Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

Facultad de Ciencias Económicas y Empresariales

TRABAJO FIN DE GRADO EN
(Doble Grado Internacional en Administración de Empresas y en Economía)

HIGH FREQUENCY TRADING: AN OVERVIEW

Pamplona-Iruña, 19 de mayo de 2019

Autor: Joxe Mari Barrutiabengoa Ortubai

Tutora: Ana González Urteaga

Módulo: Finanzas

ABSTRACT

High Frequency Trading has been a recurring theme in financial debates. Fuelled by a considerable increase in worldwide trading volume and the sudden appearance of flash crashes, these trading strategies have drawn tremendous public attention in recent years. However, in order to form a well-founded opinion on the topic, and to be able to understand the consequences that ultra-fast trading provokes in financial markets, a deep research on the topic is necessary. Is not unusual to find inaccurate or bias opinions on the subject in the media or among several market participants, that lack any theoretical or empirical basis. In other words, the academic research on High Frequency Trading, and its consequences in financial markets, is limited and usually ignored. In consequence, the aim of this paper is to provide an up to date overview on High Frequency Trading, including definitions, key concepts, historical background, strategies, and its positive and negative consequences. Furthermore, by analysing certain proposed solutions to regulate or monitor ultra-fast trading activity, the paper contributes to the ongoing discussion on the subject, evaluating which could be the pros and the cons of each alternative, and which is their effectiveness in mitigating the problems that High Frequency Trading induces in financial markets. Hopefully, after reading this paper, the reader will have a more accurate insight on the topic.

Keywords: High-Frequency Trading, Algorithmic Trading, Latency, Trading Strategies, Frequent Batch Auctions.

TABLE OF CONTENTS

1. INTRODUCTION

During the recent era, major changes have taken place in international financial markets. Annual turnover, defined as trading volume divided by the outstanding marketable shares of U.S stocks, increased from 15% in 1951 to 250% in 2010, raising the concerns of various markets participants. As Bogle (2012) mentioned, the optimal balance that investment and speculation cultures have sustained for years in capital markets is no longer in place, and investors need to adjust to this new reality. In his words, the new model of capitalism is detrimental for the investing public and beneficial for the most advance sector participants, that obtain significant benefits at the dollar by dollar expense of individual investors.

On the other hand, other experts (see Lattemann et al., 2012; Brogaard et al., 2014) believe that the increase in trading volume is justified by the evolution of algorithmic and computer trading, and in consequence, the expansion of new investment techniques, per se, do not represent any threat for the individual and most vulnerable investor.

However, it is irrefutable that a growing conflict exists between the short-term culture of science, and the long-term culture of humanities. Put another way, a clash between the intellectual, the historian, or the philosopher investor, and the statistician, the technician, or the alchemist investor (see Bogle, 2012). Accepting that providing capital to businesses is the principal aim of financial markets, a trading volume of 300 trillion dollars, 200 times higher than the equity capital that is given to businesses, could be sufficiently significant to justify the fear that experts in the field, like the already mentioned Bogle, Munger, or Buffet, have expressed in recent years (Crippen, 2014). As Beckhart and Keynes (1936) mentioned, "When the stock market takes on the attitude of a casino, the job is likely to be ill done", and when the largest financial rewards are received by investors that may extract value from the society, or at least do not make any clear contribution, the concerns are justified.

In this context, one of the most controversial changes that we have experimented in recent years is the blooming of what today are called as "High-Frequency traders", which represent almost the 50% of the intra-day trading in the world's most important stock markets (see Bogle, 2012). At first sight, high frequency traders provide valuable services to the investor, as an increase in market liquidity or a substantial enhancement in price discovery, but nanosecond trading may also create significant inequities in financial markets, as a decrease in markets' depth or an important increase in market instability (see Glosten and Milgrom, 1985; Cartea and Penalva, 2012).

In the following lines, the role of High Frequency traders will be examined, analyzing the positive and negative consequences that they generate in financial markets, and ultimately, in the average investor. So, the paper is oriented towards introducing the reader to the world of High Frequency Trading (HFT, hereafter). In other words, by providing meaningful information about the different strategies that ultra-fast trading firms employ, analyzing the positive and negative consequences of the activity, and examining the possible alternatives that may exist to mitigate the problems that HFT provokes, the aim of this report is to conduct an extensive and complete overview of the sector, creating curiosity and building a well-founded opinion of the subject among readers.

In short, the underlying conclusion of this paper is that regulators need to work to preserve the benefits of HFT while mitigating the problems. HFT will always generate debate, but ensuring a level playing field between all market's participants, stablishing appropriate risk management mechanisms (as measures to combat liquidity problems at market stress periods), ensuring supervision and communication (in order to increase trust), and preserving the economic rationale behind financial markets, the "beast" can be tamed.

This research contributes to the existing literature in two ways. On the one hand, it conducts a complete summary of the existing HFT related literature, highlighting the importance of conducting future research about the topic. On the other hand, it examines which are the main problems related with ultra-fast trading, and at the same time, it explores the adequacy of the different alternative solutions that have been presented lately.

The paper is organized as follows. First, after a brief introduction in section [1], an overview of the HFT industry and its history is made in sections [2] and [3]. Second, an analysis of the principal strategies that are used in ultra-fast trading is conducted in section [4]. Third, the positive and negative consequences that arise from the activity are presented, together with the empirical works that sustain them (section [5]). Fourth, a bunch of different alternatives, which have been presented by their ideologues as backup solutions to contain the negative consequences of HFT, are critically discussed in section [6]. Section [7] specially focuses on analyzing the approach suggested by Budish et al. (2015), who proposed an alternative market design to finish with the problems that HFT provokes in markets, "Frequent Batch Auctions". To conclude, the main lessons of the paper are highlighted in section [8] together with the concluding remarks and future research pathways.

2. WHAT DO WE MEAN BY HIGH FREQUENCY TRADING

The American Securities Exchange Commission (SEC) defines High Frequency traders as professional traders acting in a proprietary capacity that engage in strategies that generate a large number of trades on a daily basis. In order to provide a more precise definition, the SEC also lists several characteristics commonly attributed to High Frequency Trading, including: "(1) the use of extraordinarily high-speed and sophisticated computer programs for generating, routing, and executing orders; (2) use of co-location services and individual data feeds offered by exchanges to minimize network and other types of latencies; (3) very short time-frames for establishing and liquidating positions; (4) the submission of numerous orders that are cancelled shortly after submission; and (5) ending the trading day in as close to a flat position as possible (that is, not carrying significant, unhedged positions over-night)."

On the other hand, the U.S Commodity Futures Trading Commission Technology Advisory Committee (CFTC) defines HFT as a form of automated trading that employs "algorithms for decision making, order initiation, generation, routing, or execution, and at the same time, it satisfies several criteria such as use of low-latency technology, high-speed connections to markets for order entry, and high message rates" (orders and cancellations).

So, although High Frequency Trading can be classified as a subgroup of algorithmic trading, it is important to emphasize that not all the Algo Traders[1] can be classified as HF traders, as long as HF traders need to satisfy very specific characteristics as very low latency and short holding periods. Put another way, the difference in the speed of execution of the trades is the main distinctive characteristic of High Frequency traders. In the following lines, a brief definition of the main characteristics of HFT is given.

- <u>Proprietary trading</u>

  Proprietary trading happens when a firm invest for its own direct market gain. In other words, a firm can be considered as a proprietary firm when it does not invest for the benefit of external investors (with their capital), gaining different commissions for this task. So, an HFT firm invests for its own benefit, and in consequence, its revenue-flows come from the gains that is able to obtain in the different financial markets where it participates. In this sense, HFT firms are opposite to other financial market participants that have emerged in last decades, as pension, mutual, or hedge funds. This happens due to the fact that HFT have a competitive advantage that enable them to earn excess returns: They are faster than any other market participant. The average high-frequency trading algorithm can process trade orders in less than

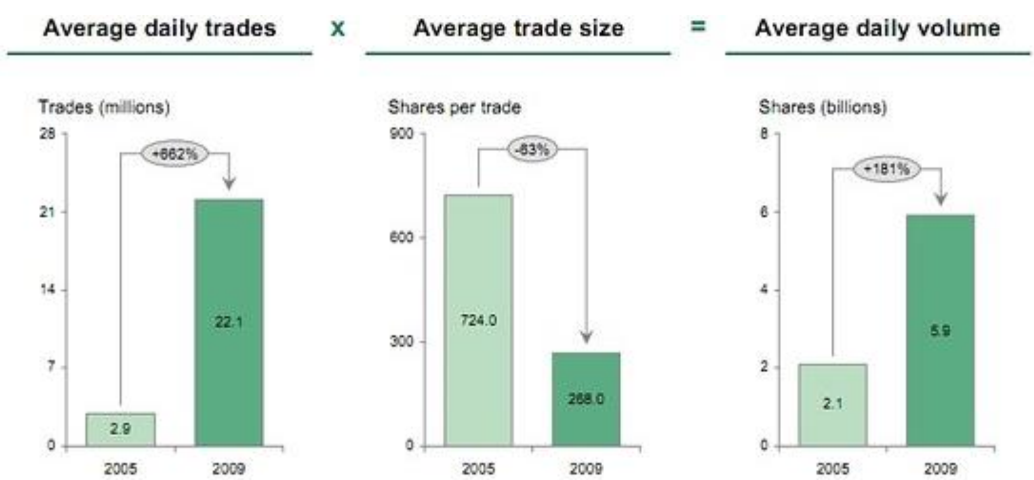
[1] Traders that that use computers to run complex mathematical formulas for trading. See a more extended definition on page 10.

400 microseconds, which is 1,000 times faster than the blink of an eye (Meyer et al., 2009).

