Bachelor's degree in Economia e Management

Chair of Financial Market Analysis

High-frequency trading and market transformation: the creation of IEX

Prof. Nicola Borri
Professor

Ilenia Fiore ID 250881
Candidate

Academic Year 2022/2023

TABLE OF CONTENTS

Introduction

In 2014, Michael Lewis, a renowned financial best-selling author, published his book "Flash Boys", shedding light on the rigging of the U.S. equity market and igniting a fierce and ongoing debate regarding the influence of technological innovation on financial markets. Lewis exposed the adverse effects of market digitalization, emphasizing their detachment from market reality, and singled out High-Frequency Trading (HFT) as a particularly guilty practice. The book revealed the disruptive nature of such trading practices, sparking significant scrutiny and discussion within the financial community.

From this book originates the idea behind this dissertation to embark on an in-depth journey to deconstruct the growing complexity of the landscape of trading in the U.S. financial market, driven by technological advancements and the emergence of new trading strategies. This thesis explores the impact that HFT, a practice characterized by rapid and automated trading executed at lightning speed, has on the market. Subsequently, the alleged transformation of the aforementioned market in terms of quality and efficiency will be discussed, before focusing on the ground-breaking project on which *Flash Boys: A Wall Street Revolt*, is also centred: the creation of IEX (Investors Exchange).

The first chapter delves into the trading scene before the rise of HFT, providing an overview of different types of markets. It examines markets that originated as auction markets, such as the New York Stock Exchange (NYSE), and those that originated as electronic exchanges, such as the NASDAQ. Furthermore, a main distinction is made between the basic order types, namely between market and pre-contingent orders, shedding light on the mechanics of trading prior to HFT.

In the second chapter, new trading strategies and markets that have emerged in the U.S. stock market are introduced. This section discusses the rise of algorithmic trading, which forms the foundation of HFT, and explores the obscure reality of dark pools. The chapter also investigates the infamous flash crash of May 6th 2010, analysing its causes and implications, following the in-depth investigations conducted by NANEX[24]. Additionally, it is pointed out how the rise of HFT was also enabled by the regulatory framework. Going into greater depth about what has changed in the institutional rules on financial markets, the role of the Securities and Exchange Commission (SEC) is explored together with important initiatives such as the introduction of the Securities Information Processor (SIP) and the Regulation of the National Market System (Regulation NMS) and Alternative Trading Systems (Regulation of ATS).

Having addressed the general framework, the third chapter then delves into the intricacies of HFT, focusing on its infrastructure, strategies, and hypothetical effects on market dynamics. It elucidates the technical characteristics of HFT, such as low latency, co-location, and the motives underlying the strategic construction of Spread Networks' fiber line. Moreover, the chapter explores various HFT strategies, namely some that purely capitalise on market misalignments, such as slow market arbitrage and liquidity providing, and others more aggressive such as flash trading, liquidity detection and quote stuffing. Furthermore, it examines the alleged impact of HFT on market efficiency, discussing the notion that "the fastest predator takes the fattest prey."

Eventually, the fourth and final chapter is centred on the Investors Exchange, a ground-breaking solution proposed by Brad Katsuyama to address the concerns posed by HFT. First, Brad Katsuyama is introduced, along with an insight into his story and his experiences at the Royal Bank of Canada (RBC). Later, the chapter delves into the unpleasant discovery made by Katsuyama regarding the rigged nature of the market and the increase in "Technical Glitches" he and his team observed when conducting several experiments on public exchanges. THOR mechanism, their initial proposal for slowing down the trading process, is discussed and questions about its sustainability are raised. The chapter then narrates the creation of IEX and its subsequent approval by the SEC as a registered exchange in 2016. It discusses IEX's commitment to creating fairness in the market, illustrated by their transparency on the fee structure, the routing mechanism and the limited types of orders accepted. Consequently, some observations and a statistical analysis of IEX's effects on market efficiency are presented, followed by a discussion of the findings. The chapter also highlights the ethical and strategic change of Goldman Sachs, from the arrest of their programmer Sergey Aleynikov (2009) to the decision to direct orders to the IEX exchange (2013), thus enhancing the latter's popularity. Finally, the chapter touches upon subsequent proposals and provides an overview of IEX's current status.

By examining the evolution of HFT and the transformative journey of IEX, this thesis aims to provide a comprehensive understanding of the impact of HFT on the financial markets and the innovative solutions that have emerged to foster fairness and efficiency. Through a detailed exploration of the topics outlined in the table of contents, this study seeks to contribute to the ongoing discussion surrounding market structure and the future of trading in the digital age.

CHAPTER I

1. The trading scene (before HFT)

The concept of trading these days is defined as the buying and selling of financial instruments, such as stocks, bonds, currencies, and commodities, with the aim of making a profit from the price movements of these assets. This definition is widely used and can be found in various sources, including financial textbooks and online financial dictionaries[1].

The trade of goods and services has been an integral part of human civilisation for centuries. The earliest forms of trading were likely conducted through bartering goods and services, before the development of monetary systems and using precious metals as currency.

As societies advanced, formal markets emerged, where traders gathered in designated areas to buy and sell goods. Among the first recorded markets, there is the *ἀγορά* (ancient Greek for "agora") of Athens in ancient Greece, which was a bustling trade centre of the polis.

In medieval Europe, trade fairs were popular, with merchants traveling from across the continent to sell their goods. As the demand for goods and services grew, so did the need for formalized trading systems. The first stock exchange was established in Amsterdam in 1602, with the formation of the Dutch East India Company. This allowed investors to buy and sell shares of the company, providing a way for them to profit from the company's success.

The 19th and 20th centuries saw the expansion of financial markets, with the establishment of commodity exchanges, foreign exchange markets, and bond markets, with notable examples including the *London Stock Exchange* founded in 1801, the *New York Stock Exchange* formally constituted as the New York Stock and Exchange Board in 1817, and the *Tokyo Stock Exchange* opened in 1878 to provide a market for the trading of government bonds that had been newly issued by the former Samurai. The rise of technology, such as the telegraph and telephone enabled faster and more efficient communication and trading.

The latest technological advancements of the 21st century have revolutionized the trading scene, through electronic trading platforms allowing traders to execute trades faster and more efficiently

---

[1] Trading: "the activity of buying and selling goods and/or services" (*Cambridge Dictionary*).

than in the past. In recent years, the rise of algorithmic trading has further transformed the trading scene, with sophisticated computer algorithms able to analyse market data and execute trades automatically at a speed impossible to attain with human capability. This trend, first initiated in the U.S. stock market, easily expanded to a global scale, resulting in the transformation of the stock market structure and trading techniques.

Before introducing today's complex structure in which the questionable practices discussed in this thesis take place, it is appropriate to take a step back and present the trading scene before the advent of algorithmic trading, i.e., the different types of markets and types of orders on the securities markets.

1.1. Types of markets

The markets on which trades are usually made can take different forms. Whereas the newly issued securities to the public are traded on the primary market, on which the initial public offerings (IPOs) take place, trades in existing securities happen on the so-called secondary market. Four types of markets can be distinguished according to the level of organization inside of them.

The *direct search market* is the least organized market, where buyers and sellers must find each other directly without any intermediary. This type of market is characterized by irregular participation and nonstandard goods and is not beneficial for most individuals or firms to specialize in it.

Moving on to the next level of market organization, a *brokered market* involves brokers who offer research services to buyers and sellers in active trading markets. These brokers possess specialized knowledge of valuing assets traded in these markets. For instance, the primary market, where new securities are offered to the public, is a brokered market where investment bankers act as brokers.

As trading activity in a particular asset type, increases, the need for a *dealer market* emerges. The dealers specialize in various assets and purchase them for their own accounts to later sell them for a profit from their inventory. The *bid price* is the highest price that a buyer is willing to pay for a share of stock or another security. The *ask price* represents the minimum price that a seller is willing to take for the same security.[2] The spreads between dealers' buying and selling prices are a source of profit. Dealer markets save market participants (i.e., traders) on research costs since they can easily look up the prices to buy from or sell to dealers. However, a fair amount of market activity is required so that

[2] See: "Bid Price/Ask Price", *U.S. Securities and Exchange Commission.*

dealing in a market becomes a lucrative source of income. Over-the-counter[3] dealer markets are where most bonds and foreign exchange trades occur.

Finally, the fourth type of market is presented in the following section.

1.1.1. Auction markets

The most highly integrated market is the *auction market* where all traders convene at a centralized location, either physically on a trading floor or on an electronic platform, to buy or sell an asset. Compared to dealer markets, the advantage of auction markets is that buyers do not have to search among various dealers to get the best price for an item. If all participants are present, they can agree on a mutually acceptable price and avoid the bid-ask spread.

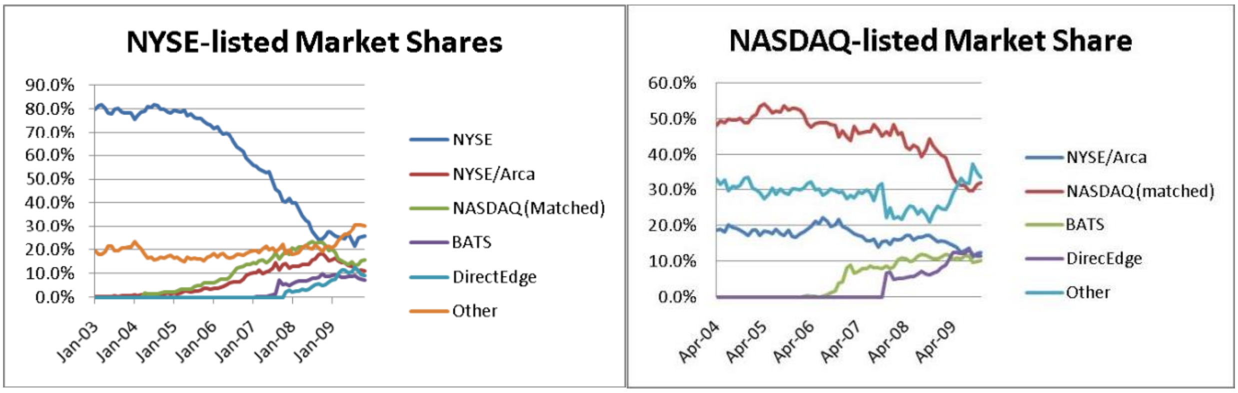
The New York Stock Exchange and NASDAQ, which were primarily a specialist market[4] and an over-the-counter dealer market respectively, are examples of auction markets, where traders come together to buy and sell shares of publicly traded companies through a centralized auction process that determines the market price for the shares. Brokerage firms with access to exchanges sell their services to individuals, by charging commissions for executing trades on their behalf.

It is precisely on stock exchanges that the practices that this thesis aims to analyse fit in. Before the fully automated markets gained market share at the expense of less automated ones, the NYSE and the NASDAQ had the greatest market shares of trading of their listed shares on the U.S. stock markets.

The decrease in the proportion of their market share from 2003 to the end of 2009 can be seen in the charts below (*See Charts 1 & 2*). To fully understand what happened after, it is necessary to focus on the rise of the other lines in the charts, especially on the one labelled “other”.

[3] Over-the-counter market: “an informal network of brokers and dealers who negotiate sales of securities.” *Essentials of investments* (2013), McGraw Hill, 62.
[4] Specialist market: “market in a stock made solely by the specialist, as no public orders, and henceforth no depth, exist in the market” (*Nasdaq.com*).

Charts 1 & 2: NYSE-listed and NASDAQ-listed Market Share



Source*: Equity Trading in the 21st Century(2010); Barclays Capital Equity Research*

Hence, let's have a brief look at the major stock exchanges mentioned above.

1.1.2. NYSE

"The New York Stock Exchange traces its origins to the Buttonwood Agreement signed by 24 stockbrokers on May 17th 1792, as a response to the first financial panic in the young nation. It set rules for how stocks could be traded and established set commissions. The Agreement aimed to promote public confidence in the markets and to ensure that deals were conducted between trusted parties"[5].

The NYSE holds the distinction of being the largest stock exchange in the U.S., measured by the market value of its listed companies. It boasts an impressive daily trading volume of over 3 billion shares. The NYSE Group was formed in 2006 through the merger of the NYSE with the Archipelago Exchange, and later in 2007, the company merged with the European exchange Euronext to become NYSE Euronext. Additionally, the acquisition of the American Stock Exchange in 2008 resulted in the formation of NYSE Amex, which focuses on small firms. NYSE Arca is the electronic communications network of the company. In 2013, Intercontinental Exchange (ICE) acquired NYSE Euronext, and it maintained the same name since.

For a long time, the NYSE relied on its specialist trading system that heavily depended on human participation in trade execution. However, it began the transition to electronic trading for smaller trades in 1976, with the launch of the DOT (Designated Order Turnaround)[6] and later SuperDOT

[5] See: *https://www.nyse.com/history-of-nyse.*
[6] DOT: "Designated order turnaround is an electronic system that increases efficiency by routing orders for listed securities directly to a specialist on the trading floor instead of through a broker" (*Investopedia.com*).

systems. These systems allowed orders to be routed directly to the specialist. The change of emphasis dramatically accelerated in 2006 with the introduction of the NYSE Hybrid Market, which allowed brokers to send orders either for immediate electronic execution or to the specialist, who could seek price improvement from another trader.

1.1.3. NASDAQ

NASDAQ, the *National Association of Securities Dealers Automated Quotations*, is a global electronic marketplace for buying and selling securities, particularly stocks. It was founded in 1971 and is headquartered in New York City. Unlike the NYSE, it operates as an electronic market, with all trades executed electronically. It is known for listing many technology companies (more than 3,000 firms) and is the world's largest electronic stock market. The latest platform, the NASDAQ Market Center, combines all previous electronic markets into a single integrated system. NASDAQ merged with OMX, a Swedish-Finnish company that controls seven Nordic and Baltic stock exchanges, in 2007 to form NASDAQ OMX Group, which also operates various European stock markets and a U.S. options and futures exchange.

There are three levels of subscribers on NASDAQ, with level 3 being the highest. Registered market makers fall under this level and so are companies that make a market in securities, hold inventories of securities, and post bid and ask prices to buy or sell shares. Level 2 subscribers receive all bids and ask for quotes but cannot enter their own quotes. These subscribers are usually brokerage firms that execute trades for clients but do not actively trade stocks for their own accounts. Level 1 subscribers only receive inside quotes and do not see the number of shares being offered. These subscribers tend to be investors who want to stay up to date on current prices but are not actively trading.

1.2. Types of orders

The number and types of orders that investors can place on the stock markets are constantly changing and evolving. The most complex models will be discussed in more detail later in this paper; however, they are based on two basic types, market orders, and pre-contingent orders.

1.2.1. Market orders

Market orders are orders to buy or sell stocks that are executed immediately at the current market price. If an investor wants to purchase shares of Apple, they will contact their broker to inquire about the market price. The broker would then provide the best bid price and ask price, which are $159.73

and $159.79 respectively at the time of this writing. If the investor wants to buy 100 shares "at market", they will purchase the shares at the current ask price of $159.79. On the other hand, if the investor wants to sell shares "at market", they will sell their shares at the current bid price of $159.73.

Table 1: Bid and Ask prices for Apple Inc. stocks on NASDAQ at close: March 24 04:00PM EDT

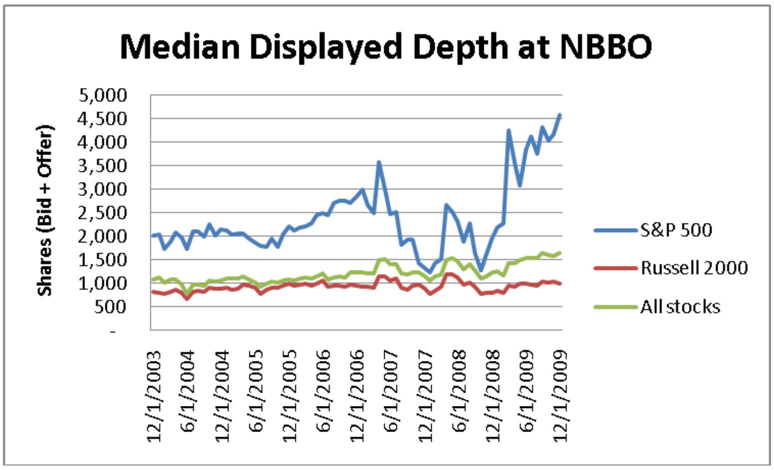
Bid	159.73 x 800
Ask	159.79 x 800

Source: *finance.yahoo.com*

However, this scenario can become more complex. The price quotes provided represent commitments to trade a specific number of shares, so if the investor's order exceeds that number, their order may be filled at multiple prices. For instance, if an order is for 1,000 shares but the quoted price is only valid for up to 800 shares, as in this case, the investor may have to pay a higher price for the remaining 200 shares. The depth of the stock, which is the total number of shares offered for trading at the best bid and ask prices, is another factor that affects liquidity.

Moreover, another trader may execute a trade at a better price before the investor's order arrives, or the best price quote may change before the order is executed, resulting in a different execution price from what was initially quoted. Precisely this is, as will be seen, one of the main aspects that High-Frequency Traders have begun to exploit to their advantage. As can be seen in the graph below (*See Chart 3*), the median depth of the biggest U.S. stocks that can be synthesised by the depth of the S&P 500, is consistently higher than the one of the Russell 2000 (an index comprising of 2,000 small-cap companies).

Chart 3: Median market depth for large (S&P 500) and small (Russell 2000) firms[7]



Source: *Equity Trading in the 21st Century (2010); Barclays Capital Equity Research*

[7] The graph above shows the median number of shares displayed at the National Best Bid and Offer (NBBO).

1.2.2. Pre-contingent orders

The orders that are launched on the exchanges by the investors can also be subject to specific prices that are set as limits when the orders are submitted. Market participants can instruct the broker to execute their buy order of a certain number of shares if and only at a specified price or below. This type of order is called a *limit buy order*. On the other hand, a *limit sell order* is set to be executed by the broker if and only when the stock price rises to a stipulated price or above. The limit order book comprises the totality of limit buy and sell orders that are waiting to be executed on a specific exchange. Let us assume that an investor intends to trade Apple Inc. shares on the Chicago Board Options Markets and decides to place a limit order. Therefore, he first decides to check the latest orders placed on the exchange. At the point of writing, the investor would face the limit order book shown below (*See Table 2*).

Table 2: A portion of the limit order book for Apple INC. on CBOE, March 27th 2023 1:45 EDT

TOP OF BOOK			
BID		ASK	
Shares	Price	Shares	Price
300	158.47	538	158.52
200	158.46	638	158.51
632	158.45	538	158.50
649	158.44	737	158.49
649	158.43	100	158.48

Source: *data elaboration from markets.cboe.com*

CHAPTER II

2. NEW MARKETS & TRADING STRATEGIES IN THE U.S. STOCK MARKET

The contemporary configuration of securities exchanges is primarily electronic. The rise of electronic trading was driven by two major factors, which are technology and regulation.

Due to advancements in technology, such as the internet and the high-speed computing, traders could swiftly compare prices across multiple markets and execute trades electronically and more efficiently. The rise of electronic communication networks (ECNs)[8] that happened at the end of the last century, which enabled traders to bypass traditional exchanges and connect directly with one another, displayed the attractiveness of the speed at which a trade can be executed. By the end of the twentieth century, the traditional exchanges became private businesses and started to specialize in specific sectors (e.g., NASDAQ is nowadays the stock exchange for the whole technological sector).

