity. Operating in a market with increasing liquidity, traders can place orders without moving the market significantly avoiding generating any significant market disruption. Moreover, starting from the observations that high-frequency trading do not carry overnight positions and then are only active during intra-day market period, Narang (2013) observes how growing influence of HFT market is not correlated with an increasing intra-day volatility. Despite some positive results, tail events that can question these average results have taken place in the market. In particular, we are referring to "**illiquidity event**", which starting from a delimited context quickly escalated into systemic crisis, reverberating on the whole system, triggering an increase in volatility and affecting financial stability. On this topic, O'Hara (2015) interrogates about how it is possible to reconcile "*the apparent evidence of improved market quality with the equally apparent evidence of decreased market stability*?".

O' Hara (2016) explains that HFT markets are a new type of financial markets which exhibit innovative features, which are essentially due to some structural elements related to:

- Nature of liquidity: HFT market makers are inter-market arbitrageurs with no binding obligations for providing liquidity.

- Nature of trading, which has been completely reshaped by electronic, algorithmic and obviously also high frequency trading as well
- The crucial role played by speed into this new market framework, where HFT "ties market together at lightning speed" (O' Hara, 2015)

Also Sornette and Von der Becke (2011) explains that HFT do not only accelerate consolidate market dynamics, by increasing trading speed, but because of the same reason they also introduce completely new ones. The Flash Crash of May 2010 provides us a useful case-study to investigate some of them. On that day, market volatility was already rising because of uncertainty related with Euro crisis,[47] a premise of paramount importance. Withing that framework, a fundamental investor willing to reduce its position in futures markets decided to sell a consistent position in E-mini futures. At the beginning, the selling pressure of this order was absorbed by HFT traders, who immediately enter into the market as counterparties detecting unusual selling pressure thanks to their microstructure models. HFTs traders then **did not start the decline**, indeed their trading systems simply reacted to a market event originated elsewhere by a non-HFT character. But the counterparties of their trading activity were other HFT traders, who were as well interested in unwinding the acquired positions quickly to not carry too much inventories. This trading conditions quickly escalated into a "*hot potato* effect", where the same securities were bought and sold by HFTs to HFTs back and forth (Nanex, 2013). Because of their cross-market market making activity, connected with their operative features, and cross-asset and cross-market correlations, related instead with market configuration (Sornette and Von der Becke, 2011) the original movement into the futures market was quickly extended to the equity market as well (SEC,2010). Their orders to sell, in reaction to the original selling pressure, ignite a "downward bandwagon effect" which quickly became a "race to the bottom" [48] because no more players who were willing to trade as counterparty.

After this step decline, trading was halt for 10 seconds because the market

[47] Greek debt crisis and consequent perspective of disintegration of the EU were severely stressing financial markets during those days.
[48] For some securities, such bottom was even defined as a quote close to zero dollar!

movements activated internal mechanism from exchanges (Kirilenko et al, 2017). During this panicking situations, several securities reached market value completely distorted with respect to their fundamental ones. Several bargain opportunities were then resting on the markets at cheap profitable prices. As soon as trading reopened and these opportunities were detected by High Frequency Traders, the same players who contributed to bring down the market, reacted for capturing these profits flooding the market with buy orders. The "bandwagon effect" was then ignited in the opposite direction and the recovery took place in a few minutes as well. To any observer who would have look at financial data only at the end of the day, **it could have seemed that nothing has ever taken place**. The reality instead was that in less than an hour, several billions of market capitalization were completely wiped out by frenetic selling pressure just before opposite upward movements capitalized a recovery. This immediate restoration is an innovative feature into financial world because there is no precedent in the history where such crash has been reabsorbed in short time.

Flash orders represent another nuance of the grey area of High Frequency Trading. According to Reg NMS, each transaction must be executed at NBBO price[49]. In order to not lose significant shares of their order flows, some exchanges introduced "**flash orders**". Through these orders, exchanges informed their liquidity providers of having received an order which could be satisfied only providing that these players agree to raise their bid or lower their ask and adjust to the current NBBO or become the new one. For a short period of time, equal to 20 or 30 microseconds, designed liquidity providers were then offered the opportunity to visualize the incoming order and react to satisfy its demand. The *original principle* of introducing flash orders was then trying to improve market efficiency. In the end, instead, flash orders favored aggressive **front-running** procedures. Receivers of this information could in fact use it not to improve their prices towards the NBBO, but to trade ahead of exchanges' clients into other trading platforms. [50]

[49] NBBO stands for National Best Bid and Offer rice.
[50] Durbin (2010) explains that actually for certain asset class flash orders have proved to generate a positive impact. For option exchanges, transferring an order when it cannot be internalized is a costly procedure. In these occasions, flash orders have then proved to be an efficient measure.