- Low capital requirements for trading

  Although broadly speaking, the initial or fixed costs are far higher in HFT than in other trading strategies, the trading capital that is needed for each trade is relatively lower than in other forms of trading. That is, infrastructure and technology requirements (networking, hardware, co-location fees or the software implementation of the algorithm) demand heavy investment, but once paid, no much trading capital is needed to operate, especially if the trades involve very specific instruments (like two S&P 500[2] ETFs[3]). Low capital requirements are based on HFT firms' business models, that incite HFT companies to go flat home every day. In other words, even if the turnover of ultra-fast trading firms is usually much higher than the one of the rests, due to the fact that their market exposure last for very small quantities of time (positions are close very fast), low capital requirements are needed to operate. In that way, with a limited capital portfolio, a HF trader can execute thousands of operations per day without any financial leverage, increasing trading volume dramatically (The portfolio's capital times the number of trades conducted). See Figure 1 for a graphical illustration.

Figure 1. Trends in trading activity 2005 vs 2009. Source: Concept Release on Equity Market Structure (SEC, 2010).



Because of that, and due the importance that closing their positions has for these firms, HF traders tend to operate in highly liquid markets, being especially prevalent in large capitalizations stocks with high on-market competition and volatility (Jarnecic and Snape, 2010). In other words, empirical work reflects that HF traders

[2] The Standard & Poor's 500 is a market capitalization weighted index of the 500 largest U.S. publicly traded companies.

[3] An Exchange Traded Fund (ETF) is a mix of securities that replicates an underlying index. An ETF is a marketable security.

usually operate with high market value stocks, also known as "blue chips", due to higher liquidity (see Figures 2,3 and 4 for further proves).

Figure 2. Relationship between HFT activity (value traded) and market capitalisation. Source: Economic Report High-frequency trading activity in EU equity markets Number 1 (2014).

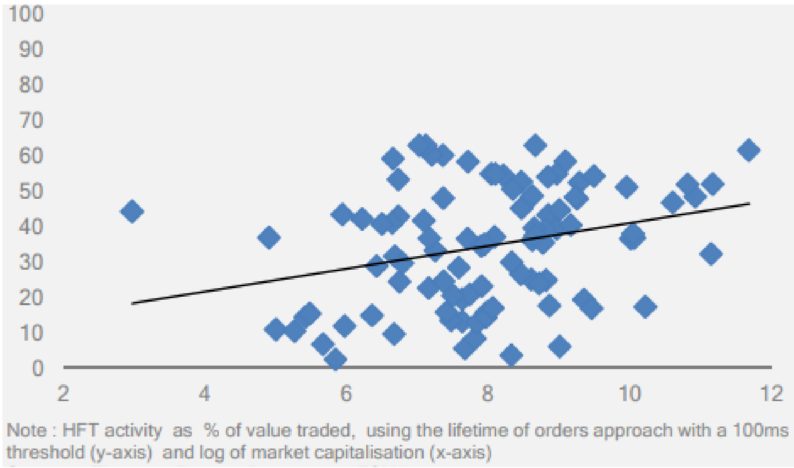
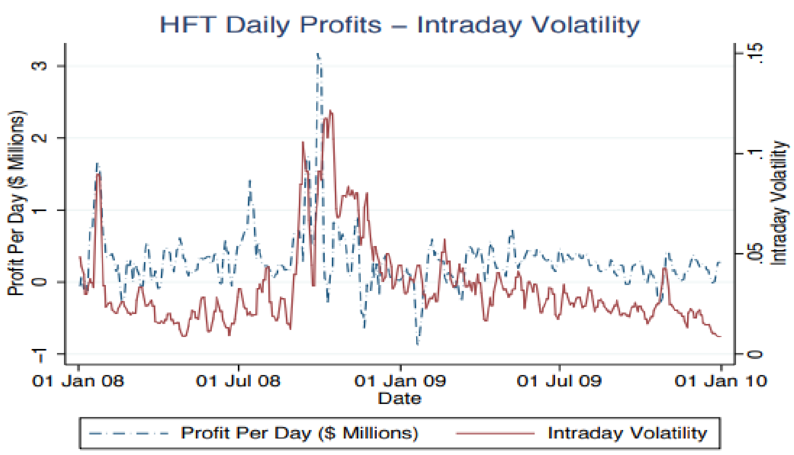


Figure 3. Relationship between HFT activity (number of orders) and market capitalisation. Source: Economic Report High-frequency trading activity in EU equity markets Number 1 (2014).

Figure 4. HFT profits – volatility. Source: High Frequency Trading and its impact on market quality (Broogard, 2010).



- Co-location

In a sector where the main differential factor is the speed advantage, every millisecond counts. In this context, co-location, or locating HFT computers in the same building (or on the same room) where the exchange's computers are, has become an incredibly lucrative business for stock exchanges, which have expanded their data centers in a very significant way to take advantage of this need. Thanks to the closeness to the exchange's computers, HF traders are able to obtain stock prices or send market orders (or cancel them) before any other market participants, and in consequence, they are willing to pay millions for this (Lewis, 2014). Put another way, by positioning their technology just some centimeters closer to the exchange's computers, some HFT capture the most up to date market pricing a few milliseconds faster than the others, and in that way, they increase their processing speed. In consequence, as they are able of seeing the dynamics of the market before everyone else, these firms obtain a competitive advantage and their gains soar.

A great example of this booming phenomenon was given by Lewis (2014), when he mentioned that NYSE Euronext spent 500 million dollars to build two centers, one in New Jersey and another one in Essex (England), each larger than two soccer pitches, with the intention to profit from co-location.

- Low latency

Latency can be defined as the time that a message needs to travel from one point to another or to execute a computer process. So, HFT low latency activities can be defined as strategies that respond to market events in a millisecond environment (Hasbrouck and Saar, 2013). However, this is an evolving definition, with the amount of time needed to be considered as low latency decreasing gradually.

In HFT, low latency is indispensable to react faster than the competitors and increase profitability. For example, if a HFT company wants to take advantage of an arbitrage opportunity, it needs to be faster than the other (*N-1*) trading firms, and for that, low latency is indispensable. Apart from co-location, HFT firms try to reduce latency periods using specialized chips that multiply the speed needed to make the calculus (FPGA), and the use of microwaves as signal transmitters (Tapia and Alubankudi, 2014). In order to permit communication, microwaves require a visual line between two points, and in consequence, signal repeaters are necessary to avoid natural obstacles. However, the latency period is considerably reduced in comparison to the optical fiber, which is the second most use mode of transmitting the orders. For example, an order needs 8 milliseconds to travel from Chicago to New Jersey using a microwave signal, 4.5 milliseconds less than employing optical fiber (Lewis, 2014).

- Unhedged positions

An unhedged or naked position is a position, either long or short, that represent a risk for the investor, as long as it has no take any step to hedge this risk. In other words, the investor has an exposure in the market that may create losses for itself if the things do not go as expected.

Usually, High Frequency Trading firms do not carry any portfolio or position over night, and in consequence, their risk of exposure is minimal. HF traders are able of doing this thanks to their speed advantage, which enables them to obtain positive benefits with almost zero exposure to market quotation.

- High message rates

High message rates include orders, quotes, or cancellations that are sent by a firm in the intraday market. High Frequency Trading Firms have a high message rate because they fulfil the following characteristics (MiFID II, 2014[4]).

1- Submission of at least 2 messages per second with respect to any single financial instrument traded on a trading venue.
2- Submission of at least 4 messages per second with respect to all financial instruments traded on a trading venue.

- Algo – Trading

HFT firms make use of mathematical formulas run by powerful computers in order to exploit or create market opportunities and take buy or sell decisions in financial markets. Thanks to preprogramed instructions, HFT firms are able of making thousands of trades per second (Tapia and Alubankudi, 2014).

In this sense, algorithmic trading can be defined as a type of trading in financial instruments where a computer algorithm automatically takes the following decisions with limited or no human intervention: When to initiate the order, the timing, the price, the quantity of the order, and how to manage the order after its submission (MiFID II, 2014). So, in algorithmic trading, the trading parameters, the confirmation of the order, and the post trade processing of transactions are handled by mathematical formulas run by computers.

[4] The "Markets in Financial Instruments Directive" (MiFID) is a European Union law that seek to provide harmonised regulation for investment services across the member states of the European Economic Area.

- Dark pools

  Dark pools are private forums for trading financial instruments, specially stocks and derivatives. The access to these private exchanges is limited for the investing public. The primary role that "Dark pools" play is to facilitate the execution of large trades to institutional investors, without making the details of the operation public. Thanks to these private exchanges, large investors avoid the adverse price impact that they could suffer with big orders. The lack of transparency is one of the key characteristics, and in consequence, they are also known as "black boxes of liquidity". Theoretically, dark pools reduce the risk of being harmed by HF traders' predatory strategies. However, ultra-fast traders have been allowed into some dark pools, making impossible for investors to know if they are trading against them, and usually receiving an inferior price when the trade is executed.

3. THE HISTORY OF HFT

The world's first stock exchange opened in Amsterdam in 1602, with the Dutch East India Company offering its shares. The Dutch East India Company is also considered the world's first multinational.

Since then, the modern financial world has experimented an incredible transformation, suffering a transition from using pigeons as order's carriers to microwaves and entering directly in the computer age. In this context, we have witnessed the incredible development of High Frequency Trading, the subject that we deal with in this paper. But what do we know about the history of HFT?

HFT, as we know it, emerged in the United States in the eighties, but its boom has taken place from the beginning of the 21st century onwards. On Black Monday, October 19, 1987, the DJIA (Dow Jones Industrial Average) fell by more than 500 points, or 22%. Back then, most of the trades were executed by a slow process that required human intervention (phone calls were needed to close trades). In this context, in a day of massive stress, traditional human market-makers felt overwhelmed by the thousands of orders that were receiving, and without prior notice, they stop answering the phone. Consequently, investors were left out in the cold, and as a response to this event, the NASDAQ revised the existing trading system (SOES) and gave preference to retail investors in the trading queue. More precisely, the instant execution of trades of 1,000 shares or less by average investors became mandatory, and in consequence, institutional investors with much larger orders had the obligation to wait. Due to this change, an opportunity was created for HFT firms, that began to appear in the market moving in and out of the stock at a more rapid speed than large investors, executing small quantity orders.