At the same time, new regulations that will be further discussed in this chapter, allowed the brokers to compete for business and initially broke the hold that dealers once had on information about the best-available bid and ask prices. They would do this all while forcing the integration of markets, making it easier to create an exchange and allowing securities to trade in ever-smaller price increments. As a result, the number of securities exchanges in the U.S. markets now registered with the SEC as national securities exchanges stands at fifteen[9].

This new structure of the exchanges changed the market and pushed even more the dispersion and the competition between new fully electronic markets and traditional exchanges, which is based on the latency of the orders placed on an exchange (*See Chapter 3.1.1.*).

---

[8] "ECNs are computer networks that allow direct trading without the need for market makers. ECNs register with the SEC as broker-dealers and are subject to Regulation ATS (for Alternative Trading System). ECN subscribers, who can bring trades directly to the network, are typically institutional investors, broker-dealers, and market makers. Individual investors must therefore hire a broker who is a participant in the ECN to execute trades on their behalf.
ECNs offer several advantages. Trades are automatically crossed at a modest cost, typically less than a penny per share. ECNs are attractive as well because of the speed with which a trade can be executed. Finally, these systems offer traders considerable anonymity". *Essentials of investments* (2013), McGraw Hill, 63.

[9] Among these are: NYSE Euronext, NASDAQ, The Chicago Board Options Exchange, and BATS Exchange. The complete list of national securities exchanges and recently approved exchange applications is available at https://www.investor.gov/introduction-investing/investing-basics/how-stock-markets-work/market-participants.

Finally, these new elements led to a downward trend in the cost of trade execution, which became just a fraction of what it was a few decades before, hence encouraging the use of more computer-assisted trading techniques.

2.1. Algorithmic trading

With the technological innovations of the late 20th century, it became possible for investors to connect remotely to order books on the market. This very first example of electronic trading later evolved into algorithmic trading.

A complete discussion of algorithmic trading, also known as automated trading or black-box trading, provided by R. Vedapradha[10], states that "algo-trading involves placing a deal using a computer program that adheres to a predetermined set of guidelines" which is indeed an algorithm. It is also pointed out that "the deal can produce profits at a pace and frequency that are beyond the capabilities of a human trader" and that "[it] makes trading more organized by minimizing the influence of human emotions".

The instructions provided to the algorithms can be based on a mathematical model, time, price, quantity, or any other factor, and these different types of algorithms can be used by distinct market participants.

As T. Hendershott[11] et al. recall, some *hedge funds* use algorithms to provide liquidity, competing with designated market makers and other liquidity providers, while for assets traded on multiple venues, *liquidity seekers* often use intelligent order routers to decide where to send an order. *Statistical arbitrage funds* use computers to quickly process large amounts of information contained in order flows and price movements in different securities, and trade at a high frequency based on patterns in the data. Finally, algorithms are used by *institutional investors* to trade large quantities of shares gradually over time[12].

[10] R. Vedapradha et al. (2023). *Algorithmic trading and its application in stock broking services*.
[11] Professor at Haas School of Business, University of California Berkeley.
[12] See: "*Does Algorithmic Trading Improve Liquidity?*" (2011).

Andrei A. Kirilenko and Andrew W. Lo, Professors at Sloan School of Management, Massachusetts Institute of Technology, identified in a brief survey[13] three developments in the financial industry over the past three decades that have resulted in the rise of algorithmic trading (AT).

The first factor they pointed out is the greater complexity of the financial system, because of general economic growth and globalization, which has also increased the number of market participants, the diversification of financial transactions, the extent and distribution of risks, and the monetary amounts involved. As the complexity of the financial system raised, a more highly developed financial technology became greater and indispensable.

The second advancement pertains to the revolutionary improvements made in the quantitative modelling of financial markets which, according to Kirilenko and Lo, have provided a strong foundation for contemporary quantitative financial analysis. Algorithmic trading is just one of the many offshoots of these developments since it is no longer possible to carry out financial analysis without the use of sophisticated mathematical models.

On the other hand, the third development is a closely related progression that pertains to breakthroughs in computer technology, spanning hardware, software, data organization, and telecommunications, which the authors attribute to Moore's Law[14]. The exponential growth in computing power per dollar has brought about significant changes in the way financial markets operate, thus impacting data storage, data availability, and electronic interconnectivity in a profound manner.

Algorithmic trading aims to eliminate emotions from transactions and ensure optimal execution of deals through the swift placement of orders. Popular trading methods include trend-following tactics, arbitrage opportunities, and index fund rebalancing. Additionally, algorithmic trading can be executed based on trading volume (volume-weighted average price) or time (time-weighted average price).
A computer software will automatically monitor stock prices and moving average indicators to place buy and sell orders when predetermined circumstances are met. This eliminates the need for manual order entry or constant monitoring of live pricing and graphs by the trader.

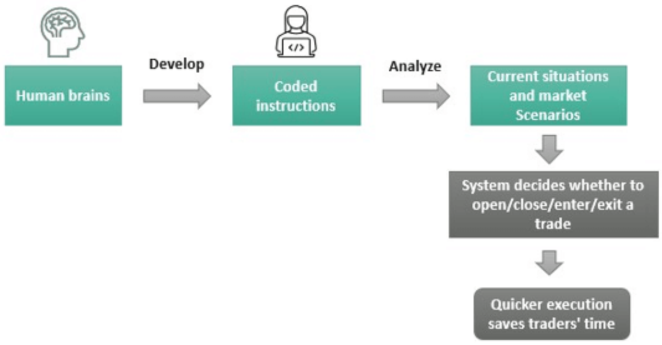
---

[13] A. Kirilenko. A. Lo, (2013). *Moore's law versus Murphy's law: Algorithmic trading and its discontents*, 53
[14] Moore’s law is the prediction made by the American engineer Gordon Moore in 1965 that the number of transistors per silicon chip doubles every year (Britannica, T. Editors of Encyclopaedia).

To make the trades start, algorithms require an identified opportunity that facilitates profitable terms of improved earnings or cost reduction, hence the occurrence of favourable patterns, which are simple and straightforward to apply through the detailed instruction that is been given to the system. A common trend-following strategy used is to monitor the 50- and 200-day moving averages. The price difference can be used as risk-free profit or arbitrage by purchasing a dual-listed stock at a cheaper price in one market and simultaneously selling it at a higher price in another market.

Among the expected results, in addition to a guaranteed profit, there is saving time for the investor. Below is a graphic representation of the working framework of AT (*See Fig. 1*).

Fig. 1: The working framework of AT



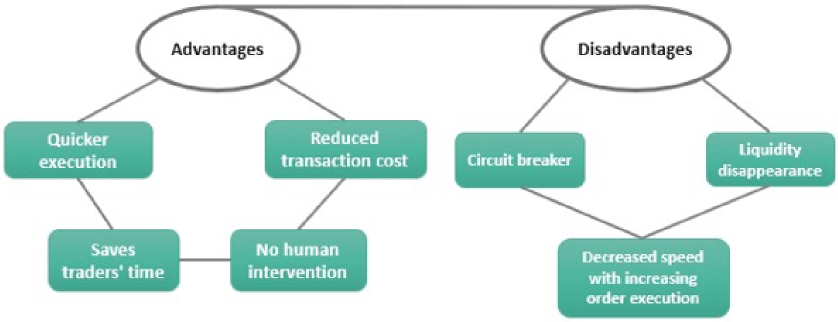
Source*: Algorithmic Trading (wallstreetmojo.com)*

Algorithmic trading can also have, contrary to initial claims, downside implications. Certain types of AT involve activities similar to traditional *market making*, where traders aim to make a profit by exploiting the difference between the bid and ask prices of a stock. Essentially, they purchase a stock at the bid price and quickly sell it at the asking price before it changes. Although this resembles the role of a market maker who provides liquidity to other traders in the stock, these algorithmic traders are not obligated to maintain both bid and ask quotes (since they are not registered market makers).

Therefore, if they withdraw their orders from a market during turmoil times, they can cause a disruption in market liquidity. This was observed during the flash crash of May 6th 2010, when the stock market experienced extreme volatility, leading to the Dow Jones index average plummeting by 1,000 points, before recovering approximately 600 points during intraday trading. This event will be further discussed in this chapter.

To conclude, the pros and cons of AT are summarised in the diagram below (*See Fig. 2*).

Fig. 2: Algorithmic Trading Pros & Cons



Source*: Algorithmic Trading (wallstreetmojo.com)*

2.1.1. High-frequency trading

At this point, it should come as no surprise that many AT strategies are speed-based. High-frequency trading is the subset of automated trading which relies on the ability of computer programs to make decisions extremely rapidly. In their research on Algorithmic trading, A. Kirilenko and A. Lo defined HFT as "a blend of technology and hyperactive trading activity […] that takes advantage of innovations in computing and telecommunication to consummate millions upon millions of trades per day.[15]"

It is however necessary to recognise that many wrongly refer to AT and HFT as synonyms, and this can happen because there is no extensive definition of HFT. On the contrary, the Division of Trading and Markets of the SEC admits that the term HFT has not been clearly defined. To try and grasp the difference between these two phenomena, one must compare the definitions given in the various papers that have been written on the subjects.

It is commonly recognised that "Algorithmic trading is based on computerized quantitative models and is used by large investors to reduce market impact" and that "[it] is not necessarily executed at high frequencies[16]".

High-frequency trading is undoubtedly grounded in the development of algorithms; however, some characterising features have been identified by the SEC. Among these are "very short time-frames for

[15] A. Kirilenko. A. Lo, (2013). *Moore's law versus Murphy's law: Algorithmic trading and its discontents*, 59-60.

[16] F. Fabozzi, S.M. Focardi, & C. Jonas (2011). *High-frequency trading: Methodologies and market impact.* Review of Futures Markets, 9(Special Issue), 7-38.

establishing and liquidating positions" and "ending the trading day in as close to a flat position as possible (that is, not carrying significant, unhedged positions overnight)"[17].

In conclusion, it is possible to state that AT and HFT are mainly distinguished by the holding period, as Algorithmic trading may have a variable holding period that can vary from hours, days to months, while High-frequency trading involves, by definition, a short holding period of just a few milliseconds and a flat position at the end of the trading session (i.e., a day).

AT made it possible to reduce market impact by spreading large orders over many small transactions, thereby originated an increase in the volume of trading which is a prerequisite for HFT[16]. The target of high-frequency traders are indeed trades that offer very small profits but also occur in considerable numbers.

Therefore, it is possible for HFTs to exploit and gain from bid-ask spreads, cross-market discrepancies, and any other small offset, present at short time horizon, which they can detect. Over and above the specific strategy they can implement, the underlying scheme is to leverage the tremendous competitive advantage of being the first to "hit the button" and execute their trades at the bid or asked price.

This is the reason why HFT firms started to compete on minimizing the distance between their servers and the exchanges on which they trade. When an order takes less than a millisecond to be executed, the extra time it takes for it to travel from a remote location could eventually make it impossible to win the race against the other traders.

The primary concept of "co-locating" is even older than the recent advancements in technology. Prior to the invention of automated markets, brokers had to manually bring their clients' orders to the exchanges quickly from the brokerage firms, which indeed placed their headquarters in New York, next to the biggest stock exchange in the U.S., the NYSE.

Nowadays, even if the trades are completely electronically executed, the need to arrive first into the exchange remains and has indeed turned into the need of being next to the computer servers of that exchange (today represented by computer servers).

[17] U.S. Securities and Exchange Commission (March 2014). *Equity Market Structure Literature Review, Part II: High Frequency Trading,* 4.

As declared in the SEC's *Equity Market Structure Literature Review Part II* (2014), "Publicly available data on orders and trades does not reveal the identity of buyers and sellers. As a result, it is impossible to identify orders and trades as originating from an HFT account when relying solely on publicly available information.[18]"
The review authored by the Division of Trading and Markets summarizes 31 papers published by authors who managed to have access to "datasets with non-public information that identifies, to a greater or lesser extent, activity arising from HFT accounts[18]".

These datasets are divided into four categories: NASDAQ Datasets, E-mini Datasets, Flash Crash Datasets, and International Datasets[19].

NASDAQ used its knowledge of the firms submitting orders to manually classify 26 firms as high-frequency traders. The factors used to identify these firms as HFT included (1) the frequency of their net trading crossing zero (i.e., flat position), (2) the duration of their orders, and (3) their order-to-trade ratio (i.e., the percentage of total volume of all posted orders to the number of orders filled). NASDAQ Datasets do not provide the full catalogue of HFT firms since they do not consider firms that also act as brokers for customers and firms that route their orders through integrated firms because this activity cannot be clearly identified.

On the other hand, E-Mini Dataset papers adopt a quantitative approach to classifying trading accounts as HFT. The SEC report states that Baron, Brogaard and Kirilenko (2012) used three criteria: "(1) trading more than 5000 contracts per day (representing a notional value of more than $300 million in August 2010); (2) holding end of day inventory positions of no more than 5% of their total volume; and (3) maintaining intraday inventory positions of less than 10% of their total volume." As a result, the group encompassed 65 HFT accounts.

According to F.J. Fabozzi, S. M. Focardi, and C. Jonas[16], "the biggest players in High-frequency trading include the electronic market-makers *Getco, Tradebot, Citadel*, and *QuantLab*; hedge funds such as *D.E. Shaw, SAC Global Advisors,* and *Renaissance Technologies*; and the proprietary trading desks of *Goldman Sachs, Morgan Stanley, and Deutsche Bank.*"

---

[18] U.S. Securities and Exchange Commission (March 2014). *Equity Market Structure Literature Review, Part II: High Frequency Trading,* 8-9.
[19] The particulars of the four categories of HFT Datasets can be found in Section III.A of the review.

2.1.2. Dark Pools

The rise of algorithmic trading, and HFT in particular, have become the subject of intense debate among all experts. It is admitted by academics that "market quality defined, for example by the size of spreads, has improved"[16]. However, the features of automated trading create new risks for large institutional investors, who fear that "they are paying a tribute to HFTs for keeping markets efficient."[16] More specifically, they are afraid that if HFTs see them posting orders, prices will move against them. This fear is completely grounded and has been the starting point of the investigation carried out by a group of investors that will be discussed in the fourth chapter of this paper.

This concern has eventually led to the creation of "dark pools" to make investors, who seek anonymity, less vulnerable to high-frequency trading when placing large orders. Dark pools have taken the role that was traditionally played by "block houses", brokerage firms specialized in matching block buyers and sellers, where blocks[20] (i.e., very large trades) were brought. The expertise of block brokers was to discreetly arrange large trades out of the public eye (i.e., in the "dark"), by identifying traders who might be interested in large purchases or sales.

The contemporary configuration of dark pools enables participants to buy or sell large blocks of securities without revealing their identity, hence without causing changes in securities prices.
Furthermore, trades are not reported until they are executed, which also limits the vulnerability of traders to the risk of being anticipated by others.

However, because the orders do not appear in a limit order book that can be publicly viewed, they do not contribute to the "price discovery"[21], hence dark pools have been criticized as exacerbating the market fragmentation.

Many dark pools claim to prevent better-informed traders (e.g., hedge funds et al.) from accessing their "pools". Unfortunately, this ideal has not always been realized; on the contrary, it has been proven several times that some dark pools were enabling high-frequency traders to participate in their market and gauge the intentions of other investors.

[20] Blocks: "large transactions in which at least 10,000 shares of stock are bought or sold." *Essentials of investments* (2013), McGraw Hill.
[21] Orders contribute to "price discovery" when they help the price reflect all publicly available information about demand for a security (*Essentials of investments, 2013*).

Michael Lewis, in his book *Flash Boys*[22], narrates the discoveries made by Brad Katsuyama and his team when they first started investigating the complex structure of the financial market around 2009. The author refers to dark pools as "another rogue spawn of the new financial marketplace". It is underlined that at the time, dark pools did not publish their rules and reported any trades they executed with sufficient delay. As a result, it was impossible to know exactly what was happening in the broader market now the trade occurred. Their appeal was built on "the amazing idea the big Wall Street banks had sold to big investors": transparency was the enemy. The assumption is that if an investor wanted to sell a million shares of a corporation (e. g. Apple Inc.) they were better off putting the order into a dark pool than going directly on a public exchange, where everyone would notice a large sell order, hence causing the Apple Inc. stocks price to plunge.

Again, the investigation conducted by Brad Katsuyama, which pointed the finger at HFTs, also revealed their presence in some dark pools that falsely claimed to exclude them (*See Chapter 4*).

2.1.3. Flash Crash analysis (May 6th 2010)

The AT industry, and High-frequency trading, were initially "largely hidden from public view"[23]. Concerns started to arise on 6th May 2010, with the so-called Flash Crash, when prices of some of the most traded companies and indexes suddenly and severely declined for about 36 minutes causing a loss of approximately $1 trillion in market value.

The research published by NANEX[24], the American financial data services provider, regarding all the flash equity crashes between 2006 and 2011, demonstrates that this Flash Crash was not the first one and, unfortunately, not the last one either. Since that spring day of 2010, an extensive number of high-profile technological malfunctions have been recorded.

A theoretical definition of Flash Crash is provided by Alfonso Puorro[25]; the analyst refers to it as a phenomenon characterised by a sudden, rapid, and often unmotivated fall in the price of one or more financial instruments, usually followed by a vigorous rebound in the minutes immediately afterward. Flash crashes can affect several securities and, in the most severe cases, can be such that they affect the performance of even the world's most capitalised indices, as it happened on May 6th 2010.

[22] Michael Lewis (2015). *Flash Boys: A Wall Street revolt.*
[23] A. Kirilenko. A. Lo, (2013). *Moore's law versus Murphy's law: Algorithmic trading and its discontents*, 52
[24] Nanex is a firm based in Chicago that provides streaming market data services, and real-time analysis and visualization tools, together with in-depth continuous research on market events and phenomena.
[25] See: "*High Frequency Trading: una panoramica"* (2013).

On the afternoon of that day, the Dow Jones Industrial Average (DJIA) dropped approximately 600 points (5.7%), and then quickly recovered (*See Chart 4*). Other major market Indexes, namely S&P 500 and Nasdaq dropped by similar amounts.

Chart 4: Dow Jones Index on May 6th 2010

Source*: Yahoo Finance*

Getting into specifics, the *Analysis of the "Flash Crash"*[26] provided by NANEX, gives a more detailed overview, also based on the report from the U.S. Securities and Exchange Commission (SEC), published in September 2010 after five months of investigation.

Beginning at 14:42:46, bids from the NYSE started crossing above the National Best Ask prices in about 100 NYSE listed stocks, expanding to over 250 stocks within 2 minutes.

NYSE quote prices started lagging quotes from other markets; their bid prices were not dropping fast enough to keep below the other exchange's falling offer prices.

With NYSE's bid above the offer price at other exchanges, HFT systems would attempt to profit from this difference by sending buy orders to other exchanges and sell orders to the NYSE. Hence the NYSE would bear the brunt of the selling pressure for those stocks that were crossed.