Moving out from the gray area of HFT, we now need to consider its darkest side. Market manipulative trading strategies are also known as **predatory strategies** because of the dynamics that they produce in the market. Companies which exploit these strategies rely on advanced knowledge of market microstructure's features and trading patterns to produce these sophisticated and adaptive algorithmic strategies. They can then model prompt and tailored reactions to current market trends. Moreover, as HFT firms, they also used sophisticated combination of orders that are frequently submitted and then immediately cancelled. These strategies exploit structural vulnerabilities of exchanges or of the whole markets and superior speed of HFT players.

Identifying the adoption of these strategies is not a straightforward process, because of data access, required trading knowledge and timescale on which they are operated. All these matters also influence regulatory actions which can be feasible implemented to address this problem. There is a whole debate going on about whether these actions should be considered illegal or not. Besides that, what matters is knowing the most popular strategies and understand which kind of consequences they can generate on the market.

We will start by addressing **layering** strategy, which is initialized by a limit order that rests on the market waiting passively to be executed by another counterparty. In the picture below, this order is represented by dots. This order is posted by the predator, who submit it just above the current level that he can observe on the Limit Order Book. The immediate impact of this action is to raise quote bid price. In our framework, the second main character is the pray, who ignores that he will be soon attacked and that he will only unconsciously react to what other traders are creating in the market.

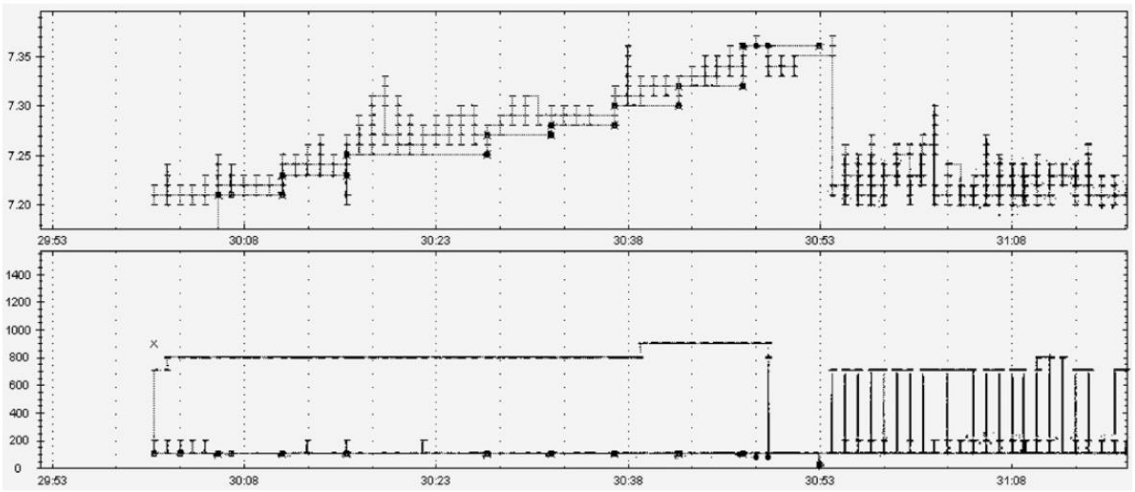

Figure 18 - Example of Layering Predatory Strategy.
In the upper panel, the vertical axis shows prices, and the horizontal axis time.
The dots are limits orders to buy placed by an agency algorithm, while the dashed lines are limit orders to buy placed by the predatory algorithm.
Source: Easley et al, 2013, pag. 212

The second order consequence of the predator's step is in fact igniting the preys' reactions, which are forced into competing by what they are observing on the market. As soon as the preys react, the original limit order to buy is cancelled. The predator then posts another limit order to buy, but this time he immediately cancels it. But not quick enough to avoid that its action could be detected by other high speed market players. However, this is exactly what a predator desires: these orders have to be seen by someone else, its preys, who need to react and continue to move the market. The preys' role in fact is contributing to amplify a market trend which is originally generated by the predator. The preys are fooled to fall in a competition among themselves and the main result of the whole process is increasing the probability of filling the predator's order once he will reach the desired price thanks to other players reaction. When the market reaches the point desired by the predator, he kill the preys by hitting off-setting orders which allow him to exit from the position with a profit. This is not exactly what we observe in the picture above, since the prey has canceled its orders and no trade had taken place. In these occasions, the predator goes back to the market to find another prey and a new race will be ignited.