But using speed advantage to exploit market opportunities is nothing new in financial markets (Techstars, 2017). As an example, in the 17th century, the Rothchild family used pigeons to arbitrage prices of the same security in different exchanges by obtaining information ahead of its competitors. However, until 1980, when the first computerized system was created by Bloomberg, the technological speed race did not begin. The system was able of using real-time data to quote stock prices and relay information (financial calculations and analytics to Wall Street firms), and in consequence, from that moment on, traders obtained instant accessibility to market data. Since then, as the pure electronic form of trading started to gain weight, HF traders have been gradually developing.

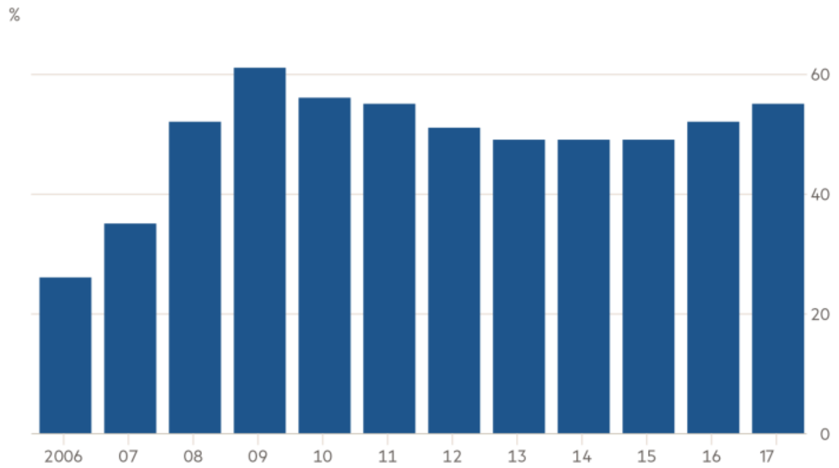
By 1998, after the SEC ruled in favor of creating electronic stock exchanges and authorized computerized HFT, High Frequency traders were already a consolidated reality in financial markets. They were able of executing trades much faster than humans, and although at that time, the latency period was still of several seconds, in a couple of years, HF traders were responsible of making at least the 10% of all trades. However, until recently, HFT has been an unknown subject outside financial experts.

From 2005 to 2009, high-frequency trading volume increased by 164% in the USA. This increase can be in part explained by the new National Market System (Reg NMS) that the SEC implemented in 2005. The initial intention of the new market system was to transform an old institution as the New York Exchange and make markets more efficient at the time of matching orders, but instead, a more complex system was created, that almost no one was able of controlling. The SEC tried to introduce competition between different exchanges in order to ensure the best price for investors, but with the creation of more than 50 public and private exchanges, the financial system became more difficult to regulate and coordinate. This is explained by the fact that the Reg NMS requires a broker to obtain the best price for their clients, the "National Best Bid and Offer" (NBBO hereafter), which is calculated by a computer system called SIP. Back in 2005, SIP had the capacity to aggregate bid and ask offers for all the U.S. stocks in all the exchanges, and thereupon calculate the NBBO, but it was not fast enough. In consequence, through co-location, HFT firms located their computers inside the exchanges, and they were able of obtaining the information (market prices), on average, 25 milliseconds faster than the SIP (Lewis, 2014). An eternity for HF traders.

In this context, HFT firms were able to find more market opportunities, and in consequence, of exponentially expanding their presence (in 2009, 60% of the trades were made by HFT firms, specially by Getco and Citadel). As it has been shown by Hoffman (2014) and Biais et al. (2015), the rents available to the fastest traders lead to over-investment in speed-enhancing technologies. Step by step, latency periods decreased to micro and even nanoseconds, and huge investments were made in order to be the fastest firm (see Figures 5 and 6 for a graphical illustration of the expansion process of HF traders).

Figure 5. Average number of HFT firms per stock (2008-2012). Source: High-Frequency Trading (Lattemann et al., 2012).

Figure 6. HFT as a share of US equities daily volume (2006-2017). Source: Financial Times.



However, after the "Flash Crash" of May 6, 2010[5], when the Dow Jones Industrial Index drop 998,5 points (about 9%) within minutes, the first warning signs emerged, becoming the regulators more aware and cautious. Furthermore, more miniature flash crashes that affected individual stocks have taken place in recent years (Malinova et al. 2012). As a consequence, HFT has been recently subject of regulatory talks. Italy was the first imposing a fee to

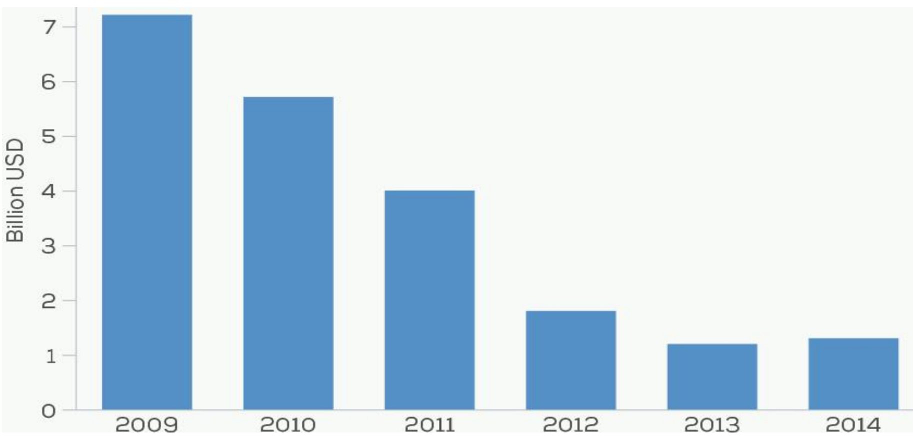
[5] The "Flash Crash" was provoked by a 4,1 billion dollar sell order instigated by Waddel and Reed (an American asset management and financial planning company).

discourage excessive market activity, and other countries as Canada or France have also implemented other fees.

Finally, it must be mentioned that European markets are more and more similar to American markets in terms of structure. However, it is evident that HFT firms have less importance in Europe ("only" 37% of the trades are made by HFT firms) than in the USA, which can be explained by the fact that the liquidation processes are more expensive in Europe.

In this regard, in Europe, MiFID has introduced a principles-based best execution regime, while in US markets, a rule-based approach has been taken. That is, instead of imposing the NBBO, MiFID requires investment firms to: "[…] take all reasonable steps to obtain, when executing orders, the best possible result for their clients taking into account price, costs, speed, likelihood of execution and settlement, size, nature, or any other consideration relevant to the execution of the order." (European Securities and Markets Authority, 2018). Hence, in order to be able to route and execute an order providing customers the best possible execution results in each of the trades, investment firms require access to a variety of real time market data (Gomber et al., 2011). In that way, as real-time market data comes at a cost, in the old continent, the size of the profitable market opportunities is smaller (Perez, 2011). However, even in US markets, driven by a huge increase in competition among HF traders, and a decrease in the volatility of the markets, profits has clearly decreased for HFT firms lately. Consider that volatility has been gradually decreasing in US markets since the financial crisis, and 2017, for example, was one of the historically least volatile markets since the 1960s. However, from 2018 on, US and worldwide markets' volatility has recovered its usual numbers, and together with the fears of a trade world between China and the US, the Brexit, the macroeconomic situations of Argentina and Venezuela, or the first signs of an economic slowdown, HFT strategies has started to recover lost ground again (Figure 7 in the shows the evolution of US HFT firms' profits from 2009 to 2014).

Figure 7. HFT firms' revenues Source: Automated Trader Magazine Issue 41 Q4 (2016).



4. HIGH FREQUENCY TRADING STRATEGIES

Gomber et al. (2011) mentioned that HF traders were responsible for almost the 60 percent of the order volume in stocks and derivatives trading. However, the term HFT cannot be describe as a single type of strategy, as HFT are a heterogenous group and they employ a variety of strategies (Hagströmer and Nordén, 2013). In other words, HFT is not a trading strategy per se, but it employs the ultimate technological tools to obtain better access to markets, market data, or order routing, and in that way, maximize the returns of already used and stablished trading strategies.

As it has been mentioned earlier, the success that HFT firms obtain using these strategies is partially driven by algorithms that are kept on secret by their owners. These algorithms are able of processing large volumes of information and executing a profitable trading strategy, thus leading to secrecy, and adding an extra difficulty to differentiate HFT strategies in a consensus way.

Nevertheless, in the paper, a differentiation is made between passive and active HFT strategies, following the classification made by Lattemann et al., (2012). Passive strategies provide trading opportunities (limit orders and quotes) to other market participants. On the other hand, active strategies seek to exploit imbalances in asset prices, discrepancies in valuation between different asset classes, or asset valuations that diverge from historical correlations. Put another way, while passive strategies follow market making strategies and provide valuable services to other broker and dealers, active strategies are constantly searching for market opportunities. Furthermore, the existence of a controversial active trading subgroup, composed by predatory trading strategies, made the classification even less straightforward. Brunnermeier and Pedersen (2005) defined predatory trading as trading that exploits or induces other market-participants' needs to reduce their position. That is, in predatory trading, a trader is able of making a profit by trading against another market-participant's position, driving an otherwise solvent but distressed trader into insolvency.

4.1 Market making strategies

Regarding passive strategies or market making strategies, HFT firms obtain cash-flows from two different sources. First, HFT firms trade on both sides of the market, placing both buying and selling limit orders that are slightly below (buying) and above (selling) the market price (Hoffman, 2013). In this regard, Kervel and Menkveld (2015) found evidence that suggests that HFT market-makers make money by earning the spread and lose money on positioning for horizons longer than five seconds (buying low and selling high). Put another way, they mentioned that HFT are incapable of forecasting price movements for periods longer than the very short time. These results were consistent with the findings that Hasbrouck and Sofianos (1993) made with human market makers.

Second, some HFT firms get paid by exchanges for providing liquidity to the market. The differences in rebates are substantial among venues, and Brogaard et al. (2014) demonstrated how significant rebates and small fees are for the profitability of HFT. The authors found that trading fees, clearing fees, and rebates can make a venue as Chi-X much more profitable than other venues as NYSE or Euronext for HFT firms. In this context, if a venue can attract HF traders by offering a favorable rebate-fee structure, it could be able of influencing where trading takes places. This is one of the reasons to understand the massive increase in trading venues that it has been experimented in the last 15 years.