Minutes later, trade executions from the NYSE started coming through in many stocks at prices slightly below the National Best Bid, setting new lows for the day.
Because many of the stocks involved were high capitalization stocks and represented a wide range of industries, and because quotes and trades from the NYSE are given higher credibility in many HFT

26 See: http://www.nanex.net/20100506/FlashCrashAnalysis_Part1-1.html

systems, when the results of these trades were published, the HFT systems detected the sudden price drop and automatically went short, betting on capturing the developing downward momentum. This caused a short-term feed-back loop to develop and panic ensued. (Nanex.net)[26]

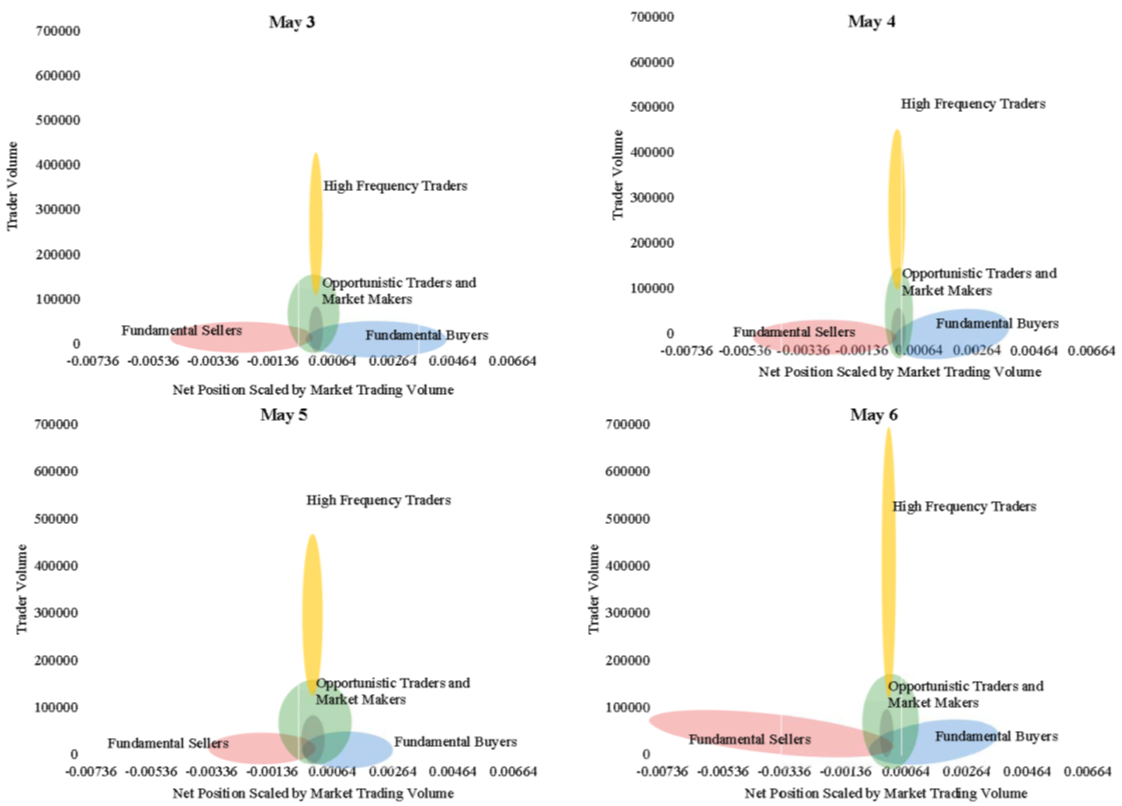
In this scenario, sellers could not match purchase orders and vice versa, resulting in irrational pricing as a result of the market makers' "stub quotes", i.e., bid and ask that they settled at prices very far from their actual value, in order to fulfil their obligation to continue listing.

The first SEC report issued after the event, pointed to a $4 billion sale of market index futures contracts by a mutual fund: Waddell & Reed Financial. Apparently, it was decided to sell through a "*sell algorithm*" programmed to execute orders in relation to the trading volume calculated every minute.

This algorithm executed the sales program in just 20 minutes and this important quantity sold was mainly absorbed by HFTs and brokers. Specifically, high-frequency operators accumulated a considerable net position of around 3,300 contracts, and between 2:41 pm and 2:44 pm they aggressively sold about 2,000 E-mini contracts to temporarily reduce their long position. As a result, a large quantity of liquidity was withdrawn from the market.

The figure below (*See Fig. 3*) shows trader categories plotted (as shaded areas) over all individual trading accounts ranked by volume and net position scaled by market volume (in the June 2010 E-mini S&P 500 futures contract) from May 3rd to May 6th 2010.

Fig. 3: Trading accounts, trading volume, and net position scaled by market trading volume[27]



Source*: The Journal of Finance (January 6, 2017)*

The SEC also mentions the lack of a clear system to pause trading during periods of extreme market volatility as one of the main factors. Eventually, at 2:47 pm, the circuit breaker stopped what is commonly defined as the "snowballing effect" and the market went back to work as usual again.

2.2. Regulation of the U.S. Stock Market

As has been mentioned at the beginning of this Chapter, the rise of HFT was also enabled by the regulatory framework. Going into greater depth about what has changed in the institutional rules on financial markets also makes it possible to answer a common question that people asked themselves when they started digging into the HFTs operate:

"How was it legal for a handful of insiders to operate at faster speeds than the rest of the market and, in effect, steal from investors?" (Michael Lewis, 2015)[28]

[27] Figure retrieved from "*The Flash Crash: High-Frequency Trading in an Electronic Market"* (2017).
[28] Michael Lewis (2015). *Flash Boys: A Wall Street revolt,* 96.

2.2.1. The Role of the SEC

The two institutional actors managing all the regulations of the financial markets in the U.S. are the Congress and the Securities Exchange Commission (SEC), which was created in 1934 by the other. Although the SEC is an independent federal identity, it must follow the Congress' guidelines.

As can be retrieved from the SEC.gov website under the section *our goals*, their common objective can be defined as a "long-standing three-part mission" in which they aim "to protect investors, maintain fair, orderly, and efficient markets, and facilitate capital formation"[29].

A critical turning point in the regulation of U.S. capital markets can be found in 1975. In that year fixed commissions on the NYSE were eliminated. This new fee system was intended to enable brokers to compete by lowering their fees. Congress also amended the Securities Exchange Act (1934) to create the *National Market System* to centralize, at least partially, trading across exchanges and enhance competition among different market makers.

In Section 11 A. (a) (1)[30] of Securities Acts Amendments of 1975 is stated as follows:

(C) It is in the public interest and appropriate for the protection of investors and the maintenance of fair and orderly markets to assure:
(i) economically efficient execution of securities transactions;
(ii) fair competition among brokers and dealers, among exchange markets, and between exchange markets and markets other than exchange markets;
(iii) the availability to brokers, dealers, and investors of information with respect to quotations for and transactions in securities;
(iv) the practicability of brokers executing investors' orders in the best market [...]

The main idea (iii, iv) was to implement centralized reporting of transactions as well as a centralized price quotation system where traders were given a broader view of trading opportunities across markets (i.e., securities exchanges).

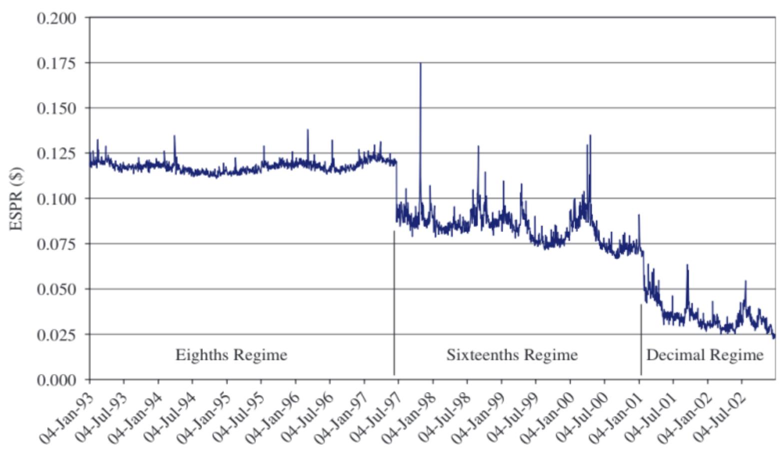
---

[29] See: https://www.sec.gov/our-goals for a more detailed description of their goals.
[30] See: https://www.govinfo.gov/content/pkg/STATUTE-89/pdf/STATUTE-89-Pg97.pdf

In 1994, a scandal at NASDAQ revealed that dealers were colluding to maintain wide bid-ask spreads. This resulted in an antitrust lawsuit being brought against NASDAQ and new order-handling rules being instituted by the SEC. These rules allowed customer limit-orders to effectively compete with dealer quotes and enabled electronic exchanges to compete for trades.

In 1997, the SEC allowed the minimum tick size to fall from one-eighth of a dollar to one-sixteenth, and in 2001, decimalization allowed the tick size to fall to 1 cent. These changes led to a further narrowing of bid-ask spreads. The effective spread, or the cost of a transaction, fell dramatically with the minimum tick size from the eighths regime to the decimal regime, through the sixteenths regime. A graphical representation of the evolution of those spreads between 1993 and 2002 is presented below (*See Chart 5*).

Chart 5: Value-weighted daily average effective spread, NYSE, 1993–2002



Source*: Journal of Financial Economics 87 (2008), pp. 249–268.*

2.2.2. Securities Information Processor (SIP)

Before continuing with the discussion of the relevant regulation in the sphere of High-frequency trading, it is necessary to introduce a brief presentation of the Securities Information Processor (SIP).

Section 22 A.[30] of Securities Acts Amendments of 1975 defines it as follows:

(22) (A) The term 'securities information processor' means any person engaged in the business of:

(i) collecting, processing, or preparing for distribution or publication,

[…] of information with respect to transactions in or quotations for any security

(ii) distributing or publishing [...] on a current and continuing basis, information with respect to such transactions or quotations.

In essence, as is also stated on the NYSE website, the SIP is the mechanism giving the measure of the entire market by compiling and processing all the bids and asks for all U.S. stocks in one single data feed.[31] The SIP disseminates and calculates critical regulatory information including, among other important information (such as short sale restriction and regulatory halts), the National Best Bid and Offer (NBBO)[32], concept on which the Reg NMS relies.

2.2.3. Regulation of NMS and ATS

The turning point in regulations arrived in 2005, with the adoption of the Regulation for Alternative Trading Systems (ATS) and the Regulation NMS, which was widely implemented in 2007.

According to P. Mahoney and G. Rauterberg (2017)[33], Regulation for ATS inaugurated a process of bringing all trading systems substitutes, that accounted for a sufficient portion of trading in the stock market, into the national market system. This would happen by bringing their best bids and offers into the public quote stream and giving the public the ability to execute against them. It is claimed by some that many of the biggest dark pools avoided falling under this category by consequently changing their business models.

Ultimately, the answer to the question that was presented at the beginning of the paragraph (*See Chapter* 2.2) can be found in the adoption of Regulation of National Market System. The effects of this measure became visible two years after being approved in 2005, when it was fully implemented.

The objective was to establish a unified electronic market by connecting exchanges through electronic means. According to the regulation, exchanges were obligated to execute the quotes of other exchanges automatically. If an exchange was unable to process a quote electronically, it would be deemed a "slow market" under Reg NMS and could be ignored by other participants in the market.

[31] Additional information about SIP available at www.nyse.com "Current capacity messages per 100 milliseconds of 392,000 for the quote feed and 86,000 for the trade feed. Capacity above peak with capacity-to-peak ratio of 2.4:1 for the quote feed and 1.7:1 for the trade feed. Current median latency of about 230 microseconds"
[32] See *Chapter 2.2.3*
[33] See: "*The Regulation of Trading Markets: A Survey and Evaluation"* (2017).

As a result, just prior to the rule's final adoption, the NYSE and NASDAQ, which were particularly at risk of being passed, purchased ECNs and readied themselves for an aggressive shift towards primarily electronic markets that would enable automated execution against quotations displayed to the public.

Before then, the brokers responsible for executing investors' stock market orders were held to a lax standard known as "best execution", which was in practice open to interpretation.

To make it clearer, if an investor wanted to purchase 10,000 shares of Microsoft Corp. at $30 a share and the broker discovered that only 100 shares were available at that price, they could decide not to purchase those 100 shares and instead wait for more sellers to appear. This allowed brokers to exercise their discretion to execute trades on behalf of their clients in a strategic manner without spooking the market. However, after brokers repeatedly abused this discretion and lost the trust of investors, the government intervened and replaced the loose concept of best execution with a stricter legal standard called "best price".

More precisely, the rule considered is Rule 611, as known as the Order Protection rule, which is an update and a reinforcement of the National Best Bid and Offer (NBBO) framework. Rule 611 was designed in part to protect investors entering market orders from receiving inferior prices. However, this was not the principal objective. Instead, it is commonly recognized that an attempt to reward and thereby encourage the provision of liquidity through limit orders.

Nevertheless, from that point on, if the above-mentioned investor still intended to purchase 10,000 shares of Microsoft Corp. and 100 shares were accessible on the BATS exchange at $30 per share, but the remaining of 10,000 shares on the other twelve exchanges were available at $30.01, the broker must procure the 100 shares on BATS at $30 prior to exploring the other exchanges.[34]

Regulation NMS was intended to create fairness in the U.S. stock market. At the same time, the regulation made it easier for HFTs, a small class of insiders with resources to create speed, to predict where the brokers would send their clients' orders, hence enabling them to front-run them, leading to a new type of inequality.

[34] Examples retrieved from Michael Lewis (2015), *Flash Boys: A Wall Street Revolt,* 96-97.

High-frequency traders were able to set up computers inside the exchanges and build their own, much faster version of the SIP, which created an advantage in latency. This meant that fast traders could stick an investor with an old price and exploit it within a twenty-five-millisecond advantage, which made volatility very valuable to them.
Eventually, by complying with Reg NMS, the so-called smart order routers simply marched investors into various traps laid for them by HFTs.

Michael Lewis, in his book, also mentions that papers of researchers have been published showing that the SIP price of stocks (e.g., Apple Inc.) and the price seen by HTFs with faster channels of market information could differ thousands of times per day (55,000 times in a single day). This was only one of the most straightforward ways a high-frequency trader could exploit the ignorance of the slower-footed investor.

Therefore, the subsequent chapter (*See Chapter 3*) will be devoted to a more detailed overview of the behaviour of, and techniques used by these traders.

CHAPTER III

3. HFT: infrastructure, strategies, and effects

In the preceding section (See *Chapter 2.1.1*) has been stated that the key feature of high-frequency systems is the speed at which they can enter, execute, modify and even delete orders in the market. Only this competitive advantage allows them to exploit inefficiencies and opportunities that would otherwise go unnoticed.

In this respect, it is fundamental to have logistical and electronic support capable of executing operations in a few milliseconds, which consists of an appropriate infrastructure. Then algorithms must be trained to implement specific strategies. Eventually, this can affect market efficiency.

3.1. HFT technical characteristics

The competitive advantage of speed is obtained thanks to two main characteristics that have been also mentioned above in this paper: low latency and co-location.

3.1.1. Low Latency

Latency can be defined as the time required to implement the series of necessary transactions to turn an economic decision into an actual negotiation, i.e., the time between the initial moment of a transition process of the order and its effective execution.

To be precise, according to A. Puorro[25], between these two temporal phases there are others that deserve to be considered, such as:

i. the speed at which traders can receive information from the market (i.e., data) and can process it (i.e., operational reaction & data analysis);
ii. the time between the data analysis and the transmission to a broker, comprising of the preparation of the trade order and the time to be physically sent;
iii. the time between the processing of the data, and the sending of the trade order to the trading venue by the broker; the broker's computer systems need time to recognize the type of order received (Buy or Sell) and its technical characteristics, which financial instruments need to be traded and on which market, and eventually send the order.
iv. the time frame that the order takes to arrive in the market from the time it is issued by the broker;

v. the time between the data is received by the market and the disclosure of this to all market participants.

High-frequency systems are characterised by a low latency: they need the least amount of time, only a few milliseconds to cover all the steps in the investment process listed above.

Hence, HFT firms must not only have a computer and technological apparatus capable of receiving, analysing, and processing market data and information in an extremely small-time interval but also avail themselves of a broker who has the technology able to minimize the latency of the processes.

Low latency can only be obtained through heavy investments in technological infrastructure and hardware to increase the speed of the whole process. Some people grasped before others that the exchanges were evolving and that were transforming into simple stacks of computers in data centres, so they decided to find the fastest road to arrive there. It is the case of the Spread Network's founder, which will be presented further on (*See Chapter 3.1.3*).

3.1.2. Co-location

The aforementioned measures are undoubtedly essential, but they are not enough to minimise latency, since "stock exchange orders are electrical impulses that, although travelling at very high speeds, encounter space limitations[35]".

The greatest enemy of speed is, in fact, distance; for this reason, HFTs have implemented a new method to tackle it and boost their speed by a few, yet extremely crucial, milliseconds: *co-location*.

The proximity of their own server to that of the exchange on which they have chosen to trade becomes a considerable advantage that, although infinitesimal, allows them to act faster than non-high-frequency traders. Consequently, they can take advantage of certain investment windows that only open for a few seconds or fractions of a second.

The mere presence of other high-frequency traders in the market makes the industry extremely competitive and the lack of the co-location requirement would result in the loss of profit opportunities that would instead be exploited by faster traders.

---

[35] A. Puorro (September 2013), *High Frequency Trading: una panoramica, 11.*

The first exchange to open to co-location possibilities has been the NYSE Euronext in 2010, with the proposal of an amendment of the Price List to reflect fees charged for Co-Lo services presented to the Securities and Exchange Commission.

On August 16th, SEC issued the filing of the exchange amendment[36], which reported the exchange plans to begin operating a data centre in Mahwah, New Jersey, from which it would offer co-location services (i.e., space at the data centre in cabinets with a power usage capability of either four or eight kilowatts).

The NYSE Price List has been updated several times since then and the current one, which also includes fees for these types of services, is available on the NYSE website[37]. However, already mentioned in the first publication was the possibility granted to users to "obtain an option for future use on available, unused cabinet space in proximity to their existing cabinet space[36]".

In fact, all the major exchanges have seized on the demand for the co-location of HFTs and gradually moved their servers off Wall Street in New Jersey, where there was plenty of space to rent for the aforementioned services. Banks, proprietary trading firms, and HFT firms especially started to pay to have their computers as close as possible to the "market".

Nowadays, the exchanges offer a "wide variety of connectivity options to fit a firm's specific network and trading strategy needs"[38]. As can be read on the NASDAQ exchange website, customers can also choose from several options, including the high-speed 40G Network, and significantly reduce latency network complexity "by utilizing a single hand-off to reach all Nasdaq markets[38]".

Hence, it is evident that without co-location the low latency would be of low effectiveness and vice versa.

3.1.3. Spread Network's fiber line

Dan Spivey is a former stockbroker who had realized, by 2008 that the speed with which trades occurred was no longer constrained by people and that there was a big difference between the trading

[36] U.S. Securities and Exchange Commission (August 16, 2010). *Notice of Filing of Proposed Rule Change Amending its Price List to Reflect Fees Charged for Co-Location Services.* Link of the pdf available in the other references section.
[37] See the NYSE Price List at https://www.nyse.com/publicdocs/nyse/markets/nyse/NYSE_Price_List.pdf
[38] See the NASDAQ brochure at https://www.nasdaq.com/solutions/nasdaq-co-location

speed that was available between the exchanges in CBOE[39] and New York[40] and the trading speed that was theoretically possible.
In more detail, given the speed of light in fibre, it should have been possible for a trader to send his order from Chicago to New York and back in roughly 12 milliseconds, or roughly a tenth of the time it takes to blink. However, the routes offered at that moment were lower than that and even inconsistent (sometimes it could take 17 milliseconds, sometimes 16, and seldom any less).