Traders who are interested in profiting from short-selling positions engage in **bear raiding**. They submit a large sell orders with the aim of triggering a reaction from other market participants, which are supposed to amplify the downward movement that the original order has created. As for layering, to be able to ignite the right reaction and profit from it, a trader needs to study carefully the Limit Order Book and choice the proper moment to submit his order. When a trader instead desires to increase his position in a stock on the buy side, he could **spoof** the market. Because he wants to accumulate a large position on the buy side, he hopes to do it at a cheap price. To pursue his goal he then needs to place a large sell orders to move down the market. Other market participants, scared by the incoming downward pressure, will react by submitting sell orders to exit from their position. A spoofing trader also posts and immediately cancel sell orders to support the decreasing market movement. He also needs to cancel them, before they get executed, to avoid any undesired positions. Once the market reaches the bottom of this spiraling down pressure, he submits a large buy order to accumulate the desired position in the traded security. **Wash sales** is useful for a trader who want to move the price of its securities upward. To ignite an increasing trend, he needs to create interest in the market by adding depth to the order book. This objective can be reached by posting simultaneously buy and sell orders on both side of the LBO to simulate volume. Finally, **quote stuffing** is implemented submitting an enormous amount of "non executable" (far from current bid and offer level) orders which are immediately cancelled. For non-informed traders, monitoring and studying these orders will become a very stressful and time consuming activity. The advantage of the player which starts quote stuffing is indeed knowing that the lion's share of orders is fake. The main goal of this strategy is "create order congestion" for slowing down the stock exchange system causing retards in order processing.

3.2 Regulatory responses

According to Easley et al (2013), we can identify three main regulatory approaches to HFT, which reflect different valuations regarding HFT's usefulness for financial markets:

1. According to the first approach, HFT is a useful technological advance into trading systems and market operativity. It only requires adjustments regarding market surveillance and coordination from regulatory bodies.

2. As for the second view, HFT has distorted markets' functioning, therefore regulatory agencies need to create a new set of rules to restore market operators' confidence into capital markets and bring back to the table a fair competition.

3. Finally, according to a group of reviewers, HFT does not deliver any real benefit to the markets and therefore it should be eliminated or at least significantly reduced.

Despite such a heterogeneity of views, regulators' actions should be inspired by the aim of **preserving benefits** provided by High Frequency Trading and implemented with the **goal of eliminating negative consequences** on the market produced by illegal market practices and unfair adoption of some technological advancements. In addition, to increase the likelihood of effectively address problems, regulators must act together in a **coordinated framework** to rule out the possibility of any regulatory arbitrage. In what follows, we will address some of the regulatory interventions that have been taken or suggested throughout the years with the aim of debating their structure, potential positive effects and real limits. We will also consider some of the struggles that regulators had experienced approaching HFT regulation, both on a theoretical and practical perspective.

Illiquidity events have brought to the public attention that regulators do not dispose of technological tools to monitor the markets in real time. Moreover, given that several glitches into trading firms' technology and exchanges' order processing infrastructure have provoked some markets' breakdown, it is also clear that regulators did not dispose of instruments to ensure that market players employ technological instruments which are not harmful for market activity.