So, while at first sight, it could be believed that market-making HF traders play the same role of more traditional market makers, as they fulfill the same duties and they also make a profit from the "bid-ask" spread and the liquidity rebates, that is not exactly the case. As Hagströmer and Nordén (2013) mentioned, passive HFT firms make use of market orders, demanding liquidity, with a higher frequency than traditional market makers. Furthermore, in the majority of the world's markets, HF traders do not have the obligation of providing continuous liquidity provision or having a minimum quote lifetime, and in consequence, a liquidity withdrawal is plausible in unfavorable times. In this regard, the new regulatory framework that has been stablished in Europe, the MiFID II, obliges stockbrokers to contractually define their obligations as liquidity providers, and it imposes legal sanctions to the market participants that do not meet their defined requirements when needed. With these regulatory changes, European countries try to avoid events like the Flash Crash of 2010. The new supervision regime is binding in the European Union and in the European Economic Area as per 3 January 2018.

4.2 Active trading strategies

On the other hand, active strategies are mainly divided in three different subgroups; arbitrage strategies, directional trading strategies, and market manipulation strategies.

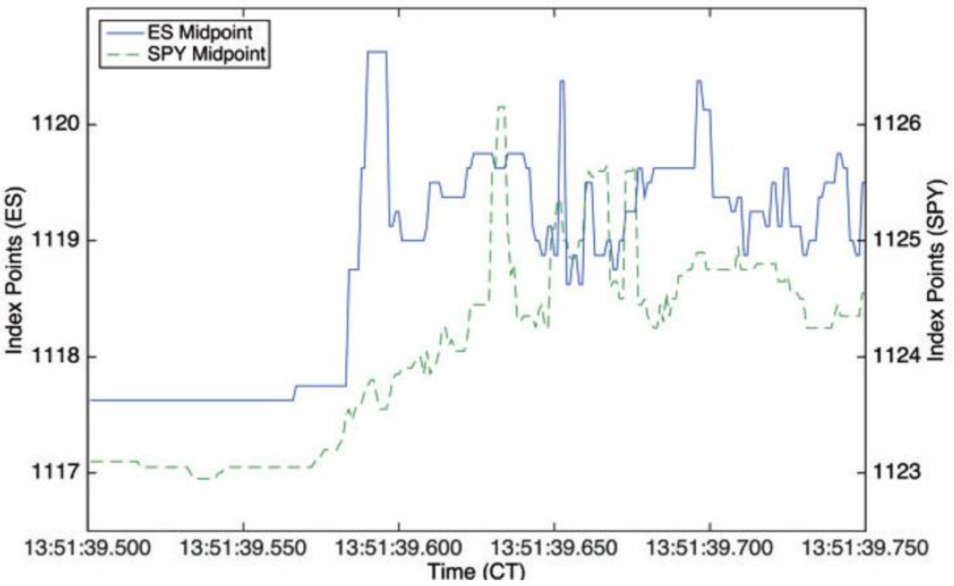
4.2.1 Arbitrage strategies

Arbitrage strategies can be defined as trading strategies that exploit specific market structure characteristics, as the excessive fragmentation of markets. So, following this definition, arbitrage strategies could be considered a subgroup of structural trading strategies, which profitably exploit certain structural vulnerabilities using their superior speed advantage. Hence, in a great extent, arbitrage strategies are rooted in the HF traders' superior ability to process and react to new information faster than the average trader. A great illustration of this phenomena was given by Arnuk and Saluzzi (2012): "By the time the ordinary investor sees a quote, it's like looking at a star that burned out 50,000 years ago".

Being more specific, Budish et. al (2015) unveiled that the prices of two securities that track the S&P 500 are perfectly correlated at the level of hour and even minute, but at the 10 and 1 millisecond level, the correlation breaks down to provide arbitrage opportunities (see Figure 8 to have a graphical representation). This result can be extrapolated to other liquid financial assets that have a very high price correlation or even to instruments that are meaningfully correlated (with correlation far from 1). For instance, even though Apple and Microsoft have an approximate correlation coefficient of 0.66, as both stocks have a sufficiently informative correlation, a large price jump in Apple's stocks may induce a race to react in the market of Microsoft. The same happens with stocks that are listed in at least two different markets, or with the price of an ETF (or a derivative), and an identical portfolio composed by these assets (or the underlying asset). Furthermore, arbitrages opportunities as the interest rate parity are also used in bond markets.

In this context, HFT firms compete to exploit these arbitrage opportunities first, utilizing their speed advantage to end with the price differences in less than a thousand of a second. Put another way, HF traders look for market participants that are offering quotes at stale prices. These strategies are known as statistical arbitrage strategies.

Figure 8. ES (E-mini S&P 500 future x10) and SPY (SPDR S&P 500 ETF) time series at HFT time horizons (250 milliseconds). Source: Frequent Batch Auctions (Budish et al., 2015).



4.2.2 Directional trading strategies

Directional trading strategies try to predict price movements based on public information. That is, by being the first detecting and analyzing public information, firms that make use of directional trading strategies take a long position if they believe that the tendency of the market or a specific security is going to be bullish, while they take a short position if they predict that the prices will fall. In this extent, some HF traders try to predict price changes over short horizons, based on sources of public information, as imbalances in the limit order book, public news, or social media announcement (Cao et al., 2009).

Using the latest technologies as machine learning, HFT firms use automated systems in order to detect keywords that could be useful at the time of predicting movements in assets' prices. As an example, on April 23, 2013, a false tweet was published in Twitter, claiming that a terrorist attack took place in the White House and that Obama was injured. Immediately, the Dow Jones plummeted more than 150 points, suggesting that several investing strategies are connected to Twitter's activity. But why is so important for a HFT firm to be the first to react when new market information is aired? The answer is simple; by being the fastest, HFT firms trade on new information before others even have the chance to react. In this respect, Scholtus et al. (2014) affirmed that among HFT firms, speed is crucially important for strategies that are based on U.S. macroeconomic news releases. Put another way, by using order-level data on the highly liquid S&P 500 ETF (NASDAQ, from January 6, 2009 to December 12, 2011) they found evidence indicating that a delay of 300 milliseconds or more significantly reduces the returns for news-based trading strategies. Furthermore, the authors asserted that in the minute following the arrival of macroeconomic news, algorithmic trading increases trading volume and depth at best quotes, but at the same time, it has a negative effect on the overall depth of the market and it also increases market volatility.

Imagine the following situation:

The European Central Bank makes an unexpected statement announcing an increase in the interest rates. In consequence, BBVA's stocks prices would rise in the stock market, and with a reasonable probability, the jump in the price (x'-x) would be greater than the bid-ask spread. If that is the case, the liquidity providers would try to cancel their old quotes at x and, at the same time, replace them with new quotes at x'. However, simultaneously, HFT firms, by executing trades at the old quotes before they are cancelled, would try to snipe the stale quotes. In this context, as the market follows a continuous microstructure, the message that is processed is the one that arrives first (the limit order book processes messages in a serial way), and in consequence, by being the fastest, HFT firms obtain extra rents (Budish et al., 2015).

In this sense, Baron et al. (2014) found evidence suggesting that HFT traders make approximately 45% of their profits by adversely-selecting slightly slower liquidity-providing HF traders, and Chaboud et al. (2014), also suggested that the fastest traders take advantage of their speed leverage to pick off slower market participants, increasing the adverse selection costs (refers to the presence of certain traders that have advance knowledge of asset fundamentals).

In short, by making use of directional strategies or by looking for price discrepancies between securities on different exchanges or asset classes, HFT firms are able to obtain a revenue, that although small in comparison to their trading volume, amount for a total of more than 5 billion dollars only in the U.S markets (Cookson, 2013). Being more precise, Brogaard et al. (2014) mentioned that approximately, HFT firms win 0.43$ per 10,000$ traded in NASDAQ. Hence, financial literature seems to indicate that informational advantage (measure in milliseconds) is enough to overcome the bid-ask spread and the trading fees, generating positive revenues.

4.2.3 Market manipulation strategies

Market manipulation strategies, also known as predatory trading strategies, are market strategies that manipulate the price of the underlying security intentionally and with the sole intention of obtaining a profit. Among them, momentum ignition and front-running are the most usual ones.

As to momentum ignition, it refers to strategies that try to attract other algorithms and traders to also trade the stock. In that way, HF traders create a price movement that does not respond to the intrinsic valuation of the stock, and they make a profit by taking a position before the price movement takes place and closing their position before the price reverts to normal.

But how is this this artificial price change create? One way, for example, is to quote enough selling orders to activate the automatic selling of a stock (investors usually make use of automatic selling orders to avoid excessive loss), that is usually programmed to generate sell orders when the price goes bellow a minimum threshold. Another related method, known as spoofing, consists on placing orders that are not intended to execute, creating a misleading impression of the stock's liquidity and, in consequence, manipulating the price of the security. For example, by placing sell orders that are below the market price, as market participants feel the selling pressure in the stock, the ask price may fall. So, if HF traders have the required capacity to quote enough selling orders, they may be able to buy the stock at a lower price, and using their speed advantage, immediately cancel their remaining orders. In other words, HFT firms could use their superior technology to manipulate prices at speeds that are undetectable by slower traders.

Obviously, these strategies could be very harmful for other market participants, as they create market stress periods where HFT or algorithmic traders trigger sell orders that create impulsive drops in financial markets (see the Flash Crash as an example). A well-known case of spoofing is the Singh Sarao investigation.

Navinder Singh Sarao operated as a trader in London until 2016, when by court order, he was extradited to the United States. Sarao, who was accused by US authorities, plead guilty of building a fraudulent investment system between 2009 and 2014, that manipulating worldwide market prices, obtained more than 40 million dollars. The use of spoofing in American markets is illegal since 2010, when the Dodd-Frank law[6] was approved. More specifically, by using "bait" orders, Sarao was able of moving investors in the intended direction, and almost instantly, in fractions of a second, of cancelling the old orders and executing the new ones (in the opposite direction).