Whereas previous connections between the two financial centres zigzagged along railroad tracks and around mountains, Spivey founded the Spread Networks LLC and started the construction of a new high-speed fibre optic cable connecting financial markets in New York and Chicago. This cable was dug in an 825-mile nearly straight line and by the end of the work in 2010, the round-trip communication time between Chicago and NY was reduced by 3 milliseconds (i.e., to 13 milliseconds).

Fig. 4: Spread Network original route

Source*: Spread Network website*

Three milliseconds cost around $300 million, but they mean an "eternity" to high-frequency traders. Later, the company reduced latency of 1 microsecond on the dark fibre service because of its continuous route improvements. To provide an idea of the power of those milliseconds, it is necessary to remark that the company started charging firms over $20 million each, to benefit from its cables.

3.2. HFT strategies and new types of orders

The synergy between low latency and co-location also allows high-frequency traders to implement new operational strategies that neither an ordinary investor nor every algorithmic trading system,

[39] Chicago Board Options Exchange
[40] As mentioned before, when referring to New York exchanges, one means in New Jersey, where they were relocated in those years.

would be able to perform. At the same time, HFTs have been able to create and launch on exchanges such as BATS, new types of orders specifically designed to their advantage. One particular type is the post-only order, the execution of which is subject to specific requirements such as that the trader is on the passive side of the trade so that a rebate offered by the trading venue can be obtained. Alternatively, a hide-not-slide order is even more intricate and enables the HFT to jump in as first in the line to purchase and obtain the kickbacks from the exchange.

Regarding the way of operating, there are a series of strategies through which HFTs seek to maximise the economic return from the time and technical-computational advantages they hold over the rest of the market, all of which seek to exploit small market inefficiencies and, in some cases, aim to exploit market liquidity with the mere purpose of detecting it. This section is a brief description of the main strategies used by high-frequency traders.

3.2.1. Statistical passive arbitrage (Latency arbitrage)

The first type of strategy is a natural field of application for high-frequency traders because it exploits statistical passive arbitrages.
Arbitrage refers to price divergences between identical assets that have no economic justification other than a momentary inefficiency in the functioning of the markets on which they are traded. This strategy allows the whole market to respect one of the main hypotheses on which it is built: the Efficient Market Hypothesis[41].

According to the economic theory, indeed, prices immediately reflect all the available information about a certain security. Therefore, arbitrage opportunities are constantly present, yet they are as rare as immediately absorbable events.

However, due to the increased speed with which traders identify and take advantage of such occurrences in the most recent years, an increasingly sophisticated level of technology to fully exploit the breadth of arbitrage is now required. On the other hand, the spread of financial instruments listed and traded at the same time on different places increased the possibility and frequency of this type of arbitrage.

[41]"Efficient market hypothesis (EMH): The hypothesis that prices of securities fully reflect available information about securities". *Essentials of investments* (2013).

The possibility of outperforming all traditional traders is granted by the situation they find themselves in, which is a double disadvantage. To make the point, when small price divergences appear among stock exchanges, HFTs immediately detect and exploit them, while the rest of the investors suffer a disadvantage in both the identification and execution phases of the transaction. As it is linked to the latency of the orders, this is also referred to as latency arbitrage.

3.2.2. Liquidity providing

It was previously stated in this work that some algorithmic trading techniques simulate the work done by designated market makers, with the only difference being that, not belonging to the latter group, they are not under the strict constraints of their regulation, regarding minimum number of contracts, maximum spread allowed, time, and so on.[42]

This is the case of this so-called liquidity providing or market making strategy, whereby the high-frequency trader is not obliged to be the counterparty to all incoming orders and is able to choose whether, how and to what extent to operate, gaining from the bid-ask spread.

The trader will accordingly place their trading proposals on the first levels of the trading books, both on the bid and ask side (i.e., providing liquidity), and wait for market participants, who act as price takers, to cross them with their buy or sell orders.

The impact this technique has on the market depends on its volatility at the time. When the volatility of the market is low, this strategy allows the HFTs to profit from the bid-ask spread, the orders and quotes are stable and do not represent too much risk for them, who are able to buy at a low price (i.e., bid price) and sell at the higher price represented by the ask. Consequently, the technique will boost market stability and the liquidity provided.

On the other hand, if the market is in a period of high volatility, the risks associated with it will rise, so traders will increasingly withdraw their orders, removing liquidity from the market to avoid losses. The volatility will eventually further exacerbate, and the market will become weaker.

[42] A. Puorro (September 2013), *High Frequency Trading: una panoramica,* 14.

3.2.3. Passive rebate arbitrage

The Passive Rebate Arbitrage represents a strategic compromise between the latency arbitrage and liquidity provision described above. The main objective of this technique is to exploit the profit opportunities offered by the particular features of the new commission structures of ECNs, which have been addressed at the beginning of this chapter.

Indeed, the presence of several trading systems parallel to regulated markets has led to increased competition between ECNs and other markets as well as between multiple ECNs. To tackle this competition, platforms have also started to create new commission structures to give high-frequency traders the opportunity to profit from commissions.

The key feature of this type of strategy lies in the fact that ECNs have started to offer commission rebates or even passive rebates every time liquidity providers offer liquidity, thus fulfilling the orders of price takers.

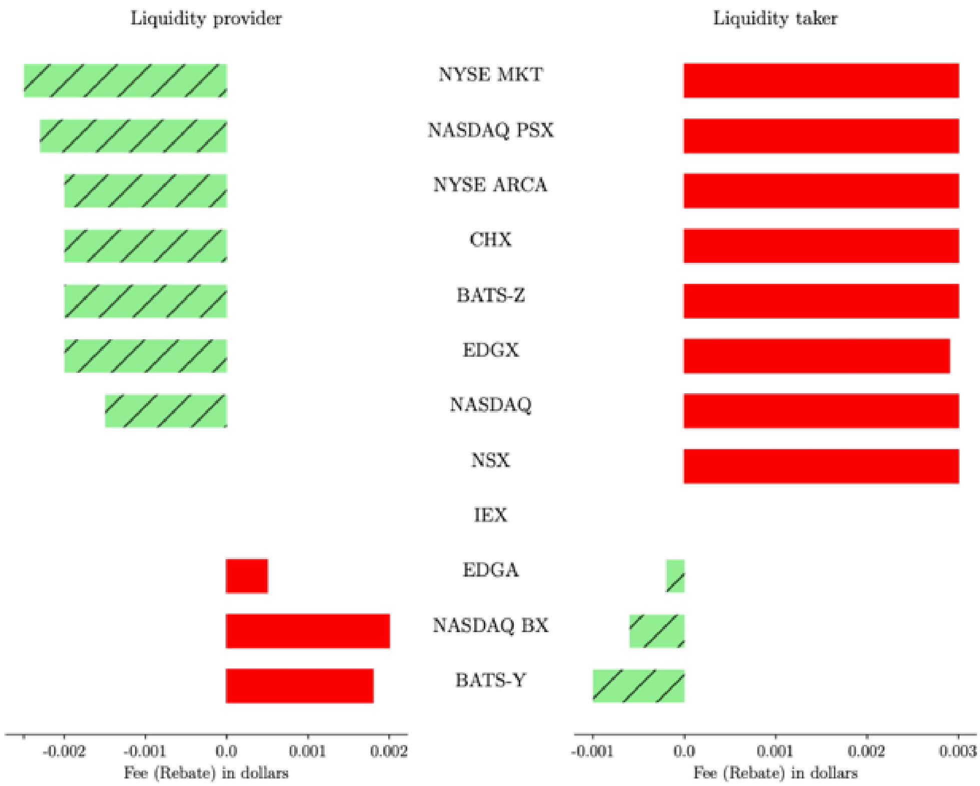
HFTs are thus prompted to offer liquidity to ECNs for the sole purpose of accumulating the rebates recognised by the multilateral system. Their technical characteristics then make it possible to maximise the returns from a passive rebate strategy, while managing to minimise the risks[43].

Many insightful studies have been conducted on the structure of the fees that exchanges charge, and on the effect they have on market quality. C. Comerton-Forde, V. Grégoire, and Zhong[44] highlighted that fee models themselves are extremely complex, with exchanges charge different fees depending on participant activity level and stock price. It is not rare to see changes in the fee models adopted by a certain exchange. The exchanges can switch from a traditional make-take fee model[45], to an inverted fee model (i.e., liquidity demanders are paid a rebate and liquidity suppliers are charged a fee); otherwise trading venues can adopt a fee-only model, where no rebates are paid.

The current structure of fee models is presented in the figure below (*See Fig. 5*). Eight of the exchanges offer a traditional make-take fee model, only three offer an inverted fee model, and two offer a fee-only model.

[43]As already mentioned in the previous paragraph, in the event of unusual orders or sudden decreases or increases in liquidity, the ability to immediately react guaranteed by the high frequency allows HFTs proposals to be quickly withdrawn from the trading books, offering the possibility of minimising the risks associated with the strategy in question.
[44]C. Comerton-Forde, V. Grégoire, Z. Zhong (October 2019). *Inverted fee structures, tick size, and market quality.* Journal of Financial Economics, 141-164.
[45]Where liquidity providers are paid a rebate and liquidity demanders are charged a fee.

Fig. 5: Fees and rebates for Tape C stocks[46]



Source: *Journal of Financial Economics 134 (2019), 145.*

3.2.4. Trading on news (Momentum Trading)

A natural operational output for high-frequency systems is the possibility of exploiting the effect that news and macroeconomic data can have on the price trend of financial instruments. This technique is in fact very simple to implement and simply takes price movements into account when making buy or sell decisions.

The fundamental precondition is to have computer systems capable of drawing operational instructions from the continuous flow of information from the main information providers (i.e., algorithms). Such systems are able to associate trading strategies with certain patterns of words in the news, adapting strategies to the resonance of the news, that is to say, the number of times the news is reported.

[46] The figure above presents the standard exchange fee and rebate schedule, in $ per share, for Tape C stocks (NASDAQ-listed stocks) priced above $1 as at 21 December 2016. Positive numbers (red solid bars) indicate a fee while negative numbers (green hatched bars) indicate a rebate.

It is therefore possible to minimise the time needed to interpret the news, as strategies are automatically and immediately associated with certain word and phrase configurations, eliminating the time needed for a human interpretation process.

Although this strategy can also be exploited through simple algorithmic trading, as it gives straightforward instructions based on the interpretation of the news provided by the so-called "financial data vendors", the high frequency allows one to maximise profits from trading on news, as the speed allows one to make the most of the rapid and sometimes violent movements that the markets manifest after the publication of macro data or important news, and of course beat ordinary investors to the punch.

3.2.5. Flash Trading

Flash trading is one of the most aggressive and controversial strategies used by HFTs and has therefore attracted the most criticism from the trading press and some academic research. This service, offered by some ECNs to high-frequency clients who wish to take advantage of it, represents a sort of "pre-emption on orders[47]" that arrive on the market and that cannot be executed. When an order arrived on the exchange cannot be partially of fully executed because it has a different price than the National Best Bid or Offer, it must be sent to other trading venues (according to regulation previously discussed in this paper), where orders at the NBBO are placed in that moment.

Instead, these ECNs, before transferring the order, offer it in visibility to HFTs that have subscribed to this flash trading service, who will then be the first ones to view any order which is not matching the market price.

Thanks to speed, it is therefore possible to exploit this information in the same manner as latency arbitrage (i.e., to profit on price arbitrages). The feature that distinguishes this strategy from the previous one is that in this case, the ECNs and the HTFs agree to profit together and make liquidity remain on that trading estate. As a result, the service provider (i.e., the stock exchange) makes its own market more robust and efficient in the eyes of other participants, and thus more appealing.

---

[47] A. Puorro (2013), *High Frequency Trading: una panoramica*, 16.

As far as HFTs are concerned, the reasons why they decide to stand as a counterpart to flash trading are various. A. Puorro[47] identified four main ones, according to their position on the security:

I. If the HFT is already long on a stock, flash trading offers the opportunity to close the position at a price before it becomes available to the rest of the market.
II. If the HFT is flat[48], the flash order offers an arbitrage possibility, as the trader knows there is pressure on the demand side a little in advance. At this point he will try to buy up an amount of shares, not exceeding the quantity in the flash order, at a price lower than that the one detected on the flash order, knowing that he can then resell them at a profit at that latter price.
III. If the HFT is short, the flash order is an informational advantage because it provides a demand-side pressure signal that they can use to close the short position before the other traders.
IV. If the HFT is active with quantities for sale on markets where its offer is not the NBBO, the system can process the information offered by the flash order and, if the demand pressure appears unsustainable, remove its quantities for sale on the other ECNs as well, where the order is likely to be sent, so as to avoid being short in a market with pressure on the long side.

In all such cases, the high-frequency traders, by exploiting an information advantage of less than one second in time, achieve a risk-free gain, also designated as a "free lunch". This gain, although minimal on a single trade, is multiplied by the opportunities on all possible securities and in several markets, so it reaches millions of profits that together constitute a considerable gain.

3.2.6. Liquidity detection

Liquidity detection is the activity by which algorithmic traders, also including HFTs, "test" certain key levels through small market orders (e.g., 100 shares) to detect stop losses or take profits on the market.

The stop loss is a conditional order used to minimise losses by closing the position at a certain price level that corresponds to the maximum loss that is intended to be tolerated. The take-profit order, on the other hand, is designed to automatically close the position in profit when a certain event occurs.

[48] When the trader it is not in a long or short position.

These AT systems, also called algo-searchers, then try to trigger the stop losses and take profits of other market participants by marking as “traded” a price deemed by the algorithms to be a triggering event of the condition underlying the SLs or TPs, in order to accumulate the liquidity offered (either buy or sell) in correspondence with such close orders.

Pre-contingent orders have, indeed, the disadvantage of representing latent trading orders. Their positioning is therefore easily predictable through the behavioural analysis of other traders, performed by algorithmic systems.

3.2.7. Ignition Momentum

The following strategy is a continuation of liquidity detection, and it is known as “ignition momentum”. This technique involves a high-frequency trader taking an aggressive position (a large long or short order), which generates a strong price movement and prompts other traders to react to the movement. Then the momentum trader can close his position, taking profits.

The ignition momentum is, therefore, characterised by an initially stable price, a sudden increase in volume followed by a strong price movement in the desired direction and a return of prices to starting levels, where the trader gains from price spikes. This strategy is accordingly a highly speculative and aggressive strategy, suitable for periods of low volatility in the market.

3.2.8. Quote stuffing

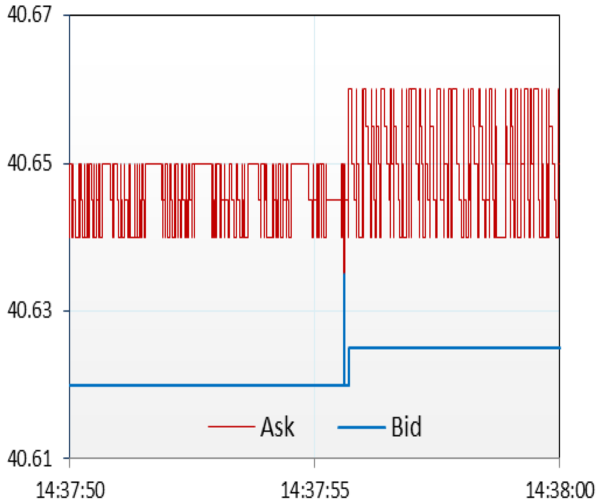
Quote stuffing is an even more aggressive strategy than the previous two, aimed at increasing the chances of latency arbitrage by artificially creating the conditions for it.

In facts, an HFT, aware of its time advantage, may dump an unmanageable amount of information on the market by placing and simultaneously cancelling thousands of limit orders. The effect is to cause a slowdown in the operation of other trading systems, which lose some of their time advantage in processing the HFT's massive amount of information, as well as a delay in the operation of the market receiving the orders.

The consequence of this is a proliferation of latency arbitrage opportunities that can only be exploited by the HFT that initiated the quote stuffing, as most other traders are likely to be computationally limited in processing and reacting in time to the information.

The chart below, presented by A. Puorro[49] (*See Chart 6*) relates to the episode that occurred on May 2nd 2012 on the Heineken share, listed on Euronext. One can see the entry and deletion of thousands of limit orders on the ask side. It must also be remembered that for this technique, limit orders are used, as they allow the limit order books to be swamped without turning into actual trades for the submitter (who can in this way cancel them).

Chart 6: Quote stuffing on Heineken share May 2, 2012



Source: Credit Suisse AES Analysis

3.2.9. Ghost strategies (smoking, layering, and spoofing)

"Ghost strategies" is a term that refers to a set of operations that differ from each other only in minor functional aspects, and are specifically smoking, layering, and spoofing strategies.

The practice of smoking consists of luring a considerable number of slow traders by placing very attractive "fake" orders, which, however, are quickly modified and generate less favourable conditions for the counterparties attracted by the initial conditions, before the latter can realise the changes.

On the other hand, layering consists of simultaneously placing an unobservable order in the trading book on one side of the market (e.g., sell) and another order, this time visible in the book, on the other side (e.g., buy) so as to make other market participants mistakenly believe that a bearish phase is about to begin. The sell order is then cancelled before it is executed due to the high speed typical of HFTs and then the buy order is secretly executed at a lower price.

[49] A. Puorro (2013), *High Frequency Trading: una panoramica*, 21.

The spoofing method is also about obtaining better buying conditions for a security that the trader considers interesting. The specific aim is to manipulate the market by placing a series of sell orders to convince other investors, as with layering, that a bearish phase is occurring. Again, orders will be cancelled before being executed.

Overall, through these strategies, HFTs continuously flood and alter the trading book, artificially simulating market situations for the sole purpose of inducing the traditional trader or algorithmic trader to make wrong trades in response to unreal situations, thus creating a sort of market deception.

The HFT then acts as a counterpart to these operations and, after accumulating a certain amount of liquidity, reproduces the same behavioural pattern in the opposite direction, prompting traders to close previously accumulated positions at a loss due to the falsity of the stimuli. Then, by creating the same situations again to lead to an increase in trading orders, they close their initial trades at a profit. The strategic objective is to exploit the predictability that characterises some institutional traders (namely mutual funds, and pension funds).

3.3. How market efficiency is affected

It must be admitted that, while providing a clear overview of the techniques implemented in high-frequency trading is difficult and not always clarifying, knowing with certainty what the effects of such practices are is an even a more challenging matter.

The academic literature is also divided on the subject: some emphasise the possible positive effects of their actions, while others, such as the press, highlight the negative impact they would have on the market, namely on quality, efficiency and volatility. Since the rise of the phenomenon, and in particular after some episodes with significant news exposure in the press, such as some flash crashes mentioned earlier in this paper, many academics and other experts have been trying, with no little struggle, to prove to one extent or another the effect of algorithmic and HFT trading practices. Indeed, the results of these studies often lead to contrary conclusions, depending on what is taken into account for the analysis.

Thus, in some cases, studies have led to the conclusion that high-frequency traders' operations have positive effects in terms of liquidity, volatility and price information efficiency. In other cases, the research result has been the opposite, showing a reduction in market quality and significant increased systemic risk. The complexity of the effects of high-frequency trading and the conflicting results of theoretical and empirical research justify the delay in regulation. The authorities have only acknowledged the phenomenon without arriving at proper regulation.