To effectively address these problems, S.E.C has then moved its attention to

solutions which would provide access to real time data regarding market activity. Thanks to collaboration with Tradeworx, an HFT firm, S.E.C now has access to a platform named **MIDAS** (short for Market Information Data Analytics System), which provides to the regulator technological and analytical capabilities to monitor the market using a technology equal to the one used by HFT firms. In particular, MIDAS will collect data on all quotes, trades, orders and cancellation in real time. Popper and Protess (2012) [51] defined this instrument as "*the S.E.C first real-time window into the stock market*". In order to also obtain public and non public information regarding the identities of traders which submitted, cancelled or execute the orders, the S.E.C has also realized a new instrument, **the Consolidated Audit Tape (CAT).** The financial market's picture produced by MIDAS and CAT together is then the most completed possible and it allows the S.E.C to see real-time through the lit markets. [52]

Illiquidity events have also shown that it is of paramount importance to design an appropriate **ex-ante regulatory framework** to reduce probability of market collapse. Opacity regarding HFT practices, more complicated liquidity' dynamics into this new market framework and current market speed makes this need absolutely urgent. There is obviously no time for human regulators to react to events when they happen in fraction of a second. Both US and European Union have already embraced this perspective taking some actions. S.E.C has substituted the long-term standing voluntary supervision of trading firm with a *compulsory* one as prescribed by the new **Regulation SCI**. It requires to each firm to demonstrate on a periodic basis that technology used for trading respect certain standards. The European Commission, through the **MiFid II directive**, requires to trading firms to provide to regulators a fully detailed description of their operations, which ranges from parameters and limits of implemented trading strategies and insights regarding risk management framework. This is clearly a burdensome and really expensive

[51] Some market operators exhibit public concerns regarding how fair could to monitor market with a technology provided by a HFT firm. However, as clearly pointed out by former Tradeoworx CEO Manoj Narang, HFT firms are the only players who possess such technologies so there actually exists no alternatives to provide these tools to regulators.
[52] Data regarding dark pools would be missing anyway because these market platforms do not allow for public disclosure for protecting their clients' activity.

procedure which imposes severe costs on trading firms; moreover, it also requires technologic and operative skills to regulators to oversee all these elements. As observed by Dave Cliff, one of the authors of Foresight Project 2012, on this perspective Mi Fid II "*is conceptually and financial infeasible*".

In a market structure perspective instead, regulators have also discussed the possibility to implement **new circuit breakers** or modifying the ones already in place. Circuit breakers define a "set of acceptable prices for a given security", by establishing a band of prices of a given size centered on the current security's price. As soon as the price moves away from the one identified as acceptable region, circuit breakers are triggered and as a result transactions are suspended. Currently, this stop to trading activity lasts on average for 5 seconds[53], to restore the original price eliminating any excessive market movements which could have produced a significant crash. However, designing circuit breakers is a complicated tasks for several reasons: there is a wide range of circuit breakers, which span from single stocks and basket of stocks to market's segment or exchange as a whole; defining a proper bound range is crucial to ensure that circuit breakers will effectively be activated; this instrument needs to be flexible, therefore market changes in infrastructures and trading patterns must be taken into account into the structuring process. The Flash Crash and other illiquidity events should provide some hints for defining the proper circuit breaker system: a market-wide perspective could eliminate the temptation to send orders to exchanges which have not been yet stopped; different levels should be established for different securities because of their idiosyncratic features; price's band could be dynamically adjusted to markets' conditions.

The second group of critics point out to the following regulatory adjustments to eliminate the inefficiencies created by HFT: *a new minimum tick size, address price discrimination by trading venues, limit on ratio of orders to execution and order transience*. Each of these proposals is motivated by potential gains regarding markets' efficiency. However, as we are going to see, in certain occasions their

[53] 5 seconds could seem a short time, but when compared to execution time of HFT trading firms, it becomes clear that it is a sufficient time.

supporters underestimate the benefits provided by HFT or misunderstand the deep roots of inefficiencies proposing adjustments which would not be really effective and even potentially further disrupting.

The **minimum tick size** can be defined as the smallest possible increment quoted price in a market. The implementation of decimalization in the US, as we described in Chapter 1, established 0.01$ as the minimum tick size for all companies whose traded price is above $1; for penny stock, instead, minimum tick size could go under 1 cent. In EU, instead, there is no specific rule for this market's feature. This creates a complicated situation, because the same stocks can be traded across different trading venues which employ different set of rules increasing the probability of cross-market arbitrage. The minimum tick size has a direct impact on transaction costs and indirect consequences on liquidity and market stability. The higher the minimum tick size, the bigger would be the spread, making the opportunity of taking liquidity more expensive and simultaneously increase profitability of providing liquidity. At the same time, a higher minimum tick size would reduce price competition, reducing number of possible quoted price improvements. As a consequence, we could also observe a decline in cancellation rate which will make price more consistent, a feature which would contribute to strengthen market stability. Despite everything seems to be beneficial within this framework, taking liquidity would become expensive and as a consequence we should expect a reduction in trading volume. In fact, small minimum tick size reduces trading costs and increases price competition, because it makes improving price easier. We would also observe more cancelled orders, because of the need of constantly update quoted price to adjust them to new information in the market. Moreover, the smaller the minimum tick size, the higher would be trading volume because of more profitable opportunities available in the market. What we can reasonably expect is that minimum tick size would not be further reduce into US and maybe adjusted to US level also in Europe. Moreover, we need to highlight that any consideration about modifying the minimum tick size should consider the trade-off that we explained earlier between trading costs and other factors which measures market efficiency (as liquidity costs or market stability).