On the other hand, utilizing another well-known predatory strategy, named front-running or tailgating, HFT firms are able to detect large pending transactions that create movements in the securities' prices and take advantage of this non-public knowledge. This strategy is especially harmful for institutional investors like hedge or pension funds, and partially explains the development of dark pools, or private trading exchanges not accessible for the investing public. To better understand tailgating strategies, imagine that a broker receives a market order from a pension fund to buy 900,000 shares of a specific stock. However, at that moment, no market has 900,000 shares available-for-sell and the broker divides the buying order in 3 different orders of 300,000 shares each, that consistently, are sent to three different stock markets. Furthermore, even if the markets have sufficient liquidity, Kyle (1985), Vayanos (1999) and Sannikov and Skrzypacz (2016) suggested that an investor with multi-unit demand prefers to split his order into several smaller orders. In this situation, and taking advantage of the disaggregation, HFT firms have the capacity to manipulate the market in at least two different ways: Changing the market price in every exchange or changing the price in the exchanges that need a higher latency to execute the order. But how do they do it?

[6] The Dodd Frank law initiated various reforms affecting almost every aspect of the financial system with the intention of preventing a repeat of the 2008 crisis

HFT firms constantly quote selling orders that work as baits. So, if they see that enough selling orders (of a specific security) have been executed as to assume that a big buying order has been placed, they immediately cancel their remaining selling orders, and thanks to their speed advantage, they buy a big portion of the remaining shares, selling back to the broker at a higher price (the reverse happens when a big selling order is placed). The price difference is minimal for each security, but due to the intervention of HF traders, the broker is not able of executing the order at the initial market price. Hence, nanosecond traders impose a tax on buyers. Lewis (2014) mentioned that this invisible tax was smaller than a 0.1% tax, but only in the U.S, accounts for approximately 160 million dollars per day (average daily volume in the U.S stock market: 225 billion dollars). The estimation was made by Katsuyama and his team (founders of an alternative trading system named IEX, see pages 33 and 34), that described the phenomenon in the following way: "It happens on such a granular level that even if you tried to line it up and figure it out, you would not be able to do it. People are getting screwed because they cannot imagine a microsecond" (Lewis, 2014).

But, even if the broker is able of executing the buying order in the exchange with the lowest latency at the initial market price, the same may not happen in the other two. This occurs due to the fact that HFT firms are able of detecting that a big buying order has been executed in the first exchange, and thanks to their speed advantage, manage to arrive earlier than the broker to the other two exchanges. Once there, they follow the same procedure, buying the shares at the market price and selling them back at a higher price. So, without any doubt, low latency periods are critical to profitably execute these strategies, and at the same time, to compete against other HFT firms. However, by front-running, HF traders are somehow cheating, and in consequence, committing fraud.

The reader may question whether these strategies, also known as liquidity detection strategies (LD hereafter) are realistic strategies in European markets. In this regard, it may be convenient to remember that, as it has been explained in section [3] of the paper, the imposition of the NBBO was one of the triggers of the success of LD strategies in US markets. The NMS regulation requires orders to be routed to the trading venue that hold the NBBO, splitting orders and routing them to different trading venues, and in consequence, offering HFT a chance to obtain a profit. The European marketplace does not enforce a "European Best Bid and Offer", but nonetheless, it meets the three requirements needed to successfully operate with LD strategies: The presence of (co-located) HF traders, cross-market trading, and different latencies from brokers to the different trading venues (The Netherlands Authority for the Financial Markets, 2016).

5. POSITIVE AND NEGATIVE CONSEQUENCES OF HFT ACTIVITY IN FINANCIAL MARKETS

As it has been mentioned in sections [2] and [4] of this work, it is undebatable that the inclusion of technology, algorithms, and HFT in last instance, has transformed the financial markets. Some argue that these changes have been beneficial for the efficiency and transparency of the markets, but not everyone agrees on that. As Bogle (2012) mentioned, Wall Street may have forgotten which its principal mission is, and it has plunged into a wave on speculation and betting. On the other hand, some tech experts (Musk or Coleman, for example) consider Bogle, Buffet, Gates, or Simpson old dinosaurs that are reluctant to accept their extinction. In this context, HFT begs the question whether it adds value to financial markets or not. In other words, investors may ask whether a social planner would be willing to spend the resources required by HFT or not.

Among the principal benefits, HFT supporters cite that liquidity is higher with HFT firms in the market, as they ease the effect of market fragmentation. Furthermore, they sustain that bid-ask spreads have experienced a decreasing pattern due to HFT activity, and finally, they also argue that HFT enhances price discovery and reduces volatility (see Malinova et al., 2012; Hendershott and Riordan, 2011). However, opponents of HFT dispute these claims and express their concerns about the market inefficiencies that could emerge as by-effects of nanosecond trading, as market anomalies created by high order cancelation rates or predatory strategies. Furthermore, they warn that episodes of market stress as the "Flash Crash" are more common now, due to the development of HFT (see Kirilenko et al., 2016; Menkveld, 2013).

Early work on the effect that algorithmic trading has in financial markets suggests that market liquidity increases thanks to a greater automation (see Hendershoot et al., 2011), but on the other hand, there is growing evidence highlighting that increases in the efficiency of automation (higher speed) could led to lower liquidity. As an example, Menkveld and Zoican (2017) found that the increased speed of trade execution on the NASDAQ OMX had a detrimental effect on liquidity (the effective spread increased by 32%). Furthermore, most of the improvements in liquidity are associated with the rise of IT in the period between 1990-2004, well before the boom in HFT, both in the evolution of bid ask spreads, and in the cost of executing large trades (see Angel et al., 2015; Frazzini et al., 2012). See Figures 9 and 10 to visualize the variations in transaction costs and in bid-ask spreads.

Figure 9. Average Transaction cost estimate for 1M shares in a 30$ stock (in basis points). Source: Equity trading in the 21st century: (Angel et al., 2015)

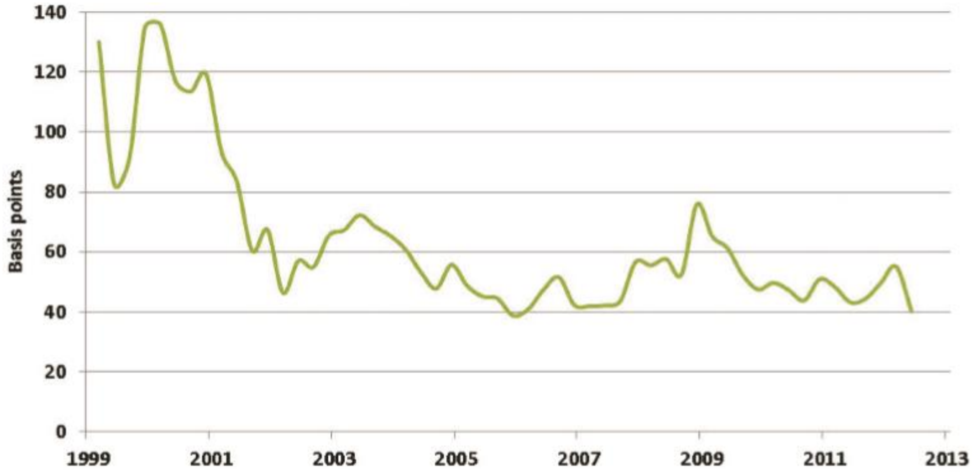
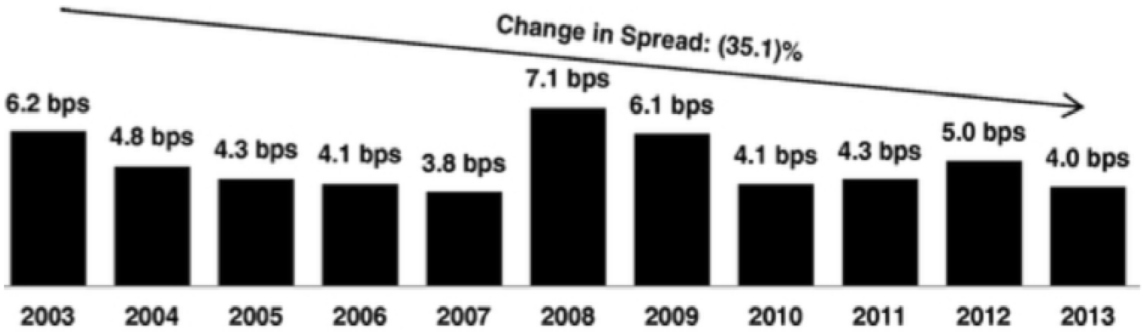


Figure 10. Average Bid/Ask spreads in the S&P 500 (2003-2013). Source: United States Securities and Exchange Commission (Virtu, 2014).



In short, the HFT phenomena conflates two distinct occurrences: the introduction of information technologies in financial markets over the last 20 years, and the speed race. At the beginning, the introduction of automation reduced the cost of attention, and inventories became easier to manage, reducing trading costs as a consequence, but with the expansion of HFT, these findings have been put into question, revealing the need to conduct further research.

In the following lines, an overview of the most significant findings is made, in order to determine if HFT is beneficial or prejudicial for financial markets. For that, an extensive overview of the literature has been made, analyzing the positive and negative aspects that this activity brings to financial markets. Overall, the academic literature is divided, but some common ground could be found between scholars; the challenge is to favor the positive aspects of HFT, trying to minimize the drawback that could arise.

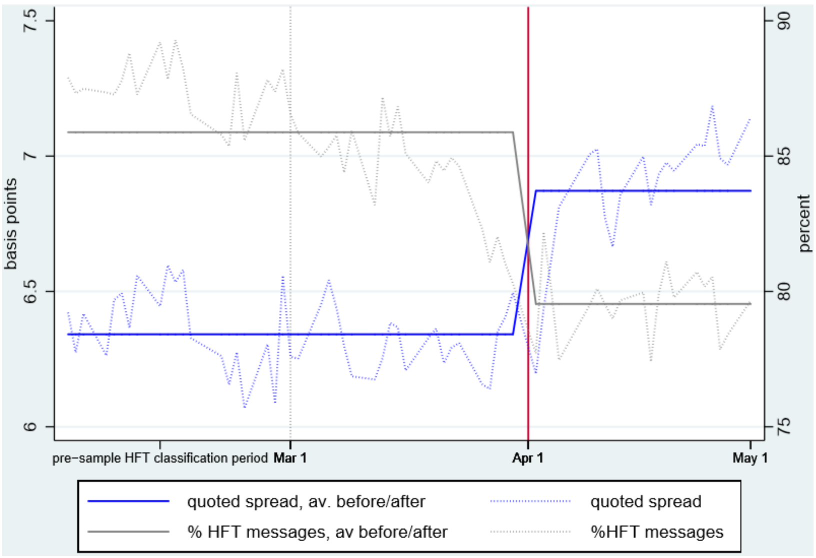
5.1 What do we know about HFT?