The first research that will be presented in this paper focuses on 42 equity markets around the world to investigate the relation between AT intensity and market quality, summarized through measures of liquidity, volatility, and informational efficiency between 2001 and 2011. Boehmer at al. report that "the average effects estimated in empirical results tend to be positive"[50].

More specifically, they conduct a two-dimensional analysis within each of the 42 markets considered using daily two-way fixed-effects regressions and documented this relationship in panel regressions that control for firm and day fixed effects. Their results show that a higher AT intensity on the market causes greater liquidity, improved efficiency, and greater volatility. However, these effects are not uniform across time and on all the stocks considered in their dataset[51]. They recognised that "AT has systematically negative effects on the liquidity of small or low-priced stocks and also increases volatility more in those stocks".

Moreover, they provide evidence that AT's liquidity provision does not apply to all firms and that, instead, it declines on days when market making[52] is unusually costly. Eventually, they state that the systematic increased volatility caused by this is not always beneficial and imposes costs on most market participants.

It is commonly agreed the little understanding of the highly secretive strategies used by HFTs leads to conflicting opinions and theories. According to F.J. Fabozzi et al., if an "high-frequency trader does pure arbitrage and is not predatory, not manipulative, there is no problem"[53]. This statement reflects the previously mentioned idea that high-speed algorithmic trading is not harmful per se. The problem is indeed "what we do not know"[53], i.e., it is not straightforward, even for the SEC and the FINRA[54], to assess the limits and the extent of the phenomenon and thus its potential distorting effects on the market.

---

[50]See: "*International evidence on algorithmic trading*" (2012), 5.
[51]They considered Thomson Reuters Tick History (TRTH) database, which contains intraday trades and quotes data for many markets around the world, intraday data from Trades And Quotes (TAQ) database, Datastream and Center for Research and Security Prices (CRSP) database, Ancerno database complied by Ancerno Ltd. More information regarding the database used in their analysis can be found in paragraph 3 of their paper.
[52]As previously mentioned in this paper, market-making strategies are an important subset of the strategies available to algorithmic and high frequency traders, but it must be kept in mind that HFTs are not subject to market makers regulation.
[53]F. Fabozzi, S.M. Focardi, & C. Jonas (2011). *High-frequency trading: Methodologies and market impact.* Review of Futures Markets, 9(Special Issue), 29.
[54]The largest independent securities regulator in the U.S.

The research also provides an example of positive impact involving the Canadian market. They argue that the entry of HFTs on the Canadian market has led to a better market: an influx of new participants to the exchange and thus a diversification of order flow and trading strategies.

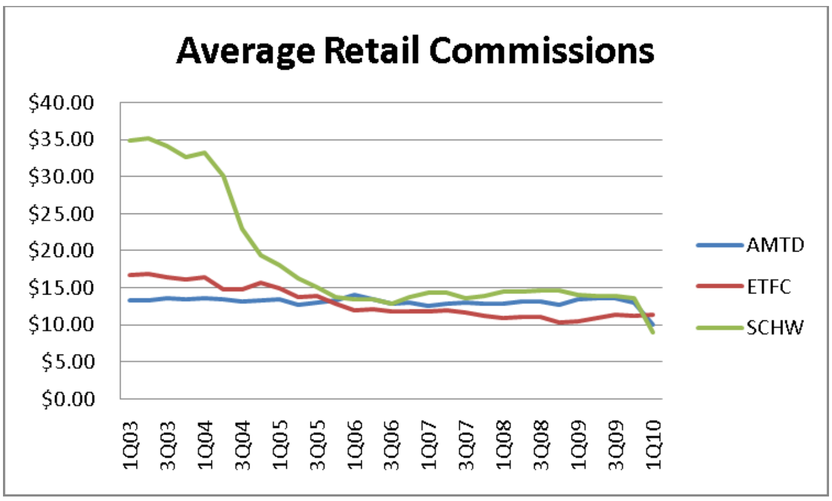
This could be the case because, thanks to these traders, big names in the U.S. have set their eyes on Canada, thereby attracting others. "As liquidity improves and trading speeds increase, the increased activity on listed stocks means that companies are trading that were previously excluded by filters that exclude stocks trading less than 1 million shares per day”[53].

Companies would therefore benefit, as they have greater access to capital and lower costs. It is also admitted that short-term volatility spikes increase, however Fabozzi et al.[53] state that “what happens on an intraday basis does not have a major impact on the long-term investor”.

Among the many studies treating this phenomenon, it is necessary to report the conclusion reached by Angel, Harris and Spatt (2010) which summarises all the main identified positive effects of HFT. First, there is an increase in liquidity available to market participants, i.e., an increasing market depth (*See Chart 3*). The second benefit pointed out is that the average bid-ask spread (*See Chart 5*) has tightened decisively, also thanks to the regulatory effort., and remains at low levels.

Moreover, transaction costs have also decreased, as commissions for retail traders have fallen substantially and continue to fall. The chart presented below (*See Chart 7 below*) shows the average commissions charged by three of the largest online brokerage firms, i.e., AMTD IDEA, E*TRADE Finl Corp, and Charles Schwab.

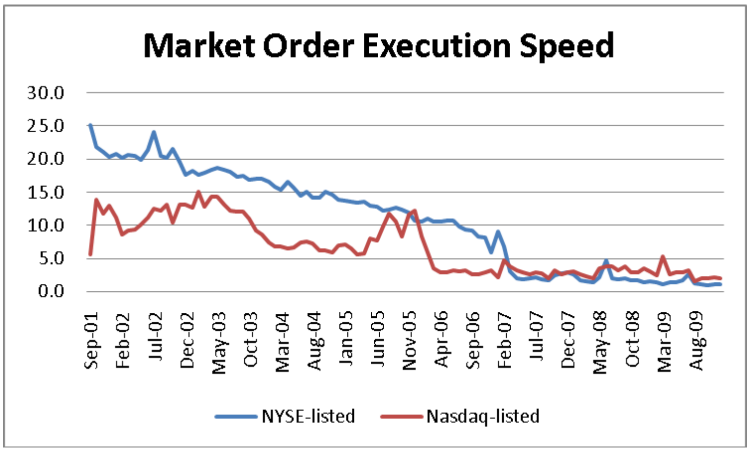
Chart 7: Retail commissions fell and remain low



Source*: Equity Trading in the 21st Century (2010); Barclays Capital Equity Research*

The willingness of HFTs to devote capital to buy when others desire to sell and vice versa, flattens out the price effects of order imbalances and further reduces transactions costs for end investors. Hence, their activity increases price information efficiency and cross-market coordination. Execution time has plummeted, allowing better control of trade execution status for retail investors. Generally speaking, the increasing automation led to a market wide decrease in the speed of execution for small market orders, which greatly facilitates monitoring execution quality by retail investors (*See Chart 8 below*).

Chart 8: Execution times fell



Source*: Equity Trading in the 21st Century (2010); Rule 605 data from Thomson for all eligible market orders*

To sum up, it can be safely argued that the main supporters of the benefits of HFTs focus on the information efficiency of prices, claiming that they improve market quality.

In point of fact, in the case of statistical arbitrage transactions, the impact can only be seen as positive, as their very presence helps to remove arbitrage opportunities more quickly, thus improving the information efficiency of prices and making the market more efficient, even with some of the strategies that have attracted the most criticism from the trade press and academic community, such as flash trading (*See Chapter 3.2.5).*

3.3.1. The fastest predator takes the fattest prey

The above conclusion conflicts with the attacks of the trade press and non-high-frequency operators, as well as some academics and the book by M. Lewis that inspired the writing of this paper[55], who

[55] Michael Lewis (2015), *Flash Boys: A Wall Street Revolt.*

believe that the more aggressive practices employed by many HFTs artificially create situations in which the probability of market inefficiency is high, with the aim to increase their profit opportunities. Michael Lewis states that they engage in a race for speed among themselves to gain the "fattest victim". This statement refers to the technical features that distinguish these traders: the ability to execute orders at such speed that no traditional investor could compete. However, the number of companies and banks with a high-frequency department substantially increased since the beginning of this phenomenon. Hence, HFTs not only profit at the expenses of slow-footed traders, but also compete to arrive first and get the "biggest piece of the cake".

Going deeper into research, some academics demonstrated different ways HFTs are distorting the soundness and stability of the market. R. Jarrow and P. Protter[56] have analysed the market and provided an econometric model representing a frictionless and competitive market, with no bid-ask spreads and perfect liquidity. They divided market participants in two types: ordinary traders (e.g., pension funds and small investors) and HFTs, who differ for speeds of transacting and the market information set which they have access to. The strategies of the ordinary traders are based on their observation of the market; hence they follow a predictable process.

The model shows that "there exist no arbitrage opportunities for ordinary traders in our economy, yet the introduction of high frequency trades both increases market volatility and generates abnormal profit opportunities for the high-frequency traders at the expense of the ordinary traders"[57].

Moreover, they highlight that the abnormal profit opportunities are generated by the trading speed advantage, hence if one removes it, then these abnormal trading profits would disappear, and the market would be fair for every investor[58].

Overall, the major negative impacts of high-frequency systems that many acknowledge are:

I. Information asymmetry
II. Adverse selection
III. Front running
IV. Shadow liquidity
V. Flash crashes

[56] See: "*A Dysfunctional Role of High Frequency Trading in Electronic Markets*" (2011).
[57] R. Jarrow, P. Protter, (2011), *A Dysfunctional Role of High Frequency Trading in Electronic Markets*, 2-3.
[58] This is the principle on which IEX was founded, as will be seen shortly in Chapter IV.

When HFTs execute flash trades, a situation of information asymmetry occurs; there is a gain possible only due to an informational advantage. In fact, having in pre-emption, albeit for a few milliseconds, the possibility of processing an order before it is sent by the ECNs to the markets on which the NBBO is present, allows an HFT two alternatives. Those are undertaking the operations described in the section on flash trading if it believes it to be profitable, or conversely, deciding not to process it on that trading venue.

In addition, knowing that he will not process such an order, he is aware that it will be sent automatically on the market where the NBBO is present; hence, before the order arrives, he can withdraw his buy or sell orders, using this informational advantage as a partial shield from unprofitable liquidity flows.

It has been mentioned above that some specific strategies, called ghost strategies, induce other market participants to make trades that, in the absence of the simulations, they would not have undertaken. From this it is impossible not to recognise the establishment of an adverse selection mechanism to the detriment of less sophisticated traders.

Front running is defined by the SEC as an illegal practice that allows high-frequency traders to make profits in an environment of strong information asymmetry. This practice is mainly exploited by all those traders who engage flash trades, i.e., it allows the HFT to know in advance that a large buy or sell order that does not meet the NBBO requirements is coming to market and gives them the choice of whether to process that order. The negative effect if the trader decides not to process it is a loss of liquidity for the market just when the market needs it. Some also compare this form of market manipulation to the insider dealing[59], i.e., the trader exploits a personal advantage having access to confidential information. As a result, the traditional trader placing the order on the market, falsely expecting to find adequate liquidity, risks obtaining a worse execution price than rationally expected.

When markets are stable and present low volatility, it has been claimed that HFTs provide liquidity, which is a good thing for the market. However, the phenomenon of shadow liquidity is related to the possibility of creating an untrue representation of the real depth of the trading book, i.e., the degree of liquidity in the market, as the "only apparent" liquidity could disappear in a matter of seconds. The

[59]Insider dealing: "a transaction for a person's own benefit, on the basis of and ahead of an order which he is to carry out with or for another (in respect of which information concerning the order is inside information), which takes advantage of the anticipated impact of the order on the market or auction clearing price" (*FCA Handbook*).

presence of this ghost liquidity on the trading book encourages an increase in uncertainty and volatility in the market, prompting often traditional traders to close their orders as quickly as possible.

Eventually, whenever the market is too volatile and there is a lack of sufficient liquidity because HFTs stop providing it, flash crashes will occur, the resulting costs of which are passed on to all market participants.

CHAPTER IV

4. An honest solution to HFT: IEX by Bradley Katsuyama

In the previous chapters of this work, the increased complexity of the financial system and the changing trading mechanisms have been extensively illustrated, together with their main suspicious effects on the markets in the past few years. The introduction of the use of highly sophisticated computer technologies by high-frequency traders has resulted in an intensification of market inefficiencies at the expense of the ordinary investors.

However, a constructive rather than destructive use of these advances has been wished-for by some. In this chapter, the most impactful "solution" to high-frequency trading, Bradley Katsuyama's ground-breaking project, will be described.

In order to better explain the IEX Group case, it is first necessary to introduce the main character of this story. Bradley Katsuyama is a Canadian trader from Toronto who used to work for the Royal Bank of Canada. After starting his career as a trading associate in 2001, Brad was sent by the RBC Capital Markets division to New York, where he gradually became in charge of Energy Equity Trading first, and director of Technology Trading later[60], guiding more than twenty traders. He was quoted by his former colleagues as the "golden child" on the trading floor on One Liberty Plaza, as he managed to be a successful trader while he fully embraced the culture "RBC nice"[61], which is the attitude that characterised the traders of this bank from the other Wall Street aggressive traders.

He has been the first man to notice that the U.S stock market was beginning to behave oddly by 2007 as he observed that heavily traded stocks of a company were inexplicably highly sensitive to his trading orders, as would happen with large orders for infrequently traded stocks, even though there should be plenty of demand and supply to satisfy his orders[62].

Thus began his persistent confrontations with the technical support team first, and the developers then. In all such discussions, what he was told was that the problem was either him (i.e., it was a "user error") or at most the "slow market", i.e., the slow signal transmission due to the distance between the trading floor in New York and the markets, which were now in New Jersey (*See Chapter 3.1.2*).

---

[60]Detailed information about his career is available on the IEX website and on Katsuyama's LinkedIn profile (https://www.linkedin.com/in/bradley-katsuyama).
[61]Michael Lewis (2015), *Flash Boys: A Wall Street Revolt,* 27.
[62]See: Michael Lewis (2015), *Flash Boys: A Wall Street Revolt,* 32.

However, he was not the only one in that department to whom this happened, and no programmer could contradict his evidence that in fact, he was the "news" that made prices move, as these strange variations only occurred when he pressed the Enter button to launch an order.

4.1. The unpleasant discovery

His biggest question was why the market display no longer represented reality as it used to. Whereas at first, it seemed to interest him alone, while everyone else was only focused on the money, by 2009 the issue also began to concern the Bank's direction, which, due to a year of underperforming trades, put him in charge of the electronic trading department, allowing him to conduct his own tests on why, and to what extent, the market was rigged. Hence, together with Rob Park[63], the first partner in the team he started, he came up with some theories to test. As Lewis reports, the "RBC agreed to let his team lose up to $10,000 a day"[64].

They began their experiments from the most obvious place, the public markets, i.e., the stock exchanges directed by NYSE, NASDAQ, BATS and Direct Edge. They realised that, when sending the orders to different exchanges, the percentage of the orders filled at the market prices was low, whereas when they sent the orders to one exchange only instead, the market on the screens resulted real. Moreover, the greater was the number of exchanges on which they sent orders at the same time, the lower was the percentage of the orders filled immediately at the price on their screens. There was one single exception, and it was that no matter how many exchanges they sent orders to, they always managed to get one hundred per cent of the offer on BATS. This issue, initially unexplained, became clearer when they began to observe other factors that could influence the implementation of the orders, such as the distance from their desks inside the World Financial Center to the various exchanges.

4.1.1. Brad's problem at RBC and the increase in "Technical Glitches"

The occurrences that began to attract Brad's notice, even when they were acknowledged by most, were still depicted as "technical glitches". This expression referred to any malfunction in the stock market associated with the obscure and complex mechanisms on which it is based, so as to render incomprehensible instead of clarifying the origin of such events. In fact, when some sort of

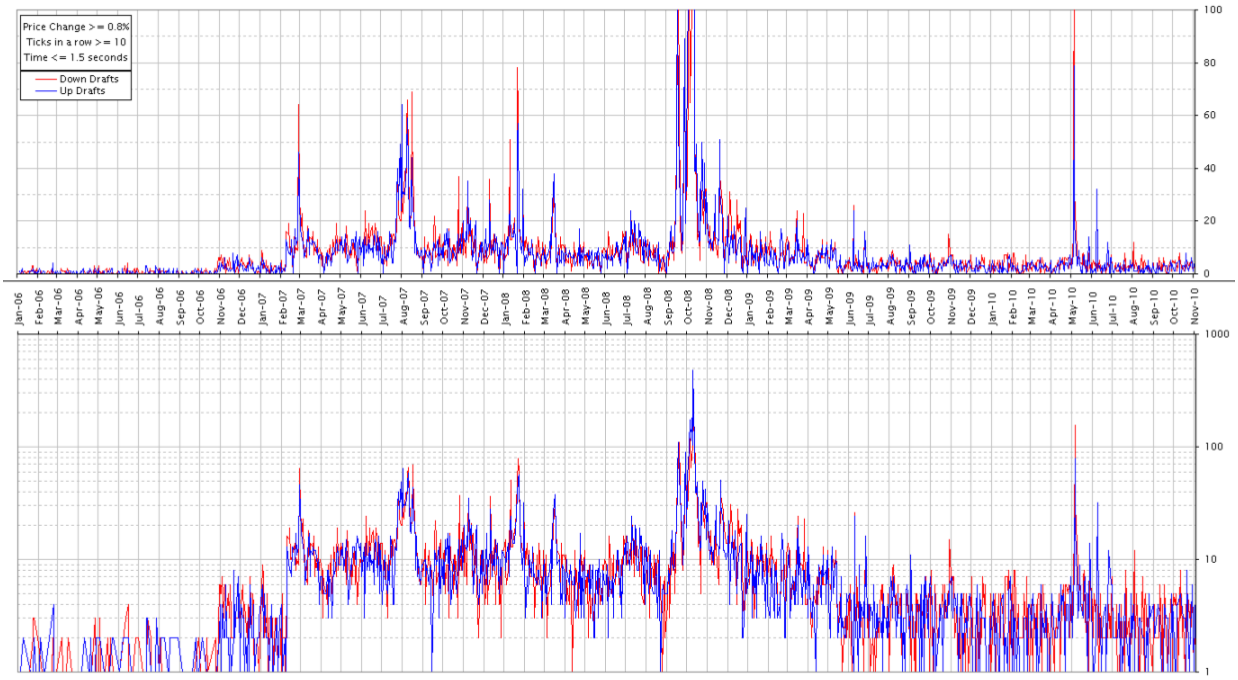
[63] Rob Park is the first person Brad Katsuyama hired. At the time he was the Head of Global Algorithmic Trading for RBC Capital Markets, where his team was instrumental in the creation of client-facing algorithms and smart order routing strategies. Prior to working at RBC, Rob co-founded Quantera Financial Corp., a proprietary high frequency trading firm that specialized in statistical and machine learning algorithms (*IEX website, Leadership and board*).
[64] Michael Lewis (2015), *Flash Boys: A Wall Street Revolt,* 46.

breakdown occurred and Katsuyama and his team personally tackled the head of the stock market, he had no idea of either the cause or how to fix it, and he attributed this to what might better be described as a normal symptom of the growing instability of the markets. Lewis argues that “in a stock market now defined by technology accidents, nothing actually happened by accident”[65].