Price discrimination by trading venues can be observed in two different forms: maker-taker pricing scheme or volume discount. The idea behind the maker-taker pricing scheme is to pay an amount of money (rebate) to operators who provide liquidity, and charge the ones who take liquidity with a parallel amount (fee). Providing a rebate should encourage liquidity supply and increase turnover into Limit Order Book. From a theoretical point of view, it could also stimulate liquidity provision during stressful market conditions. Some trading venues, at their inception, have adopted this organizational scheme to attract market volume by trying to incentivize market operators to trade through their platform. Despite this market design has potential positive applications, we need also to consider any drawback and here we list the following two: maker/taker pricing scheme can foster unfair competition among trading venues and can also impair price discovery mechanism. In particular, as observed by Easley et al (2013) "*broker may opt to send orders to venue offering suboptimal execution in return for rebates are not passed on to the original investors*". Speaking about unfair competition, we need to recall that HFT is a low margin business. A rebate further increments the profitability of such kind of opportunities, rendering some of them profitable only because of the rebate itself. Therefore, High Frequency Traders can engage in transactions accepting worse price than they would theoretically accept reaping off market share from other professional market makers. Moreover, trading for the sake of earning a rebate does not add any information to price discovery mechanism and it is instead a detrimental activity which implies a waste of market resource. To sum up, a maker taker pricing scheme is certainly an interesting scheme that exchanges could adopt to attract market volume, but it needs to be designed properly to avoid unpleasant consequences.

A **minimum resting time** has been proposed to tackle the problem of excessive order cancellation rate. Orders which are submitted and immediately cancelled are detrimental for market liquidity quality because they produce what we called "phantom liquidity". In addition, a market operator cannot trustfully rely on what he observes on Limit Order Book, because due to high cancellation rate he should increase his monitoring market activity to assess LBO's trustworthiness. Some

regulators have then proposed a minimum resting time[54] to address this problem. The potential advantage of this proposal would be to restore market confidence in the Limit Order Book (especially in its highest levels), because both visible depth and quoted market prices would become more representative of actual depth and prices that will be obtained by trading, respectively. On the other hand, a minimum resting time would also increase the cost of providing liquidity. A passive limit order which rests on the LBO waiting for execution has in fact a Free Option Risk (FOR) embedded into it because it offers a certain payoff to active market participants. Establishing a binding minimum resting time period would then increase this FOR. A logical consequence is that market players would be less incentivized to provide liquidity and will also require a higher reward (i.e increase the quoted bid-ask spread) to provide liquidity. This would be most dangerous for the market into high time of volatility, when uncertainty already reduces liquidity providers' willingness to post quotes. Moreover, we need to consider that, with exception of predatory strategy, a high rate of order cancellation is perfectly coherent with market functioning. With markets that move at speed of light, we can expect that both fundamental or microstructure news arrive with extremely high frequency. HFT want to minimize staleness of their quotes to avoid being sniped off and cancelling orders which are not anymore representative of current market conditions is the only way to do it. A minimum resting time, despite some advantage, seems then unfeasible to implement because of negative consequences on liquidity costs.

A high rate of order cancellation is characteristic of HFT's trading activity. Because information flow constantly moves the market, HFT market makers are engaged in cross-market activity and competition is more than ever based on execution speed, HFT traders are frequently forced to cancel their orders and update their quotes. Therefore, among the debated solutions, regulators has also discussed the opportunity to establish **an upper limit on order rate cancellation.** To not impair market making activity and price discovery mechanism, a proper upper limit order rate cancellation should be conceived as sufficiently high, heterogeneous and

[54] It could be dependent on factors such as trade side, volatility or contingent market conditions. Typically a uniform time span is proposed, which is on average 500 milliseconds.

flexible enough to adjust to different type of traded securities and competition between market players who trade such kind of asset.