5.1.1 Asset pricing: Liquidity and price discovery

Malinova et al. (2012) cited that for every 1% decrease in message intensive traders' activity, bid-ask spreads increase by 3 basis points. The study analyzed the effects of a regulatory change that took place in Canada on April 2012, where a per-message fee of 0.00022$ that affected algorithmic quoting activities was imposed. Message intensive traders were defined as traders that had a high message-to-trade ratio, being market messages, messages composed by order submissions, cancellations or modifications that a broker or a dealer generated. So, after conducting the analysis, the authors found evidence suggesting that due to the regulatory change, the number of orders and order cancellations decreased by 30%, increasing the bid-ask spread by 9%. Figure 11 graphs the effect of the fee in the market spread.

Figure 11. Time-weighted quoted spread vs %HFT. Source: Do Retail Traders Suffer from High Frequency Traders? (Malinova et al., 2012).

- The figure plots the percent to messages that are generated by traders classified as HF traders.



These findings are consistent with the theoretical predictions made by Copeland and Galai (1983), Foucault (1999), or Bernales and Daoud (2013), who suggested that if liquidity providers cannot modify their orders so frequently (due to the per-message fee in this specific situation), the chances of quoting more stale orders increase, and therefore, an increase in the bid-ask spread can be observed. The HFT firm Getco, on a comment letter to the Investment Industry Regulatory Organization of Canada (IIROC) explains the logic behind this relationship in the following way: "If additional risk is placed during the time that quotations are made, additional risk compensation will be needed, widening the bid ask spread" (IIROC, 2011). Furthermore, Malinova et al. (2012) obtained evidence that suggests that while institutional trader's intraday results from market orders increased, retail orders were better off with a higher level of activity from message-intensive HFT firms.

On the other hand, Budish et al. (2015) contradicted these statements and affirmed that among others, the effect that HF traders have in market liquidity depends on the transparency regime governing the market. With transparency, increasing the number of HFT is good for liquidity, but with opaqueness, the market may be unstable and liquidity hump-shaped in the proportion of HFT. This suggests that the liquidity impact of HF traders should be examined considering the effect of frictions in the access to market information. Furthermore, the authors mentioned that being HFT dependent in liquidity terms is risky, as crashes could suddenly emerge. Why? Because a shock in market parameters could trigger a change in HFT firms' strategies and provoke a liquidity withdrawal.

However, liquidity is not the only relevant function for asset pricing, as price discovery also plays an essential role for incorporating information in prices (see O'Hara, 2015). Consider that due to the pressure arising from the liquidity demand by long term investors, the temporary price impact of large trades causes noise in prices. In this extent, if HF traders trade in the direction of permanent price changes and in the opposite direction of transitory pricing errors, HF traders benefit price efficiency and decrease long term investors' trading costs.

In this regard, Brogaard et al. (2014) found evidence that suggests that HF traders benefit price efficiency, concluding that the efficiency-enhancing activities of HFT play a greater role than the efficiency harming ones (e.g. predatory strategies that attempt to manipulate prices). In other words, the authors could not find any significant evidence suggesting that, on average, nanosecond trading is linked with price inefficiencies, so they concluded that overall, manipulative strategies do not play a dominant role in HFT firm's strategies. Hence, the study indicated that when prices deviate from their fundamental values, due to pricing pressure, HF traders demand liquidity to push prices back to their efficient levels. Brogaard et al. (2014) also highlighted that HF traders do not reduce their liquidity supply in the days of market stress, being their role in price discovery similar to the one that they have in lower volatile days. In this regard, the German research institute KIT analyzed NASDAQ data, and concluded that HFT worked as a buffer against decreasing stock prices in the years 2008 and 2009 (Riordan and Zhang, 2011).

So, according to these scholars, HFT decreases pricing errors, and at the same time, it increases the information impounded into prices, aspects that help a better resource allocation in the economy. These finding are consistent with other author's predictions that speed increases both, market liquidity and price efficiency (see Brogaard et al., 2014; Baron et al., 2014; Weller, 2013). Furthermore, Hendershott and Riordan (2011) also obtained evidence suggesting that HFT facilitates price discovery, as it increases the informational efficiency of prices. Nevertheless, they rose a new question, as they affirmed that they were not able of determining whether or not the information that HFT incorporates into prices, only a few milliseconds faster, enhances social net benefits or not. For that, the authors considered that ultra-fast trading imposes cost to investors, in the form of intermediation, technological expenditures, or regulatory difficulties.

Menkveld (2013) went a step further, and he rebated the positive effects of HFT in price discovery. In his words, traditional market-makers maintain an inventory control, or selling when accumulating inventory and buying when inventory becomes too low. However, HFT are subject to capital constraints, and in consequence, they do not hold large or negative positions. So, when HFT have such constraints and they play such an important role in trading, their limitations affect prices.

5.1.2 Market efficiency

Gomber et al. (2011) cited the difficulty to perform profound research about the topic, mainly due to the lack of available empirical data, but he argued that as part of the value creation chain, HFT contributes to increase efficiency and to reduce explicit and implicit transaction costs. Other recognized scholars as Riordan, Riess or Krogmann (Lattemann et. al, 2012) also highlighted the contribution of HFT to the efficiency of securities trading. In their opinion, nanosecond trading leads to a faster processing of information, to an increase in liquidity, and thus, it adds value to the economy.

Nonetheless, the study published by (Borkovec et al., 2014) displayed evidence that contradicts these affirmations, as they concluded that the trading costs that investors bear in U.S. stocks markets, had if anything, risen in the period between 2007-2014. The analysis was made using trading data from institutional investors, and costs were calculated in the following way: $Total\ Cost\ =\ Implementation\ Shortfall\ +\ Commissions$, being the "Implementation Shortfall" the difference between the trade and the benchmark price.

Regarding competition among HFT firms, back in 2014, the Bank of Canada arrived to the conclusion that competition decreases bid ask spreads and increases market efficiency. When a new HFT firm begins trading, as there are particular trades that every HFT firms want to engage, it changes the trading environment and it disturbs incumbent HFT firms' behavior. Consequently, incumbents' HF traders lose part of their market share, and competition in providing liquidity tights the spreads. Furthermore, their price predictability decreases, and markets become more efficient, decreasing at the same time, revenues for HFT firms (Brogaard et al., 2014). Being more precise, revenues to market-makers have dropped from 1.46% in 1980, to just 0.11% in 2006 (Budish et al. 2015), and a great part of this variation can be attributed to the development of HFT firms. So, as new entrants compete for order flows, and more firms impound information into prices, the price path becomes more random. As part of its framework, Fama (1970) highlighted the notion of an efficient market. In his words, an efficient market is a market in which securities' prices reflect all available information.

Again, these findings are consistent with much of the literature. Among others, Huang (2002), Mayhew (2002), and Battalio et al. (1997) proved that when exchanges compete, market participants benefit, and Klock and McCormick (1999), Weston (2000), and Van Ness et al. (2005) demonstrated that market making competition increases liquidity. The same logic applies for the competition between HFT firms.

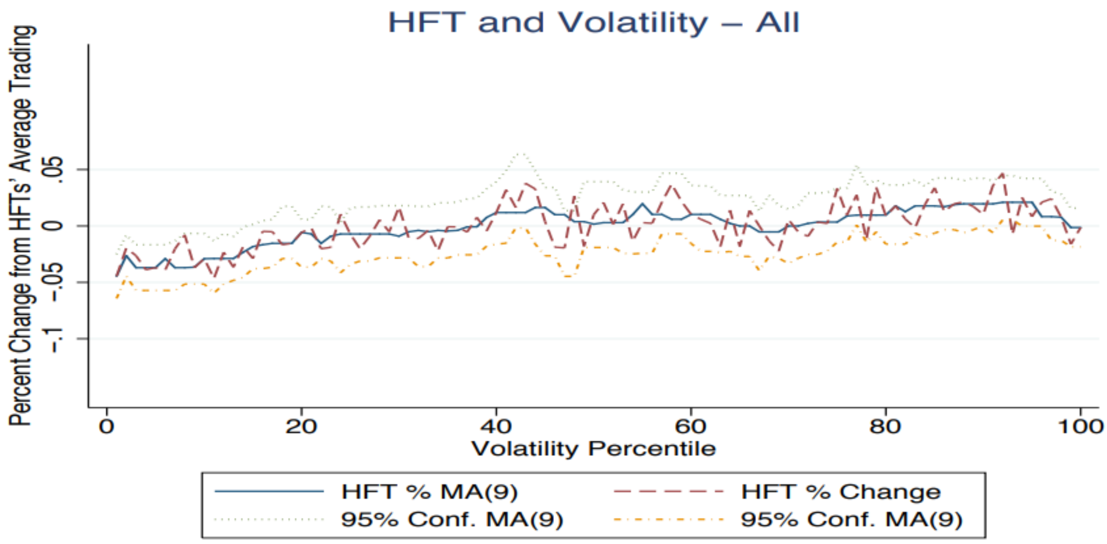
On the other hand, Breuer (2013) recognized that, although thanks to HFT, information can be processed faster and more efficiently than ever before, from the perspective of financial economics, considerable doubts could arise regarding the benefits that the activity provides. As the author mentioned, more and faster information does not necessarily lead to a correct determination of the intrinsic value of financial instruments. This happens because ultra-fast trading uses short term information, which primarily consists on short term volume and time series data, and thus, does not help to the evaluation of the intrinsic values.