The loopholes left by regulation in those years, thus enabling the shady practices detailed in the previous chapter, predictably resulted in a series of technical glitches such as flash equity failures on listed equities individual stocks, sometimes also occurring during IPOs[66].

A research conducted by NANEX[24] in 2011, analysed all the potential "mini crashes" in individual stocks which occurred between 2006 and 2010[67]. As summarised in the line graph presented below (*See Chart 9*), the number of up or down drafts considerably increased during the years 2007 and 2008, even before the well-known Flash Crash of May 6th 2010 discussed in the second chapter (*See Chapter 2.1.3*).

Chart 9: daily occurrence of flash equity failures events, between Jan 1st 2006 and Nov 3rd 2010



Source*: NANEX.net*

[65] Michael Lewis (2015), *Flash Boys: A Wall Street Revolt,* 211.

[66] e.g.: “the NASDAQ computer caused $500 million in losses for traders when Facebook's (now Meta) initial public stock offering was launched. The IPO was delayed for 30 minutes on May 18, 2012. Traders could not place, change, or cancel orders. A record 565 million shares were traded when the glitch was corrected (*What is a Flash Crash, thebalancemoney.com*)”.

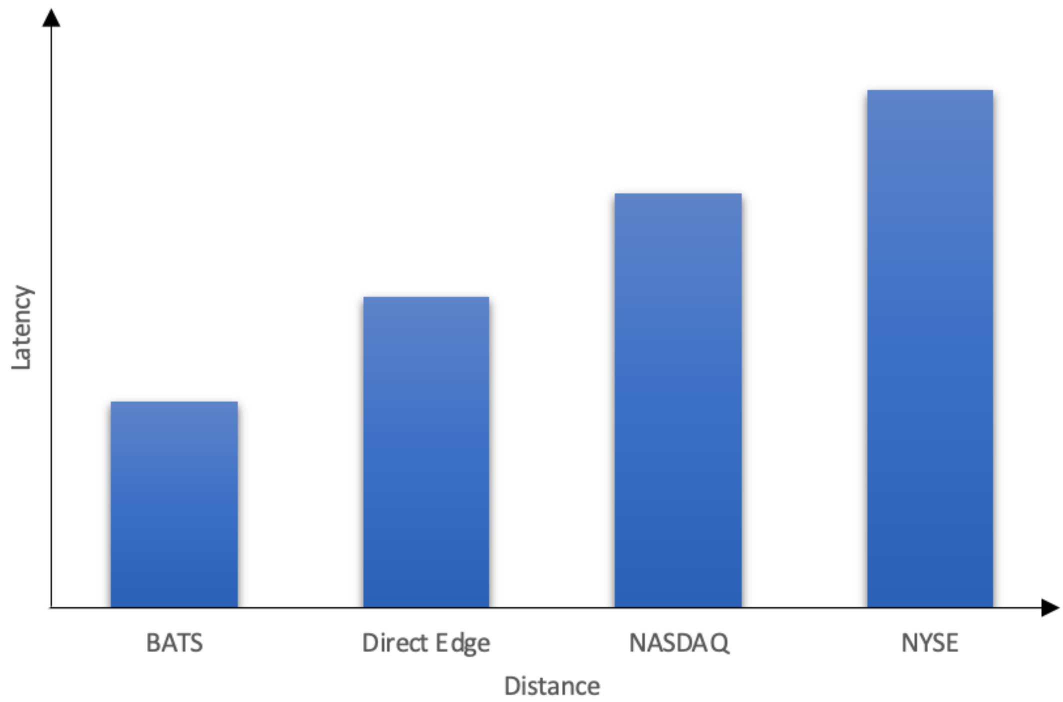
[67] See: http://www.nanex.net/FlashCrashEquities/FlashCrashAnalysis_Equities.html

The parameters used to qualify an event as a down-draft or an up-draft candidate in the research are that the stock had to tick down (up) at least 10 times before ticking up (down) within 1.5 seconds and the price change had to exceed 0.8%.

4.2. Thor mechanism

As they tested multiple combinations of exchanges on which to send orders, delays in the arrival of the orders on the exchanges were noticed by Brad and his team. These time lags were directly proportional to the distance between them. A depiction of this information applied to four of the exchanges is presented in the graph below (*See Chart 10*).

<u>Chart 10: Increments in travel time</u>



Source*: data elaboration from Flash Boys (2015)*[68]

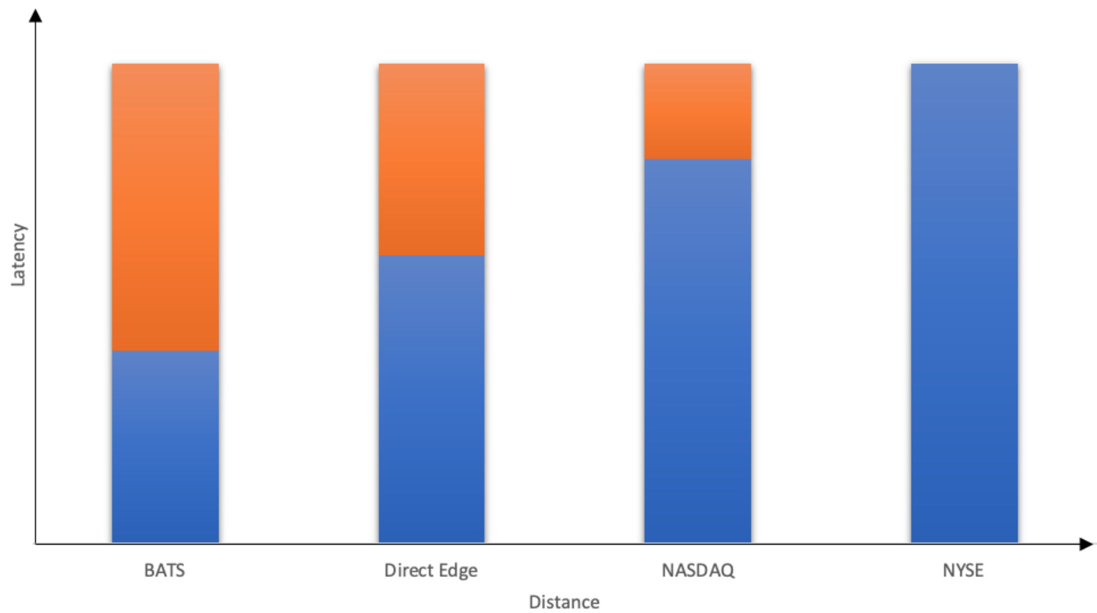
As Lewis reports, the difference in travel time between the nearest exchange, which was in fact BATS, and the furthest, was absurdly small. The signal took only 2 milliseconds to arrive at Weehawken, NJ, where BATS was located, and 4 milliseconds to arrive at NASDAQ, site in Carteret, NJ. Yet this gap drastically changed the outcome of their attempts. The time also had a certain volatility depending on the network traffic and glitches in the pieces of equipment between any two points.

[68] See: Michael Lewis (2015), *Flash Boys: A Wall Street Revolt*, 49.

Therefore, they built a program that would allow them to arrive at all exchanges at the same time and they tested their hypothesis that there were certain market manipulations by speedier operators based on the delays which they and all other investors experienced. Having reached an affirmative result, they raised the dilemma of whether to use this knowledge to join whatever game was being played in the stock market or for some other purpose[69].

The choice has been to initiate some sort of educational campaign and sell this tool to other investors. The name under which this tool was sold is THOR and its aim was to enable institutional and ordinary investors to defend themselves from unfavourable stock price fluctuations.

Chart 11: THOR Mechanism



Source: *data elaboration from Flash Boys (2015)*[68]

With the implementation of THOR, the signal travel time represented in the previous graph (*See Chart 10*), now seemed to be the same on all exchanges (*See Chart 11*). The program, written to build varying lags, had great potential, as it introduced delays that were inversely proportional to the time that had been observed, to ensure that the moment of arrival was identical to all exchanges.

However, it was infrequent that it would function perfectly, and thus problems arose in selling it to the world's biggest money managers. From a technical point of view, the software could not be

[69] Michael Lewis (2015), *Flash Boys: A Wall Street Revolt,* 50.

implemented much, but at the same time, it had exogenous limitations, i.e., the true path of the fiber optic cable along which the orders travelled was unknown and difficult to guess. Hence, THOR was inconsistent because the travel times were never the same. Depending on the traffic, it could take 4 milliseconds as well as 7 milliseconds to reach the NYSE and the time it took any high-frequency trader's signal to travel from the first exchange to the final one was 465 microseconds. In order to interact with the market as it appeared on the screens it was required to arrive at all exchanges within 465 microseconds and considering that the already existing paths along which the signals travelled had incredible deviations to make the routing more intricated, the only way to do that was to build and control their own fiber.

4.3. The creation of IEX

By 2011 it had become clear that Thor's mechanism alone would not be enough to change the entire system, even though the Royal Bank of Canada had become the number-one-rated stockbroker in America[70]. This invention was supposed to be deployed as the backbone of a new exchange to institutionalise fairness in the U.S. stock market.

After two years of development, programming and investor seeking, on Friday 11th October 2013 the group announced on their website the production launch of the IEX exchange (i.e., Investors Exchange) scheduled for 25th October of the same year[71]. IEX began as an ATS (Alternative Trading System[72]) that offered a more balanced marketplace through simpler market structure design and cutting-edge technology to any authorized broker-dealer. As stated on the official website of the platform, "IEX is the first equity trading platform founded entirely by buy-side investors such as mutual funds, hedge funds[73], and family offices"[71].

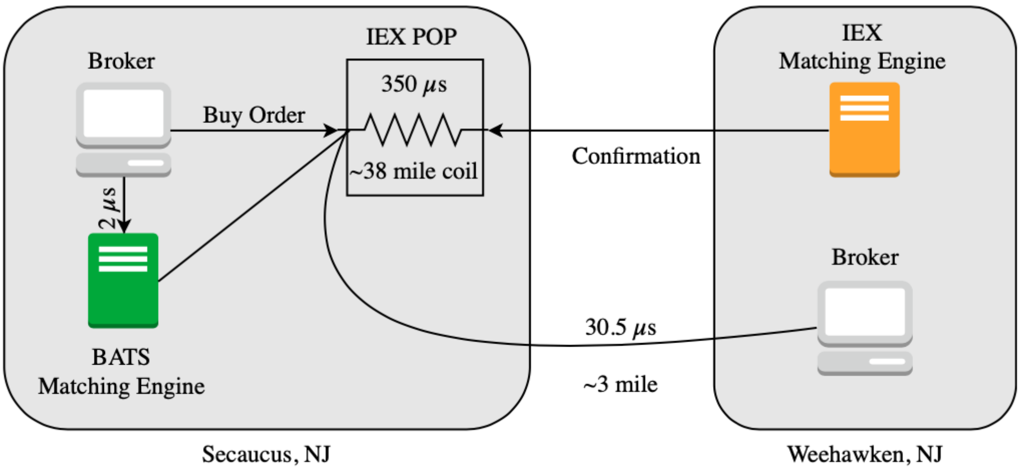
In practice, IEX implemented a "speed bump", primarily to solve the previously mentioned perceived problem of inter-exchange communication. As Mahoney and Rauterberg[74] reported, "the speed bump applies to communications entering and leaving the IEX matching engine and means that when an order arrives at IEX, the system will wait for 350-microseconds to post and/or to execute it".

[70] Michael Lewis (2015), *Flash Boys: A Wall Street Revolt,* 103.
[71] See: https://iextrading.com/trading/alerts/2013/001/
[72] "Alternative Trading Systems (ATSs) are SEC-regulated electronic trading systems that match orders for buyers and sellers of securities. An ATS is not a national securities exchange, and current ATSs are "dark pools". However, an ATS may apply to the SEC to become a national securities exchange" (*Investor.gov*).
[73] In the first round of investors, Lewis mentions Greenlight Capital, Capital Group, Brandes Investment Partners, Senator Investment Group, Scoggin Capital Management, Belfer Management, Pershing Square, Third Point Partners.
[74] P. Mahoney & G. Rauterberg (2017), *The Regulation of Trading Markets: A Survey and Evaluation,* 45.

Furthermore, "counterparties are only notified after a 350-microsecond delay". As a result, "people participating in an order do not learn about the execution for some time, hence a large trader has enough time for its orders to arrive at other exchanges or for IEX to route the remainder of an order to other exchanges".

In order to undermine the strategies of the HFTs, the matching engine was placed at a sufficient distance from the "point of presence", i.e., where traders connect to the IEX. In addition, the possibility of co-locating was not allowed. Being the match engine located in Weehawken, NJ, and the point of presence in Secaucus, NJ, the distance between the two data centres was less than ten miles. Yet, to create a 350-microsecond delay on the exchange, the route of the new optic fibre had to be around 38 miles long. Hence, a small compartment the size of a shoebox was assembled, with miles of fibre coiled inside[75]. A portrayal of this connectivity map is presented below (*See Figure 6*).

Figure 6: IEX Exchange connectivity map



Source*: Hu, Evidence from IEX becoming an exchange, 43*

In April 2015, IEX converted from ATS to lit Electronic Communications Network (ECN), which enabled the display of orders, dissemination of displayed trades, and provision of top-of-book quotes (through the TOPS system)[76].

One year later, as released in the Financial Times, after months of negotiations over the compliance of the unique model of the IEX with U.S. rules, on June 16th 2016, "the Securities and Exchange Commission gave its approval […] putting IEX on an equal footing with the New York Stock

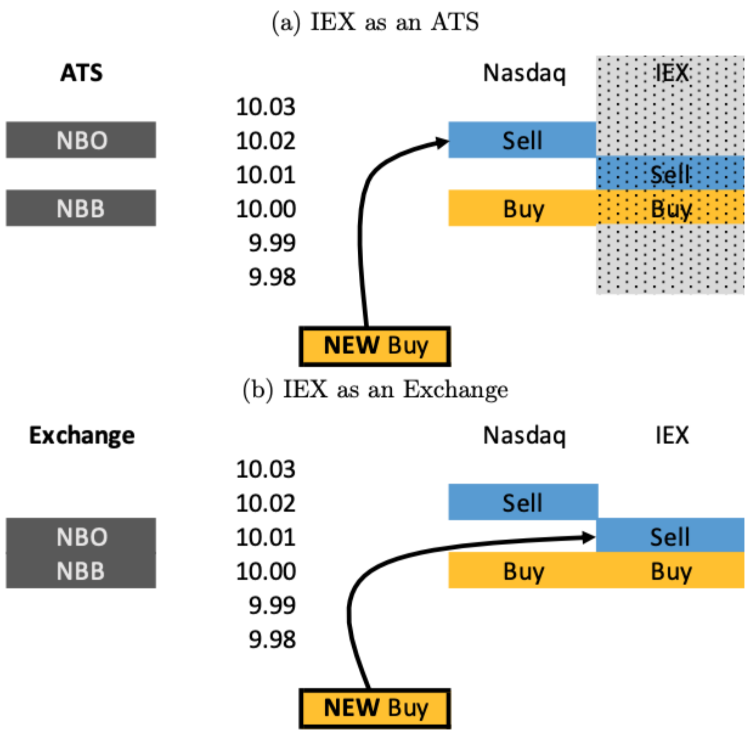
[75] See: N. Bullock & P. Stafford (2016), *US exchanges: the 'speed bump' battle,* Financial Times.
[76] See: https://iextrading.com/trading/alerts/2015/005/

Exchange, Nasdaq, and Bats Global Markets. U.S. regulators also opened the door for other exchanges with speed bumps, potentially ending a decade-long period of prioritising speed in stock trading”[77].

Following the criticism aimed at the exemption of the router from the speed bump during the application process, IEX made modifications to its system, resulting in IEX's own order routing technology being subjected to the 350-microsecond speed bump as well, to comply with the NBBO principle and the Order Protection rule of Reg NMS, i.e. Rule 611, previously discussed in this work (*See Chapter 2.2.3*).

However, before becoming an exchange and obtaining a protected quote, brokers had the option to send buy orders to NASDAQ, even if IEX had the best available offer[78]. Conversely, when IEX became an exchange and obtained a protected quote, brokers were no longer permitted to "trade through" IEX's best offer. To comply with this rule, brokers may route buy orders to IEX. These two scenarios are illustrated in the figure below (*See Figure 7*).

Figure 7: Examples of routing before and after IEX becomes an exchange



Source*: Hu, Evidence from IEX becoming an exchange, 44*

[77] N. Bullock (2016), *IEX trading venue wins battle to become an exchange,* Financial Times.
[78]Since IEX was not an exchange, it lacked the authority to establish the National Best Bid or Offer.

4.3.1. Creating fairness

The overmentioned fairness on the IEX exchange (first as a dark pool) is achieved and preserved by preventing co-location or privileged access to exchange data, not paying kickbacks to banks and brokers that send the orders and limiting the types of orders accepted and operated on this trading venue.

In the above-mentioned launch announcement of October 11th 2013, the emerging dark pool published a detailed book rule, explaining the fair cost structure applied to investors and the characteristics of the various trades.

Unlike the complex fee structures of other exchanges, IEX exchange charged subscribers a standard rate of $0.0009 per executed share (i.e., 9 one-hundreds of a cent per share) matched on the IEX platform, both for adding and removing liquidity, thus removing a disparity in the costs incurred by liquidity makers and takers. Moreover, for shares executed at away trading venues, i.e., shared routed on other exchanges, IEX charged an additional $0.0001 to the away venue cost to the exchange. This simple fee structure can be summarised in the tables below (*See Tables 3 & 4*).

Tables 3 & 4: IEX Pricing Schedule

Executed Shares Matched on the IEX Book	**Fee per executed share**
Standard Rate	$0.0009
Shares Priced below $1.00	0.30% of the total dollar value

Executed shares Routed to Away venue	**Fee per executed Share**
Shares Priced $1.00 or above	Away venue cost to IEX plus $0.0001
Shares Priced below $1.00	Away venue cost to IEX plus $0.0001

Source*: data elaboration from IEXtrading.com*

With regard to orders, there has also been a reduction in the complexity. On the first day, there were only three types of orders allowed: market, limit, and midpoint peg orders[79]. The updated catalogue of orders executed on the IEX exchange today is presented on the official website. The traditional orders explained in the "standard series"[80] section will be reported below.

I. Market Order: as previously explained in this paper, a market order is an order to buy or sell a share at the best market price available at the time the order is executed (*See Chapter 1.2.1*).

II. Limit Order: a limit order is a pre-contingent order to buy or sell a share at a certain price or better. The order is only executed if the market price reaches or exceeds the designated limit price (*See Chapter 1.2.2*).

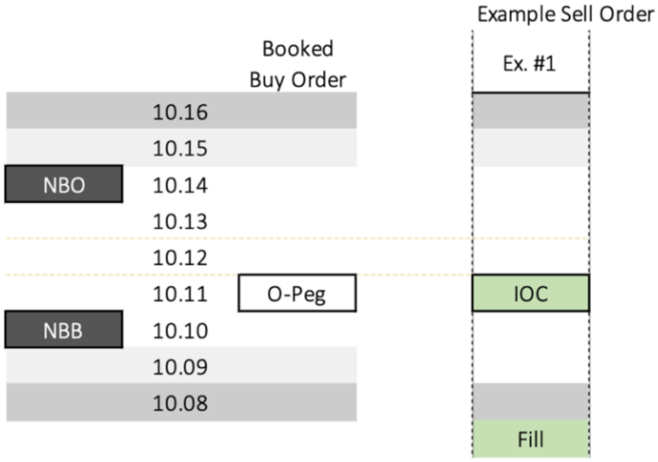
III. Midpoint Peg: an M-Peg order is a non-displayed[81] order type pegged to the midpoint of the NBBO and can be subject to the order's limit price. The focus of a midpoint peg order is to achieve better price execution than a traditional market order, thereby avoiding any significant deviation from the target price. However, it is important to note that the actual execution of a midpoint peg order will depend on the liquidity and depth of the market at the time.