3.3 WHAT'S NEXT?

In this last section, further proposals which have been advanced to address some of the current market structure problems are considered. In particular, we will focus our attention on the **frequent batch auction model** proposed by Budish et al (2015), the speed bump suggestion proposed by the **dark pool IEX** and remind the importance of **market microstructure models and machine learning techniques** to understand the world of High Frequency Trading and gain valuable insights for its functioning.

As observed by Budish et al (2015), cross-market arbitrage opportunities arise because correlation breaks down at the lowest time scales on which market can operate. If prices of the same asset traded across different venues or cross-correlated securities are certainly consistent (i.e correlation is barely equal to 1) when measured over a day or an hour, this relationship does not hold anymore when markets are analyzed on fractions of a second perspective. Price inconsistencies manifest and they cannot be completely removed away because into the current market design there is no tool which ensure price consistency on a such small period of time. As explained by the authors, the only impact on these arbitrage opportunities produced by High Frequency Traders has been to simply reduce the amount of time they last in the market. This decline has been triggered by a fierce arms race which is actually nothing more than a "*symptom of a flawed market design*" (Budish et al, 2015, pag.1557).

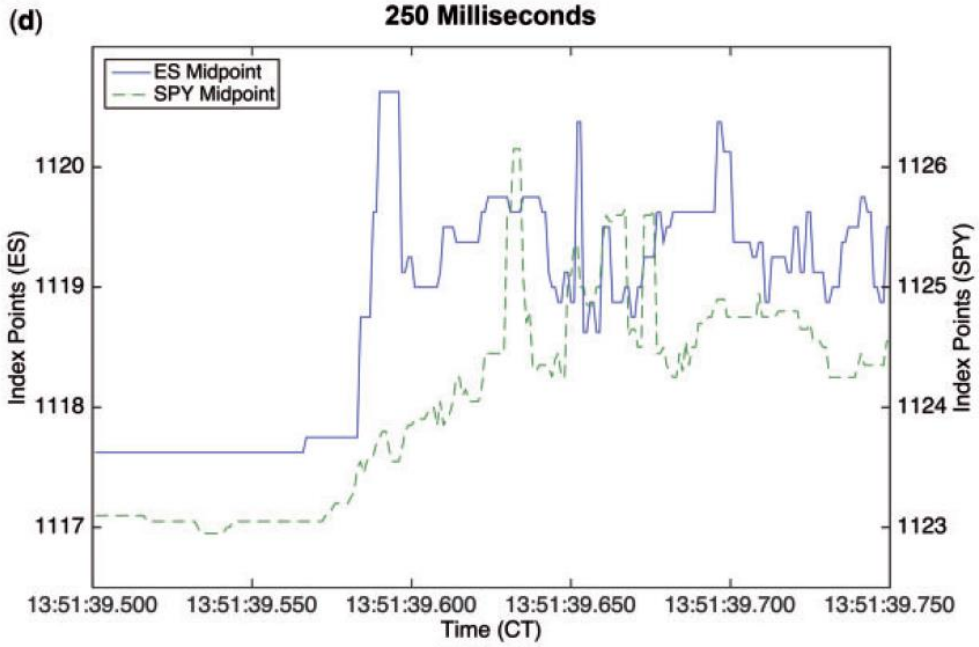

Figure 19 – Correlation breakdown at 250 milliseconds frequency.
Source: Budish et al (2015), pag.5

For this reason, Budish et al (2015) recommend moving from the current continuous trading process centered on a limit order book with price-time priority rules to a market structure based on frequent batch auctions. [55]According to their proposal, the trading day would then be divided into a huge number of small discrete time intervals. All quotes received during a trading window would then be processed as if they were arrived at the same time. The importance of speed would then be reduced to a marginal level and competition could then be focused specifically on price improvements. However, this model implies an epochal transaction from a consolidated historical rooted model as the centralized Limit Order Book and should then take time to be implemented.