5.1.3 Volatility

Kirilenko et al. (2016) also criticized HFT activities, affirming that although HFT did not trigger the flash crash, their response to the selling pressure was key to increase market volatility. As they see it, HFT has substantially replaced human market-makers in international financial markets, which could be a double-edged sword. So, agreeing with Budish et al. (2015), they sustained that much of the liquidity provision in today's markets depends on HF traders' activity, and at times of stress, HFT firms become demanders of liquidity, not providers, which causes significant volatility as liquidity deteriorates. These findings were consistent with Gai et al. (2013) that found evidence suggesting that HFT activity is associated with higher levels of quote cancellations, more short-term price volatility, and little depth at the best price. Cartea and Penalva (2012) also found that HFT increases the volatility of prices. Figure 12 graphically represents the positive relationship between volatility and HFT activity that was found by Broogard (2010).

However, Hasbrouck and Saar (2013) conducted a study that gave favorable evidences in favor of HFT activity. They made an analysis in a normal month and in a month with economic uncertainty (stocks prices declining more than 10%), and the results were clear. In a market dominated by HFT, more HFT activity is preferable to less. What is more, as HFT increase their activity, the spreads were lower, the market depth increased, and short-term volatility decreased in both market situations. Remember that as it was mentioned by Glosten and Milgrom (1985), narrower spreads indicate less adverse selection cost for market makers. Hagströmer and Nordén (2013) also found that an increase in the market making activity of HFT decrease short-term volatility.

Figure 12. HFT and its impact in volatility. Source: High Frequency Trading and its impact on market quality (Broogard, 2010).



5.2 Summary

The question of whether financial markets were better or worse off before the inclusion of HFT remains unanswered. In this sense, if HF traders use their speed advantage to crowd-out liquidity provision, or if steeping in front of limit orders is not expensive, investors may not benefit from the reduction of the spread, for example. Put another way, the overall welfare may decrease.

However, the effect is not the same for all investors. While the greatest benefits are enjoyed by retail investors (lower spreads), institutional investors may find that trading large positions is more difficult, feeling the need to invest in more advanced trading technology. Nevertheless, one may question the accuracy of the argument that HFT helps retail investors and harms institutional investors, as most citizens do not have brokerage accounts, and therefore, their savings are part of big pension or mutual funds. In this extent, if HFT damages large institutional investors, teachers, savers or pensioners could be considered as HFT taxpayers.

All in all, we do not know if HFT adds value to financial markets, and the debate is still unsettled. Scholars have found evidence both in favor and against HFT, and contradicting results are prevalent in the literature. In general, it seems like HFT increases market volatility in certain market situations, increasing the systematic risk, but it also seems to increase price efficiency and to reduce the bid-ask spread. However, regarding liquidity, the effects are still questionable, as HFT can be both liquidity providers and demanders, depending on the market situation and the firm's strategies.

In this situation, the question that rises is if the benefits that HFT firms provoke could be strengthened, minimizing the possible negative consequences. In that extent, many market participants have advocated for more strict regulation, but again, the utility of a more severe regulation remains an open question. In the following lines, possible alternatives to this problem are analyzed, concluding the paper with a deep analysis on the possible pros and cons that the frequent batch auction alternative, as a new market design, could have at the time of reducing HFT negative aspects and enhancing the positive ones. This alternative was proposed by the professors Budish (University of Chicago), Shim (University of Chicago) and Cramton (University of Maryland).

6. POSSIBLE SOLUTIONS TO HFT PROBLEMS

Among the alternatives that have been proposed to combat the negative consequences of HFT, more severe regulation emerges as the easiest and the most viable one. However, market regulation could lead to dramatic changes in market behavior, and inappropriate or excessive regulation, might even have more negative than positive effects for market participants. In consequence, financial regulation must address market transparency, accessibility, and competition, maintaining investor's trust as a central asset of any market, and without excessively altering market behavior (Schwartz and Francioni, 2004). In other words, the main objective of market regulation is investor's protection, and for that, providing an efficient and fair market is essential, avoiding at the same time, fraud, manipulation, or insider trading.

Tobin taxes, or financial transactions taxes, were originally suggested as the cure against massive speculation and volatility (see Tobin, 1978; Summers, 1986). In this extent, Stiglitz (2013) proposed Tobin taxes as a viable solution against HFT, and in 2013, Italy was the first country to impose the tax on its financial markets. Nevertheless, the results were not optimal; investors suffered an increase in trading costs, and the prevailing feeling was that the increment in trading costs was more pronounced than the benefits from less sniping. In order to eliminate the incentive to invest in speed technology, Tobin taxes need to be greater than the maximum benefit that can be obtained with HFT strategies. So, if Tobin taxes are too high, financial regulators bear the risk of making investors worst off. Similar alternatives, as imposing a tax on speed (Biais et al., 2015), suffer the same practical problems, as long as to reduce arms race expenditures by 90%, regulators are obliged to impose a tax of 900% on speed expenditures (Budish et al., 2015), that without any doubt, wil cause unexpected changes in market behavior.

On the other hand, another possible solution could be to impose minimum quote lifetimes or maximum message to trade ratios to decrease the excessive message-cancelation rates that HF traders have installed into financial markets, and at the same time, ensure continuous liquidity provision (Figures 13 and 14 show the effect that Algorithmic Trading and High Frequency Trading have had in quote changes and lifetime). The first alternative prohibits immediate order cancelations, as orders must rest in the book for the time that regulators decide, whereas the second approach prohibits ratios (messages divided by complete trades) that are above a threshold imposed by the regulators (Brown and Yang, 2016).

However, both measures impose additional risk to HFT firms, and in consequence, both possibilities may create an adverse effect on market stability and quality. Volatility may increase, and liquidity may decrease because HF traders will be less able to manage intermediation risk, and simultaneously, as HF traders will not be able to change their limits orders so frequently, an increase in transaction costs (increase in the bid-ask spread as a compensation) could be also experimented (Lattemann et. al, 2012). In other words, nanosecond traders will not be able of reacting quickly and adequately to market exogenous information, presenting a free option to other market participants. To better understand the problem, and to appreciate how these proposals seem to misunderstand the cause and effect of the problem, see the following example:

Imagine that there is a jump in the price of asset y within the minimum lifetime period, or when the market maker has already reached the maximum message to trade ratio. At that moment, the HFT firms that work as liquidity providers in the market will not be able of cancelling their stale quotes, and in consequence, sniping opportunities will increase.

Figure 13. Distribution of lifetime of orders (cancelled or modified orders, in second).
Source: Economic Report High-frequency trading activity in EU equity markets Number 1 (2014).

- Lifetime of orders: Time elapsed before the order is modified or cancelled. Firms identified as HFT appear to send orders with shorter lifetime (40% less than 0.2 seconds), compared to Investment banks (40% less than 5 seconds) and other firms.

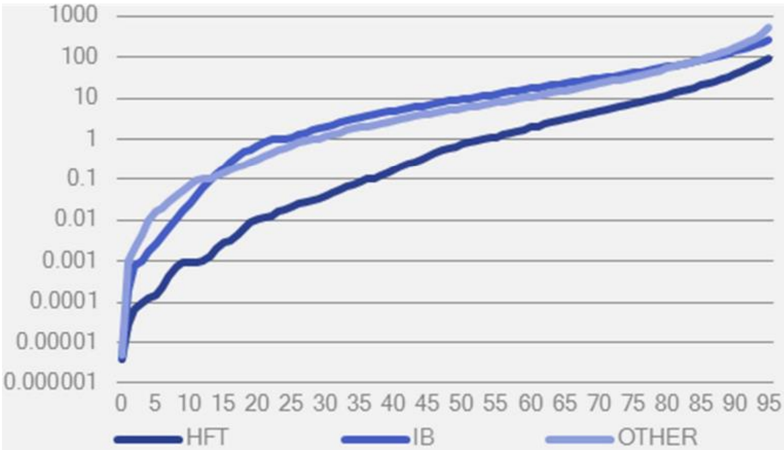


Figure 14. Number of US quote changes per million shares traded (2003-2016).
Source: Financial Times.

In this context, Harris (2013) proposed random message delays as an alternative solution. In his words, "regulatory authorities could require that all exchanges delay the processing of every posting, cancelling, and instructions they receive by a random period of between 0 and 10 milliseconds". However, random message delays have clear drawbacks, to the point that they only add additional randomness to the markets. If a HFT firm try to snip a stale quote, and at the same time, the liquidity provider aims to cancel its stale quote, the random message delays incorporate more uncertainty to the speed race, but they do not solve the core problem. What is more, HFT firms will submit more and more redundant orders in order to avoid the possibility of having their orders cancelled, increasing the message to trade ratio dramatically.

Something similar will happen if HF traders have the obligation to provide liquidity in market stress situations, or in other words, if an obligation of fulfilling the role of more traditional market-makers is imposed to them, even in plumping markets. Hence, enforcing a legal obligation to constantly provide quotes may not be the best idea to avoid a sudden liquidity withdrawal. Furthermore, scholars wonder whether any rule can bind HFT firms to buy when selling pressure is high. These firms are characterized by rapid closing positions to avoid risk, and consequently, they rather pay a fine than going against their business model. Remember that theoretically, HF traders do not have any preferential access to market information that is restricted to other market participants. Put another way, their advantage is solely based on the more advance technology, and in consequence, without such privileges, there is no clear basis to impose traditional market makers' obligations to HFT firms.

Finally, Gai et al. (2013) proposed to deregulate the minimum tick size in financial markets. The authors found evidence that suggests that HFT firms make more trades in stocks where the spreads cannot get smaller, and in consequence, they proposed reductions in the minimum tick size to enhance competition on price instead of competition on speed (among liquidity-providing HF traders).

Summing up, it looks like no magic solution exists once again, and in consequence, it could be assured that more severe regulation also comes with its drawback. In this context, the decision to introduce tighter regulation or not should be make considering the social preferences. Does the social utility increase with regulation even though negative effects exist? The answer is still unknown. However, one thing is clear; if regulators decide to impose stricter regulatory rules, these must be applied to every market and market participant to the very same extent, otherwise, they are worthless.