IV. Market Peg: Market Peg is a "non-displayed buy or sell order type that is pegged to the contra-side primary quote[82] (NBO or NBB respectively), with an optional, user-specified passive offset amount. This type of order is designed to make it easier to access the far side of the NBBO and bring the benefits of the IEX Speed Bump to this functionality"[83].

V. Offset Peg: an O-Peg order is a non-displayed order pegged to the National Best Bid for buys and the National Best Offer to sell, with a positive or negative offset amount. This type of order "is designed to bring the benefits of the IEX Speed Bump to trading inside the spread (non-midpoint), helping traders manage adverse selection risk, particularly in stocks with wider spreads"[84]. For further understanding, examples of O-peg sell and buy orders from the book rule are presented in the figures below (*See Figures 8 & 9*).

[79]See: Michael Lewis (2015), *Flash Boys: A Wall Street Revolt,* 178.
[80]See: https://www.iexexchange.io/products/order-types#comparison
[81]An order that is not publicly displayed in the market order book. This type of order is also known as a "dark order" or "hidden order", and it can be advantageous for traders who wish to maintain the confidentiality of their trading intentions or avoid negative price impacts caused by the disclosure of such orders.
[82] The contra-side primary quotes are represented by the NBO (National Best Offer) for buy orders and the NBB (National Best Buy) for sell orders (*Investors.gov*).
[83]IEXexchange.io (2022) *MARKET PEG, Overview*. Link of the pdf available in the other references section.
[84]IEXexchange.io (2022) *O-PEG, Overview*. Link of the pdf available in the other references section.

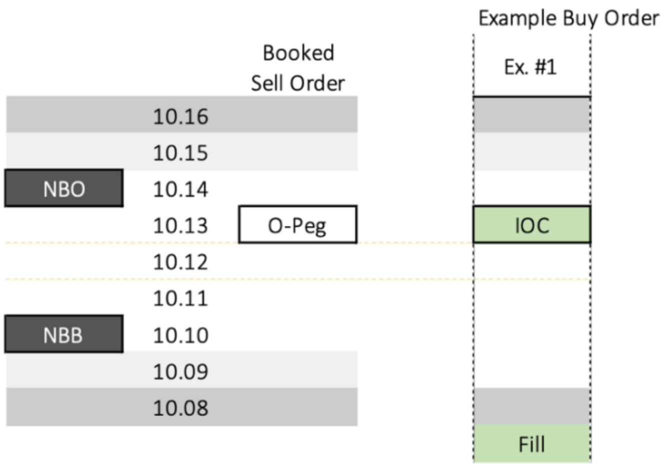
Figure 8: Buy Order with Offset



Source*: O-PEG, Overview, IEXexchange.io*

In this first example, an O-Peg buy order is booked with a positive $0.01 offset from the NBB, which stands at $10.10. Assuming that an IOC[85] sell order with a $10.11 limit crosses the spread, the trade occurs at a resting price of $10.11.

Figure 9: Sell Order with Offset



Source*: O-PEG, Overview, IEXexchange.io*

If, instead, a trader wants to sell a share, they could book a Peg sell order with a negative $0.01 offset from the NBO, which stands at $10.14. In the event that an IOC[85] buy order with a 10.13 limit crosses the spread, trade occurs at a resting price of $10.13.

[85] "An Immediate-Or-Cancel (IOC) order is an order to buy or sell a stock that must be executed immediately. Any portion of an IOC order that cannot be filled immediately will be cancelled" (*Investors.gov*).

4.3.2. Early observations and statistical analysis of the effects on market efficiency

Having regard to the aforementioned features, some observations about the claimed outcomes need to be made. The first remarkable observation is suggested by the preliminary statistics published by the IEX Group on the website one working week after the launch of the platform. On November 4th 2013, in fact, the first statistics of interest were published[86], using the SEC Market Structure Division's ATS statistics as a benchmark for this ATS performance.

Indeed, it is necessary to recall that one of the "symptoms" of the market being rigged was the failure that the investors were experiencing to completely fulfil large orders on the exchanges. Accordingly, smaller orders were quite frequently placed by HFTs on the exchanges to detect liquidity and front-run the investors with high-speed technologies.

The key statistics of interest from the first week of trading, considering the SEC ATS White Paper[87], can be summarised in the tables below (*See Tables 5 & 6*).

Table 5: Average order size

observed measure	IEX	All ATS
Average Order Size	1403	374
Average Aggregate Fill Size	382	232

Source*: data elaboration from SEC ATS White Paper,* 8; *IEXexchange.io*

This table suggests that the average size of the order placed on the IEX exchange was more than triple the size of orders on the remaining ATSs, which results from the absence of the small orders of HFTs discussed earlier. At the same time, a larger average order size does not necessarily guarantee that there is sufficient depth to satisfy them at the NBBO, hence the average aggregate fill size (which measures the average size of trades that were completed in a single transaction by aggregating the quantity of each order involved), is still larger than the average of all ATS (approximately 1.5 times), but the two values show a smaller distance.

[86] See: https://iextrading.com/trading/alerts/2013/003/
[87] See: https://www.sec.gov/files/otc-trading-white-paper-03-2014.pdf

Table 6: Percentages of different size order fills

observed measure	IEX	Median ATS
% of Fills: 100 Shares	61.10%	67.17%
% of Fills: 5,000-9,999 Shares	0.55%	0.18%
% of Fills: >= 10,000 Shares	0.48%	0.08%

Source: data elaboration from SEC ATS White Paper, 11; *IEXexchange.io*

Analysing the percentages in the table above, it can be assumed that in a market where fairness is promoted, improper practices are undermined and clear rules are observed, a higher percentage of submitted orders are fulfilled, given that, as opposed to the presence of HFTs as discussed in the previous chapter of this paper (*See Chapter 3.3.1*), no episodes of sudden lack of liquidity occur, i.e., shadow liquidity. In fact, the median percentage of ATS of small orders that are successfully completed is higher than that of IEX (67.17% compared to 61.10%), whereas the trend is reversed as the order size increases. Indeed, for particularly large orders (>= 10,000 shares) the percentage of fulfilled orders is about six times that of ATSs.

This correlation has also been investigated in a few published studies. Concretely, the last part of this section will be aimed at the discussion of the study conducted by Edwin Hu, from the U.S. Securities and Exchange Commission, on the evidence resulting from IEX on market quality and price discovery[88].

In "Intentional Access Delays, Market Quality, and Price Discovery: Evidence from IEX Becoming an Exchange" (March 2019), Hu uses four data sources, primarily trades and quotes from the NYSE TAQ database from August and September 2016 and top-of-book quotes provided by the IEX Group.

First, summarised predictions of the impact of IEX's speed bump on market quality are presented and the discussion is formalised into three main hypotheses.

I. H0 No Change: if the IEX access delay is *de minimis*[89], then no change in market quality is expected, hence the IEX speed bump has had "no impact" on HFT market-making abilities.
II. H1 a. Market Quality Deteriorates: If the speed bump impedes the provision of low-latency liquidity and obstructs price discovery, it can result in reduced liquidity and limited price discovery for securities with a significant market share on IEX.

[88] See: Edwin Hu (2019), *Intentional Access Delays, Market Quality, and Price Discovery: Evidence from IEX Becoming an Exchange.*
[89] "The staff guidance states that delays of less than one millisecond are at a de minimis level" (*SEC.gov*).

III. H1 b. Market Quality Improves: if IEX introduces a general slowdown in the markets, thereby reducing adverse selection caused by high-frequency traders, securities with a significant market share on IEX might witness an improvement in market quality, without any negative impact on price discovery.

The initial step involves testing the potential influence of the speed bump's introduction on trading activity. This is accomplished by conducting a first-differences regression on stock-day trading activity indicators to estimate the change in average market quality, if any, before and after the IEX exchange phase-in (1)

$$(1)\ MQ_{i,t} = \alpha_i + \tau Delay_{i,t} + \beta X_{i,t} + \varepsilon_{i,t}$$

Considering that α_i is a stock fixed effect which accounts for individual-stock heterogeneity, *X* is a vector of control variables and *Delay* is an indicator for when a given stock begins trading on IEX's exchange, τ captures the average change in market quality (when IEX becomes an exchange)[90]. The results of this first-differences regression are summarised in the table presented below (*See Table 7*).

Table 7: Speed bump impact on trading activity

	IEX Trades (1)	IEX at NBBO (2)	ISO/Order (3)	% ∈ 350μs (4)	Alg. Trd (5)
$\tau Delay$	0.028***	0.019**	−0.218***	−0.438***	0.185***
	(8.645)	(2.508)	(−4.836)	(−2.902)	(4.892)
β					
Market cap	−0.034**	0.013	−0.209***	0.294	−0.424***
	(−2.498)	(0.868)	(−3.427)	(0.558)	(−7.380)
Volume	0.050***	0.014***	0.700***	1.095***	−0.682***
	(17.375)	(8.861)	(53.137)	(15.660)	(−65.357)
Volatility	0.0002***	0.0001**	−0.001***	−0.001	0.0004
	(3.577)	(2.478)	(−2.637)	(−0.979)	(1.614)
Fixed Effects	Stock	Stock	Stock	Stock	Stock
Observations	196,694	196,694	194,218	196,694	196,589
Adjusted R^2	0.635	0.618	0.879	0.626	0.882
Note:	Standard errors clustered by stock-day *p<0.1; **p<0.05; ***p<0.01				

Source*: Hu, Evidence from IEX becoming an exchange, 30*

[90]*Delay* is a dummy variable, respectively equal to zero before IEX becomes an exchange and one after. *X* is a vector of controls: log market cap, log share volume, stock return volatility. Trading activity variables are named in the column headings (*Evidence from IEX Becoming an Exchange, 2019*).

The provided table presents the findings for key trading activity measures of interest, namely: IEX trades, IEX time at NBBO, ISO to Order ratio, the percentage of share volume traded within 350 microseconds of an NBBO update, and the algorithmic trading proxy based on Hendershott's methodology (i.e., the message volume normalized by dollar volume).

The results indicate that following IEX's transition into an exchange, the average number of trades on IEX per day for each stock increased by 28. Additionally, the time spent at the NBBO (National Best Bid and Offer) experiences an approximate 2% increase and the ISO[91]-to-order ratio demonstrates a decrease of around 22 basis points[92]. Among these results, it should also be noted that the percentage of share volume traded within 350 microseconds[93] decreases by approximately 44 basis points. Furthermore, the algorithmic trading proxy registers an increase of roughly 18.5 messages per dollar traded.

All these findings exhibit statistical significance (the p-value is almost always below 0.01), with standard errors clustered at the stock-day level. However, most of the estimates are relatively minor, indicating that a delay of 350 microseconds might be considered negligible in terms of trading access. Although the impact on trading appears to be small, it remains crucial to understand whether such a minimal access delay is advantageous or detrimental to liquidity.

In this regard, estimates of the impact of the IEX speed bump on liquidity are also performed through first-differences regressions, focusing on the dependent variables of quoted and effective spreads[94]. The first-differences estimate (τ) is computed using the following (2)

$$(2)\ L_{i,t} = \alpha_i + \tau Delay_{i,t} + \beta X_{i,t} + \varepsilon_{i,t}$$

As for the previous regressions, the *Delay* is zero before IEX becomes an exchange and is one after. *X* is a vector of controls (log market cap, log share volume, the standard deviation of daily stock returns, and log number of trades on IEX), with additional controls (applied when written YES), including IEX time at the NBBO and AT proxy.

[91]Intermarket Sweep Orders (ISOs) are orders used by high-frequency trading firms in anticipation of future price movements.
[92]One basis point (bps) is equivalent to 0.01%.
[93]It should be considered that such a high speed is associated with HFT.
[94]The quoted spread is the difference between the ask and the bid, whereas the effective spread is the difference between the trade price and the NBBO midpoint price.

The table provided below (*See Table 8*) presents the findings obtained; the columns with odd numbers represent share-weighted spreads[95], while the columns with even numbers represent equal-weighted spreads.

Table 8: Speed bump impact on liquidity

	Quoted Spread		Effective Spread		Quoted Spread		Effective Spread	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
τDelay	0.127	−1.133**	−1.015***	−0.994***	−1.226**	−2.176***	−1.623***	−1.448***
	(0.196)	(−2.099)	(−3.101)	(−3.518)	(−2.414)	(−5.644)	(−6.740)	(−6.583)
β								
Market cap	−27.297***	−24.140***	−19.484***	−20.620***	−28.787***	−25.455***	−20.271***	−20.622***
	(−7.668)	(−8.072)	(−8.692)	(−8.690)	(−8.419)	(−8.728)	(−9.103)	(−8.749)
Volume	−7.433***	−5.639***	−2.044***	0.018	−2.489***	−1.689***	0.305	1.488***
	(−12.124)	(−14.082)	(−6.802)	(0.058)	(−2.976)	(−3.165)	(0.650)	(3.307)
Volatility	−0.005	−0.004	−0.001	−0.005	−0.002	−0.002	0.0003	−0.003
	(−0.798)	(−0.632)	(−0.151)	(−0.857)	(−0.382)	(−0.337)	(0.065)	(−0.578)
IEX trades	1.279***	1.135***	0.120	−0.319***	1.011***	0.831***	−0.044	−0.295***
	(5.521)	(7.513)	(1.100)	(−2.876)	(4.663)	(6.483)	(−0.436)	(−2.906)
Fixed Effects	Stock	Stock	Stock	Stock	Stock	Stock	Stock	Stock
Controls	No	No	No	No	Yes	Yes	Yes	Yes
Observations	160,501	160,496	160,501	160,501	160,074	160,069	160,074	160,074
Adjusted R^2	0.836	0.848	0.849	0.816	0.843	0.856	0.854	0.824

Note: Standard errors clustered by stock-day *p<0.1; **p<0.05; ***p<0.01

Source*: Hu, Evidence from IEX becoming an exchange, 32*

Across all the measures, except for the share-weighted average quoted spread, i.e., *τDelay* column (1), there is a decrease of approximately 1-2 basis points in the spread measures. As E. Hu argues[96], considering that spreads are relatively small (the average quoted or effective spread is around 50 basis points), this translates to an estimated improvement of liquidity by roughly 2% for the average stock.

To further investigate the effect on liquidity, an additional investigation can be made. According to previous theoretical predictions, certain latency arbitrage strategies are believed to impose a cost on liquidity provision by means of adverse selection. To determine if spreads decrease because of changes in adverse selection, it is necessary to break down the effective spreads into two components: the non-informational component (i.e., realized spread[97]) and the adverse-selection component (i.e.,

[95]To compute the share-weighted spreads, E. Hu aggregated spreads to the stock-day level.
[96]See: Edwin Hu (2019) *Intentional Access Delays, Market Quality, and Price Discovery: Evidence from IEX Becoming an Exchange,* 17.
[97]"Realized spreads are measured as the spread from the trade price to a subsequent midpoint price, normalized by trade direction" (*Evidence from IEX Becoming an Exchange, 2019*).

price impact[98]). It is reported by E. Hu that event-time realized spreads are more reliable estimates of the costs of liquidity provision. Hence, the table presented below (*See Table 9*) exhibits estimates of the effects of IEX's speed bump on realized spreads and price impacts.

Table 9: Spread decomposition in realised spread and price impact

(a) Realized Spread – First-differences

	(1)	(2)	(3)	(4)	(5)
τ	−1.379***	−1.299***	−1.232***	−1.180***	−1.152***
	(−7.034)	(−7.002)	(−6.480)	(−6.050)	(−5.918)
FE	Stock	Stock	Stock	Stock	Stock
Ctl.	Yes	Yes	Yes	Yes	Yes
Obs.	160,074	160,074	160,074	160,074	160,074
Adj. R^2	0.801	0.776	0.750	0.724	0.699

(b) Price Impact – First-differences

	(1)	(2)	(3)	(4)	(5)
τ	−0.069	−0.149	−0.216*	−0.268*	−0.296*
	(−1.048)	(−1.487)	(−1.690)	(−1.813)	(−1.847)
FE	Stock	Stock	Stock	Stock	Stock
Ctl.	Yes	Yes	Yes	Yes	Yes
Obs.	160,074	160,074	160,074	160,074	160,074
Adj. R^2	0.384	0.457	0.498	0.520	0.541

Source*: Hu, Evidence from IEX becoming an exchange, 34*

As for in Table 8, also in Table 9 (a) and (b) the first-differences estimate (τ) is computed using the specification (2). Additionally, column numbers (n) correspond to the spread measure computed based on the n^{th} tick after each trade.

From this table, it is revealed that on average, there is a decrease of approximately one basis point in realized spreads across all time horizons, along with a decrease of around 0.2 basis points in adverse selection. The estimates for price impacts are statistically significant, indicating a deviation from zero after the third trade.

Collectively, these findings suggest an enhancement in liquidity following IEX's transition into an exchange. The results align with the notion, provided by E. Hu[99], that reduced speed restrains market makers from promptly updating their quotes, resulting in lower realized spreads, while simultaneously alleviating adverse selection through lower price impacts.

4.4. Goldman Sachs: an ethical change

This paragraph will explain the particular role Goldman Sachs played in the rising exchange in 2013 and the ethical choice it made, which helped IEX Group in the "creating fairness" project undertaken

[98]"Price impacts are defined as either the difference between the effective and realized spreads, or equivalently, the midpoint to subsequent midpoint spread, normalized by the trade direction" (*Evidence from IEX Becoming an Exchange, 2019*).
[99] See: Edwin Hu (2019), *Intentional Access Delays, Market Quality, and Price Discovery: Evidence from IEX Becoming an Exchange.*

years earlier. To better understand the bank's attitude prior to this change, however, it is first necessary to recall the case, narrated in Flash Boys[100], concerning Sergey Aleynikov.

Lewis tells the story of this young Russian immigrant who arrived in New York in 1990 and started his career on Wall Street almost by chance, thanks to his special writing code skills. In 2007, he started to work at Goldman Sachs as a programmer, at the very beginning of the most serious financial crisis of the century. During those years, the HFT firms were arising, and they started to compete on speed to grab the largest profits.

However, Goldman Sachs was not the "fastest predator"[101] since the stock market had become a war of robots, and their robots were slow compared to the other ones. In 2008, the position of Goldman Sachs in the world of high-frequency trading, referred to by the author as "insecure"[102], only granted the firm $300 million in profits from these practices, as nothing compared to the top 10 HFT firms, among which was Citadel with $1.2 billion revenues, which had newly built and more efficient codes.

Solving these technical problems to increase their speed and building a "ticker plant" (i.e., a software to translate and summarise data from all the exchanges as it happened on the SIP[103], but faster than the latter), were the reasons why he was hired. The issue was that the existing code was an amalgamation of 60 million lines and 15 years of attempts to fix it, and Aleynikov's bosses would not let him start a new code from scratch as they had no long-term interests.