Other more-market oriented solutions have also been proposed. IEX, a dark pool founded by Brad Katsuyama, has introduced a "**speed bump**" mechanism which

[55] The full reconstruction of the model is certainly beyond the scope of this presentation. Here will be discussed only the main features and potential results of the suggested changes.

applies to each received quote a pre-determined delay during the processing procedure (Lewis, 2014). Despite it could seem a counterintuitive procedure in a market where speed is of paramount importance, this solution allows the exchange to minimize each unfair advantage which market players would have gained as a consequence of incredibly fast infrastructure technology. IEX's matching engine in fact creates a leveling playing field where orders are treated as if they were arrived at the exchange at the same time. Moreover, in order to avoid privileged access to market data feed, IEX has decided to not offer the opportunity to trading firms to co-locate their servers into the exchange's headquarter. Rebates are prohibited as well to avoid detrimental trading activity carried out for the sake of earning's rebate.

Despite any regulatory or market-based adjustment which could be implemented to tackle some of the unfair advantages enjoyed by HFT firms or drawbacks provoked by their activity, the current understanding of HFT's business certainly needs to be improved. To proceed in these directions, further advancement **into agent-based models, market microstructure models and machine learning techniques** which could potentially replicate HFT strategies should be pursued. Agent-based models allow to study the interactions between market players and predict which would be the likely results of their interplay. Market microstructure models could improve our understanding of liquidity origination process and formalize the behavior of MMs actors to design appropriate policy. Machine learning techniques implemented to replicate the trading patterns of HFT according to different strategies could help to identify which variables play a fundamental role in determining short-term price movements and incorporate them into both agent-based and market microstructure models. Advancements in these directions will also help regulators to fully understand the HFT business and in particular design appropriate ex-ante measures to build a more resilient structure. In addition, this knowledge could then also be used by other market operators to design appropriate reactions to escape HFT's predatory techniques.

CONCLUSION

The current shape of financial markets has been determined by decades of profound and epochal changes. Regulation and technology, as always, have driven this transition. New trading venues have been created, new rules have been introduced and there is a whole new automated electronic infrastructure which underpin the global financial markets. Within this new market design, High Frequency Traders play a fundamental role in providing liquidity, incorporating information into price quickly and eliminating cross-market arbitrage opportunities.

Despite they has brought to the market significant advantages, their activity has also proved to be potentially detrimental for market quality and dangerous for financial stability. Because of cross-assets and cross-markets correlations, small contextualized event can quickly escalate into complex and large impact disruptive dynamics which could wipe out billions of dollars in a fraction of a second. The Flash Crash has been the most important illiquidity event in the last two decades, but similar manifestations have taken place on a smaller and/or individual scale. With respect to all previous crash in the history of finance, the same activity of HFT traders which contribute to propagate downward pressure is also at the basis of following reverting comeback.

The current market regulation clearly needs some updates to face these new dynamics in an efficient way. But regulatory adjustment alone could not fix the problems. We need a better understanding of HFT dynamics and market microstructure models could provide fundamental knowledge in these directions. In the meantime, some structural changes in market configuration have already taken place and they could help to reduce market imbalance and eliminate unfair competitive advantages.

BIBLIOGRAPHY

1. Abergel, Bouchaud, Foucault, Lehalle, Rosenbaum, 2012, "Market Microstructure – Confronting Many Viewpoints", John Wiley & Sons
2. Ait-Sahalia and Saglam, 2013, "High Frequency Traders: Taking advantage of speed", NBER Working Papers, No 19531.
3. Aldridge, 2010, "High Frequency Trading – A practical Guide to algorithmic strategies and Trading Systems", John Wiley & Sons
4. Aldridge and Krawciw, 2017, "What investors should know about Fintech, High Frequency Trading, and Flash Crashes", John Wiley & Sons
5. Angel, Harris, Spat, 2015, "Equity trading in the 21st century: an update", Vol.5, No.1, Quarterly Journal of Finance
6. Arnuk and Saluzzi, 2012, "Broken Markets- How high frequency trading and predatory practices on Wall Street are destroying investor confidence and your portfolio", FT Press
7. Banks, 2014, "Dark Pools – Off exchange liquidity in an era of high frequency, program and algorithmic trading", 2nd edition, Palgrave Macmillan
8. Blocher, Cooper, Seddon, Van Vliet, 2016, "Phantom liquidity and high frequency quoting", Journal of Trading, Volume 11, No.3, 6-15.
9. Brogaard, 2010, "High Frequency Trading and its impact on market quality",
10. Bloomberg, 2011, "High-Frequency Trading is 77% of U.K Market, Tabb Group Says"
11. Budish, Cramton, Shim, 2015, "The High Frequency Trading arms race: frequent batch auctions as a market design response", Volume 130, Issue 4, The Quarterly Journal of Economics
12. Chan, 2013, "Algorithmic Trading: winning strategies and their rationale"
13. Citigroup Transaction Services, 2012, "Market Fragmentation: does it really matter?"
14. Durbin, 2010, "All about High Frequency Trading", Mc Graw Hill
15. Easley, Do Prado, O'Hara, 2013, "High Frequency Trading: New realities for traders, markets and regulators", Risk books
16. Eichenbaum and Evans, 1993, "Some empirical evidences on the effects of monetary policy shocks on exchange rates", NBER Working Paper no 4271.

17. Foucault and Menkveld, 2008, “Competition for order flow and smart order routing systems”, The Journal of Finance, Volume 63, Issue 1, pages 119-158.
18. Foucault and Moinas, 2018, “Is trading fast dangerous?”, Toulouse School of Economics
19. Ganesh and Rakeja, 2018, “Deep Neural Networks in High Frequency Trading”, IEEE Transactions on Neural Network and Learning Systems
20. Gregoriou, 2015, “Handbook of High Frequency Trading”, Elsevier Inc.
21. Grilli and Roubini, 1993, “Liquidity, Capital controls and exchange rate”, Journal of International Money and Finance, volume 12, issue 2, 139-153.
22. Harris, 1998, “Optimal dynamic order submission strategies in some stylized trading problems”
23. Harris, 2003, “Trading & Exchanges – Market microstructure for practitioners”, Oxford University Press
24. Hasbrouck and Saar, 2010, “Low latency Trading”, Working Paper, New York University
25. IBM, 2008, “Tackling latency: the algorithmic arms race”, IBM Global Business Service
26. Kearns and Nevmyvaka, 2013, “Machine Learning for Market microstructure and High frequency trading”
27. Kim, 2007, “Electronic and algorithmic trading technology – The Complete Guide”, Elsevier Inc.
28. Kirilenko, Kyle, Samadi, Tuzun, 2017, “The flash crash: high-frequency trading in an electronic market”, Journal of Finance, volume 72, issue 3, 967-998.
29. Kolanovic and Krishnamachari, 2017, “Big Data and AI Strategies – Machine Learning and Alternative Data approach to investing”, JP Morgan
30. Lehalle and Laruelle, 2018, “Market microstructure in practice”, 2nd edition, World Scientific Publishing
31. Lewis, 2014, “Flash Boys – A Wall Street Revolt”, W.W Norton & Company, Inc.
32. Lyons, 2001, “The microstructure approach to exchange rates”, MIT Press
33. Malone et al., 1989, “The logic of electronic markets”, Harvard Business Review

34. Menkveld, 2016, “The economics of High Frequency Trading: taking stock”, Annual Review of Financial Economics, Vol 8: 1-24.
35. Narang, 2013, “Inside the Black Box – A simple guide to quantitative and high-frequency trading”, 2nd edition, John Wiley & Sons
36. Nimalendran and Ray, 2012, “Information linkages between dark and lit trading venues”
37. Nanex, 2013, “Perspectives on High-Frequency trading”, Zervoudakis, Lawrence, Gontikas, Al Merey.
38. O’Hara, 2015, “High Frequency Market Microstructure”, Volume 116, Issue 2, Journal of Financial Economics
39. S.E.C, 2010, “Finding regardings the market event of May 6,2010”
40. Sornette and Von Der Becke, 2011, “Crashes and High Frequency Trading – An evolution of risks posed by high-speed algorithmic trading”, The future of Computer Trading in Financial Markets – Foresight Driver Review – DR 7
41. Sutton, 1998, “Reinforcement Learning – An introduction”, MIT Press
42. The New York Times, 2018, “Time split to the Nanosecond is precisely what Wall Street wants”
43. The New York Times, 2012, “To regulate rapid traders, S.E.C turns to one of them”
44. Wang and Zheng, 2015, “High Frequency Trading and Probability Theory”, Volume 116, Issue 2
45. Zhang, 2010, “High Frequency Trading, stock volatility and price discovery”, Yale University School of Management.
46. Zubulake and Lee, 2011, “The high frequency game changer”, John Wiley & Sons