Hence, for almost two years, the programmer struggled to fix the code resorting to open-source software[104] for patching material. Later, when Sergey decided to leave for a new job offering him to build a new platform[105], he emailed himself some open-source code he had been working on and for that, he was convicted to spend eight years in a federal prison.

Beyond the various considerations that have been made about the handling of the trial and the jurors' lack of experience in computer programming, the implication that must be highlighted in this story is that at Goldman Sachs, not even the higher levels (such as Sergey's former boss, Adam Schlesinger),

---

[100] See: Michael Lewis (2015), *Flash Boys: A Wall Street Revolt,* Chapter 5.
[101]See *Chapter 3.3.1. "The fastest predator takes the fattest prey".*
[102]Michael Lewis (2015), *Flash Boys: A Wall Street Revolt,* 140.
[103]See *Chapter 2.2.2.*
[104]Open-source software are developed by collectives of programmers and made freely available on the Internet. The tools and components used are then made available so that everyone can adapt and benefit from them. According to Goldman Sachs policy, instead, everything that was touched to restore its code, automatically became its property.
[105]He accepted a job from an HFT firm, Teza Technologies, run by Misha Malyshev.

understood the functioning of the technology they were trying to improve. Hence, this was what really made the firm slower and less profitable than the others.

The ethical change referred to in the title of this section, is the one undertaken in 2013, a few weeks after IEX's debut as ATS (*See Chapter 4.3*), by two new partners at Goldman Sachs, Ron Morgan and Brian Levine. Lewis reports that weeks after the launch, in fact, "it was clear that banks were not sending their orders on IEX"[106], but of all of them, Goldman Sachs managers, concerned about the poor performance of their HFT department, which was characterised by frequent technical glitches, decided to make a decision that was both moral and strategic.

On December 19th 2013, around 3:10 PM, Goldman Sachs sent what is recalled as "the first different-looking stock market order"[107]. Two new people at the firm had been given a new authority and used it to take a longer-term approach, as opposed to what had been done in earlier years, which was narrated in this paragraph.

Goldman Sachs' decision to start sending orders on the IEX, in addition to having a direct impact on increasing the exchange's market share, was an acknowledgement to investors that this new trading venue was the key to understanding the overly complicated market described and analysed throughout this paper.

Indeed, statistics[108] published by the IEX Group regarding that date show that the trust it received had also been rewarded with increased fairness. On December 19th, 92 out of 100 of those trades was traded at the midpoint (i.e., "the fair price"), compared to only 17% of the orders on the other dark pools and even a smaller proportion on the public exchanges.

4.5. What happened next: IEX nowadays

In light of the most important and certainly the most controversial market structure development under discussion in this chapter, it is reasonable to question what further evolution has taken place and whether other proposals or rules for deliberate exchange access delays have been advanced. Therefore, before ending the paper with an overview of the current framework of the IEX, three

---

106 Michael Lewis (2015), *Flash Boys: A Wall Street Revolt,* 214.
107 Michael Lewis (2015), *Flash Boys: A Wall Street Revolt,* 240.
108 See: Michael Lewis (2015), *Flash Boys: A Wall Street Revolt,* 242.

concrete proposals, that followed the approval of IEX as an exchange in 2016 will be referred as follows.
The first proposal was made in June 2016 when the SEC approved[109] NYSE Arca's Discretionary Peg (DPEG) order. This order was similar to IEX's DPEG order[110], which aimed to protect investors from "crumbling quotes".

However, it is reported by E. Hu[111] that IEX believes that the NYSE Arca DPEG might not effectively protect investors because IEX's algorithm is specifically calibrated for its own data centre latencies, so it may not work optimally for NYSE Arca's system. Moreover, the NYSE still provides co-location and direct feed services that give an advantage to fast traders. Ultimately, NYSE Arca lacks a "speed bump" to adjust its pegged orders before high-frequency traders take advantage of outdated quotes.

A second proposal came in January 2017, when NYSE announced[112] that it would seek approval to convert its NYSE MKT floor cash equities exchange into a fully automated electronic exchange (i.e., NYSE American) with a DPEG and a speed of 350 microseconds for incoming orders, outgoing data, and external routing. NYSE American resembles IEX in its 350-microsecond speed and DPEG orders, however, it also offers co-location, charges for direct feed access, and pays rebates to designated market makers. Hence, unlike IEX, NYSE American does not object to HFT techniques.

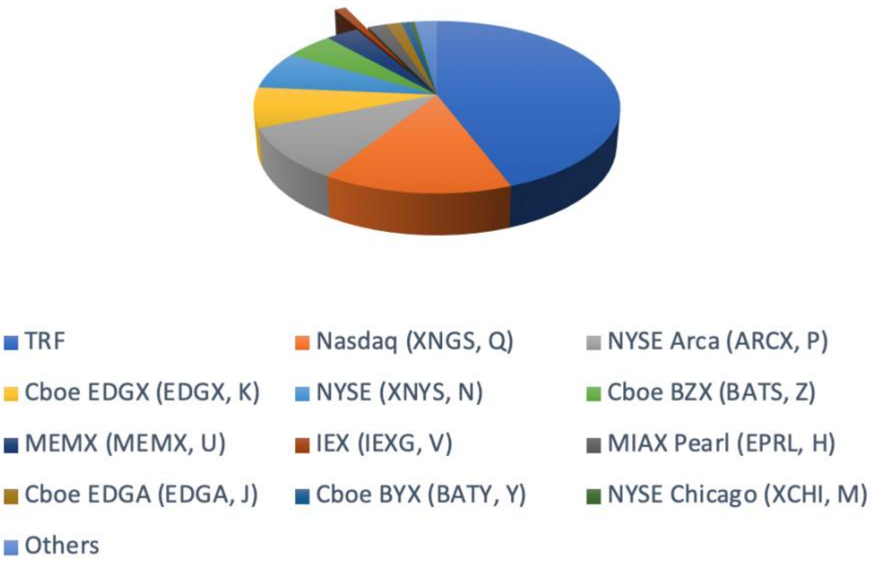
Finally, on November 30th 2016, NASDAQ introduced a new order type called "extended-life order" (ELO), which required a minimum resting time of one second without any cancellations or changes. The purpose of this proposal[113], which was later approved by the SEC on July 7th 2017, was to prioritize the execution of ELOs over other displayed orders at the same price. However, it has been argued by some that implementing minimum resting time policies would raise costs for high-frequency market makers, without effectively protecting traditional market makers from adverse selection.

[109] See: https://www.sec.gov/rules/sro/nysearca/2016/34-78181.pdf
[110] A Discretionary Peg rests outside the NBBO and is willing to step to the midpoint–except when the Signal is on. More information available on https://www.iexexchange.io/order-types/d-peg
[111] See: Edwin Hu (2019) *Intentional Access Delays, Market Quality, and Price Discovery: Evidence from IEX Becoming an Exchange,* 52.
[112] See: https://www.sec.gov/rules/sro/nysemkt/2017/34-79998.pdf
[113] See: https://www.sec.gov/rules/sro/nysemkt/2017/34-79998.pdf approved in July 2017

Having established once again, therefore, that IEX Exchange is the most revolutionary and effective project in terms of the projection of principles, this paper will be concluded with some hints at the latest market statistics at the time of this writing, provided by IEX Group[114].

The following pie chart (*See Chart 12*) represents the U.S. equity market trading scene, observed through real-time market shares of the exchanges (May 24th 2023) on IEX's official website.

Chart 12: Real-time market share on May 24th 2023



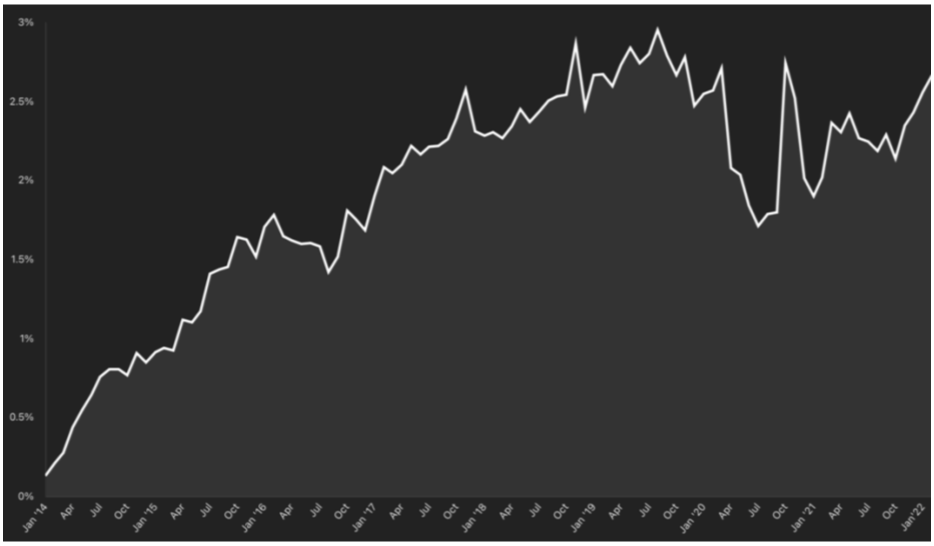
Source*: data elaboration from IEXtrading.com*

The above chart indicates that the current market share of the IEX Exchange is just above 2% (it is 2.046% precisely). Moreover, looking at the recent statistics[115], it can be observed that in the last month, IEX reached the peak of 2.5% market share on May 4th.

Finally, the line graph presented below (*See Chart 13*), also taken from the exchange's statistics, displays the historical evolution of IEX's market share.

[114] See: https://iextrading.com/stats/
[115] See: https://iextrading.com/stats/

Chart 13: IEX market share 2014-2022



Source*: IEXtrading.com*

Apart from a brief downward spike during the year 2020, the market share of IEX exchanges shows a positive increase from the time of creation until recent times. This constant upward trend is further proof of the success of the project born from the intuition of Bradley Katsuyama and suggests good chances for further growth. All of this demonstrates that the efforts to create a trading platform based on transparency have solicited greater trust in large institutional investors such as Goldman Sachs and all the supporters of IEX Group. This has and will potentially continue to change the market for the better, thus rendering it understandable again, despite its cumulative overcomplexity of the past decades.

Conclusion

In conclusion, this thesis has explored the impact of high-frequency trading (HFT) on market complexity and the creation of IEX as an honest solution to address the challenges posed by HFT.

Beginning from the examination of the trading scene before the emergence of HFT, with an overview of different auction markets, such as NYSE and NASDAQ, and the various types of existing orders, including market orders and price-contingent orders, the dissertation delved into the evolution of the U.S. stock market and trading strategies, highlighting the rise of algorithmic trading, HFT, dark pools, and the Flash Crash of May 6th 2010. This particular episode raised some concerns about the regulatory framework governing the U.S. stock market, namely the role of the SEC, the limits of the SIP (Securities Information Processor), and the implications of the implementation of Regulation NMS and Regulation of ATS.

Getting into the specifics of the technical characteristics of HFT, it turned out that the combination of low latency and co-location makes HFTs key players in equity markets. As a matter of fact, they are willing to invest exorbitant amounts of money in order to realise and benefit from the fastest possible fibre optics, and the Spread Network's fiber line has conveyed the idea of the importance of milliseconds. These new operational strategies are based on a speed that neither an ordinary investor nor any algorithmic trading system, can match. At the same time, HFTs have been able to create and launch on the trading venues, new types of orders characterised by specific conditions designed to their advantage, as it happens with the post-only order and the hide-not-slide order.

It has been outlined how HFTs seek to maximise the economic return from the time and technical-computational advantages they hold over the rest of the market, with a series of ingenious and intricate strategies, all of which seek to exploit small market inefficiencies and, in some cases, aim to exploit market liquidity with the mere purpose of detecting it. The revealed competition among HFTs on the notion that the fastest predator tends to take the fattest prey potentially leads to distortions in market fairness. Indeed, the discussion of various experimental studies on this subject has highlighted certain widely acknowledged effects on market efficiency, such as information asymmetry, adverse selection, front running, shadow liquidity and flash crashes.

The final chapter focused on the case of IEX as an "honest solution" to the challenges associated with HFT. The chapter introduced Brad Katsuyama, the founder of IEX, and his journey in discovering the rigged nature of the market. To solve the market discrepancies they were testing between exchanges, Katsuyama and his team's initial proposal was to implement a slow-down mechanism, THOR, which introduced delays that were inversely proportional to the time being observed, to ensure that the instant of arrival was identical across all exchanges.

However, the software had exogenous limitations, namely that the true path of the fibre optic cable along which the orders moved was unknown and difficult to estimate, which made the travel times, and therefore THOR, inconsistent. To interact with the market as it appeared on the screens, it was necessary to arrive at all the exchanges within 465 microseconds by building and controlling a new fibre line.

IEX was approved by the SEC as a registered exchange in 2016 and its measures to foster fairness in the market include a clear and equitable fee structure for liquidity makers and takers, limited types of orders accepted, and the absence of co-locating possibilities.

The statistics released after the first week of trading on IEX seem to indicate that this market was "fairer" compared to other exchanges. To support this notion, a statistical analysis conducted by Edwin Hu, a member of the SEC, has been examined to determine the effect of IEX's introduction of the speed bump on market efficiency. The first-differences regressions which tested the effect of the treatment variable (i.e., "Delay") on market quality, liquidity, and quoted and effective spreads appear to confirm a positive impact of the speed bump (rejecting the null hypothesis in favour of hypothesis H1-b).

Goldman Sachs' ethical change and its decision to direct orders to the IEX exchange, showcased the trust and validation of IEX as the right choice for fair and ethical trading. The status of IEX as a market player, with a 2% market share in U.S. equity trading and consistent growth since its foundation, underscores its ongoing relevance and success.

In conclusion, the creation of IEX has provided a significant and impactful proposal to redeem the financial markets from the challenges posed by HFT. By introducing fairness, transparency, and ethical practices, IEX has set a precedent for market integrity and trust. While subsequent proposals brought by others have attempted to address HFT-related concerns, IEX remains a pioneering force in reshaping the landscape of equity trading. As the market continues to evolve, IEX's dedication to honesty and fairness positions it as a key player in promoting a more equitable and efficient financial ecosystem.

References

A. Kirilenko, A. Andrei et al. (January 6, 2017) *The Flash Crash: High-Frequency Trading in an Electronic Market.* Journal of Finance, Forthcoming.

A. Kirilenko. A. Lo (2013). *Moore's law versus Murphy's law: Algorithmic trading and its discontents.* Journal of Economic Perspectives, 27(2), 51-72.

A. Puorro, Banca d'Italia (September 2013). *Questioni di Economia e Finanza. High Frequency Trading: una panoramica*

C. Comerton-Forde, V. Grégoire, & Z. Zhong (October 2019), *Inverted fee structures, tick size, and market quality.* Journal of Financial Economics, 141-164.

Congress, Public Law 94-29 94th*. Securities Acts Amendments of 1975.*

E. Boehmer, K. Fong, & J. Wu (March 2012). *International evidence on algorithmic trading*. In AFA 2013 San Diego Meetings Paper.

E. Budish, P. Cramton, & J. Shim (February 3, 2015), *The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response*.

E. Hu (March 15, 2019) *Intentional Access Delays, Market Quality, and Price Discovery: Evidence from IEX Becoming an Exchange*.

F. Fabozzi, S.M. Focardi, & C. Jonas (2011). *High-frequency trading: Methodologies and market impact.* Review of Futures Markets, 9(Special Issue), 7-38.

J. Hasbrouck, G. Saar (2013), *Low-latency trading*. Journal of Financial Markets, Volume 16, Issue 4, Pages 646-679.

J. J. Angel, L. Harris, & C. S. Spatt (February 23, 2010), *Equity Trading in the 21st Century*. Marshall School of Business Working Paper No. FBE 09-10.

P. G. Mahoney, G.V. Rauterberg (April 19, 2017)*, The Regulation of Trading Markets: A Survey and Evaluation.* Virginia Law and Economics Research Paper No. 2017-07.

R. A. Jarrow, P. Protter (June 29, 2011), *A Dysfunctional Role of High Frequency Trading in Electronic Markets*. Johnson School Research Paper Series No. 08-2011.

R. Michie (1986). *The London and New York Stock Exchanges, 1850–1914*. The Journal of Economic History, 46(1), 171-187.

R. Vedapradha, R. Hariharan et al. (2023). *Algorithm trading and its application in stock broking services.* In E3S Web of Conferences (Vol. 376). EDP Sciences.

T. Chordia, R. Roll, & A. Subrahmanyam. (2008). *Liquidity and Market Efficiency.* Journal of Financial Economics, 87.

T. Hendershott, C. Jones, & A. Menkveld (2011 February*), Does Algorithmic Trading Improve Liquidity?* The Journal of Finance, Vol. 66: 1-33.

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

U.S. Securities and Exchange Commission (September 2010). *Findings Regarding the Market Events of May 6, 2010.*

Z. Bodie, A. Kane, & A. Marcus (2013). *Essentials of investments: Global edition.* McGraw Hill.

Other References

Britannica encyclopaedia: https://www.britannica.com/topic/New-York-Stock-Exchange

CBOE: https://www.cboe.com/us/equities/market_statistics/book/aapl/

Financial Conduct Authority: https://www.handbook.fca.org.uk/handbook/MAR/1/3.html

Financial Times (1): https://www.ft.com/content/f809ab30-e21c-11e5-9217-6ae3733a2cd1

Financial Times (2): https://www.ft.com/content/9d97bef0-34a9-11e6-ad39-3fee5ffe5b5b

IEX files (1) https://assets-global.website-files.com/635ad1b3d188c10deb1ebcba/63c7dba8f9d8c7411b1837cc_IEX_Market_Peg_Order.pdf https://iextrading.com/trading/alerts/2015/005/

IEX files (2): https://assets-global.website-files.com/635ad1b3d188c10deb1ebcba/63c7dbb4a3a64a1582bf0cc4_IEX_Offset_Peg_Order.pdf

IEX.io (1): https://www.iex.io/leadership-and-board/rob-park

IEX.io (2): https://www.iexexchange.io/products/order-types#comparison

IEX.io (3): https://iextrading.com/trading/alerts/2013/003/

Investor.gov: https://www.investor.gov/introduction-investing/investing-basics/how-stock-markets-work/market-participants

LinkedIn: https://www.linkedin.com/in/bradley-katsuyama

NANEX (1): http://www.nanex.net/FlashCrashEquities/FlashCrashAnalysis_Equities.html

NANEX (2): http://www.nanex.net/FlashCrashEquities/FlashCrashAnalysis_Equities.html

NASDAQ: https://www.nasdaq.com/solutions/nasdaq-co-location

NYSE: https://www.nyse.com/data/cta

SEC.gov (1): https://www.sec.gov/rules/sro/nyse/2010/34-62732.pdf

SEC.gov (2): https://www.sec.gov/files/otc-trading-white-paper-03-2014.pdf

SEC.gov (3) https://www.sec.gov/rules/sro/nysearca/2016/34-78181.pdf

SEC.gov (4): https://www.sec.gov/rules/sro/nysemkt/2017/34-79998.pdf

SEC.gov (5): https://www.sec.gov/rules/sro/nysemkt/2017/34-79998.pdf approved in July 2017

The Balance Money: https://www.thebalancemoney.com/what-is-a-flash-crash-3306184#toc-some-past-flash-crashes

Wall Street Mojo: https://www.wallstreetmojo.com