Nowadays, the most conservative regulation is the one of the European Union, with US regulation getting stricter every time. However, the regulatory approach to HFT has nonetheless been different across every region, driven by the absence of a definition and universally recognized measures to identify it. As an example, Asian countries, with Japan and Singapore as the main leaders, have supported the growth of HFT in their markets, while European authorities have been rather cautious of adopting measures that could incentive HFT activity.

In this respect, the entry into force of MIFID (unified normative), re-shaped EU markets in 2007, but until the introduction of MIFIDII in January 2018, European authorities have not been able to correctly rule on HFT and algorithmic trading. MiFID II introduced closer regulation and monitoring on algorithmic trading, requiring new and detailed requirements on HF traders and on trading venues. As an example, a cap was imposed on the amount of trading that is permitted on dark pools. Being more precise, if the trading of any stock in dark pools exceeds 8% of the activity across all European exchanges, then, market participants will be banned from trading that stock in dark pools for the next six months (MiFID II/MiFIR series, 2014).

On the other hand, in the United States, the FINRA[7] and the SEC (as the ultimate regulator), approved a rule change on April 7, 2016, that has affected developers of algorithmic trading strategies. The new rules were designed to prevent fraudulent and manipulative acts, to protect average investors, and to promote, in the greater extent possible, equitable principles of trade. For example, the new rules require HFT proprietary traders to become members of FINRA, thereby increasing regulation of market participants that have presence in off-exchange trading activities, and increasing costs for these firms (SEC, 2016)

Finally, regarding Asian countries, thanks to a greater acceptance among regulators, and driven by the widely accepted feeling that if they do not adapt to current trends they could be left out in the cold, HFT presence across Asian markets has grown in recent years. As an example, countries like Indonesia and Malaysia are experimenting an important increase in the number of algorithmic traders, and while their markets are not yet deep and fast enough to support HFT requirements, nanosecond trading will gradually increase in these countries in the forthcoming years. However, a more advance country like Japan, that for years has been the country to follow in this sense, has started to be less proactive in terms of trying to meet HFT needs. HFT accounted for about 70% of the orders on the Tokyo Stock Exchange in 2016, a data that raised concerns about the stability of the market and the protection of retail investors. So, in the new regulations that came into force in 2018, the Japan's market regulator, The Financial Services Agency (FSA), made sure that HFT participants have from that date, the legal obligation to be registered (Twaronite, 2017).

On the other hand, far from regulation, other alternatives also exist. For example, Katsuyama, Park and Ryan founded an alternative trading system named IEX (Investors Exchange) in 2012. IEX aims to finish with predatory strategies by imposing 350 microseconds of latency for all orders (the delay is created through coiling 61km of cable), which is the difference between the time the signal is sent, and the time that it is received

[7] The Financial Industry Regulatory Authority is a nongovernmental organization that decides and enforces the rules governing brokers and broker/dealer firms in the USA.

(logged on the exchange). So, thanks to the delay, the exchange is able of calculating the NBBO before a HFT firm can act upon the price change. The exchange is fully owned by investors, it prohibits co-location, and it does not pay any rebate for providing liquidity. The idea was also adopted by the "Toronto Stock Exchange (TSX)", and the logic behind the delay is to finish with the speed and informational advantage that same HFT firms have. As an example, imagine that a firm is able of obtaining information that determinates the future price movement of a stock before the rest of the investment public (a few milliseconds earlier). With the speed bump, this advantage will be vanished before the order is executed, and in consequence, the market has time to react, reducing adverse selection problems, and accordingly, decreasing spreads and trading costs.

However, the idea of the speed bump also receives critics, as it provides artificial or stale quotes by forcing delays in execution, and consequently, distorting the dynamics of the market. Among others, Baldauf and Mollner (2015) mentioned that since IEX applied a delay to immediately executable orders but not to posting or cancelling messages, liquidity providers continue having a head start when there is a price movement. Put another way, the speed bump is not able of transforming the competition in speed into competition on prices, and therefore, it is not able of eliminating the incentives to invest in speed. Furthermore, a delay in execution makes less convenient to trade, as it increases uncertainty around execution, and it may reduce the proportion of traders that trade for non-informational reasons. In this regard, there exists evidence suggesting that a decrease in the proportion of noise-traders would then lead to an increase in bid-ask spreads and to a thinner market depth (Glosten and Milgrom, 1985).

In this regard, the concept of the speed bump has been used by the betting exchange Betfair for more than a decade. Being more concrete, Betfair delays any order submitted inplay by 5-9 seconds before it is executed in the exchange, but as in IEX, cancellations are not subject to the speed bump. Brown and Yang (2016) analyzed the effectiveness of the method, reaching some interesting conclusions: Although the speed bump protected slower traders, no clear effect on market quality was found, and after same time, fast traders begun to develop successful strategies to circumvent the delay. Fast traders sent limit orders both in favor and against an event (providing liquidity on both sides of the bet), waited for the outcome, and canceled one of the orders before it was logged into the exchange and picked off by a market order. In other words, with this new investment technique, the speed bump does not solve the adverse selection problem, as it only transfers the risk from limit order traders (liquidity providers) to market orders traders. An effect that was already suggested by

Thaler (1988) in his winner's curse theory. Conceptually, as information will be in hands of everyone before the order is executed, receiving information a few seconds faster is worthless in monetary terms, but as orders can be cancelled if new information arrives during the speed bump, the logic does not apply. So, after conducting the analysis, the authors found that although the ability of traders to predict fundamentals clearly decreased with the implementation of the speed bump, protecting slower limit order traders, the evidence were mixed when it comes to market quality (spreads, depth, and the frequency and size of the trades). Within season, larger delays increased the bid-ask spread, decreased trading, and made order execution more difficult.

In short, Brown and Yang (2016) concluded that as long as it is a simple way to avoid the speed bump, it is reasonable to think that the execution delay is not the best alternative in order to combat high speed traders' investment techniques. Furthermore, hedging and liquidity provision are more frequent in financial markets than in betting exchanges, and in consequence, the speed bump may hurt liquidity in a greater extent in financial markets. As an example, if a liquidity provider cannot hedge its position immediately, it may increase bid-ask spreads in order to compensate for the risk assumed.

Finally, Budish et al. (2015) proposed an alternative market design to finish with the problems that HFT provokes in markets. Arguing that the continuous limit order book has core design problems, the authors affirmed that the continuous market is flawed, and that the negative consequences of HFT can be avoided with a proper market model. In the new market design, the trading day would be divided into extremely frequent but discrete time intervals. So, due to the interest that the new proposal has raised, and because we have considered that it has potential to be treated as an interesting alternative, the frequent batch auctions market design is extensively analyzed in the following section of the paper.

7. THE FREQUENT BATCH AUCTIONS ALTERNATIVE

Financial markets constantly updated its order book throughout the day. Put another way, under what is known as the continuous limit order book market design, investors have the constant possibility to buy or sell traded securities at any instant during the trading day. Moreover, as it has been mentioned in previous lines of this paper, the continuous market design creates arbitrages opportunities, as correlation between two very correlated securities breaks down at high frequency time scales. These arbitrages opportunities are only available to the fastest firms, and although competition to be the fastest is fierce, the size and frequency of arbitrage opportunities have not changed, leading some scholars to ask if the current market design is the appropriate one.

Among these scholars, Budish, Cramton, and Shim stand out with their contributions to the paper "The High Frequency Trading Arms Race", where they suggested that financial exchanges should use frequent batch auctions every tenth of a second. In that manner, with the new market design, all the orders that are received during the batch interval are executed simultaneously, matching the orders at a uniform price and avoiding serial processing (Budish et al., 2015).

7.1 Why the continuous market design is flawed

The authors analyzed the existence and duration of arbitrage opportunities in two almost perfectly correlated assets, the E-mini S&P 500 futures contract (ES) and the SPDR S&P 500 exchange traded fund (SPY). Both, the ES and SPY are based on the underlying Standard & Poor's 500 stock index. Empirical data from the Chicago Mercantile Exchange and New York Stock Exchange was used, recording all the activity that occurred in the order books of both exchanges, message by message, with a milliseconds resolution, leading to very interesting results.

The duration of the arbitrage opportunities between the two assets decreased considerably from 2005 to 2011, from a median of 97 milliseconds in 2005 to a median of 7 milliseconds in 2011. However, the profitability of the arbitrage opportunities remained invariant, with a profit of 0.08 index points per unit traded (See Figures 15 and 16). This reflects the substantial investment in speed that HFT firms have made in recent years, but at the same time, it confirms that the profitability of arbitrage opportunities has not suffered significant changes. With these results, the authors approximated the gains that HFT firms can make from the ES-SPY arbitrage, which amounts for more than 75 million dollars per year.

Figure 15. Median duration of the ES and SPY arbitrage opportunities (2005-2011). Source: Frequent Batch Auctions (Budish et al., 2015).

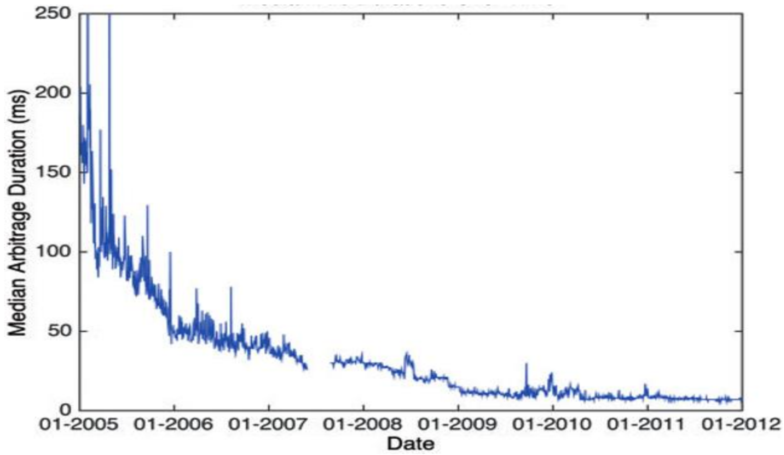
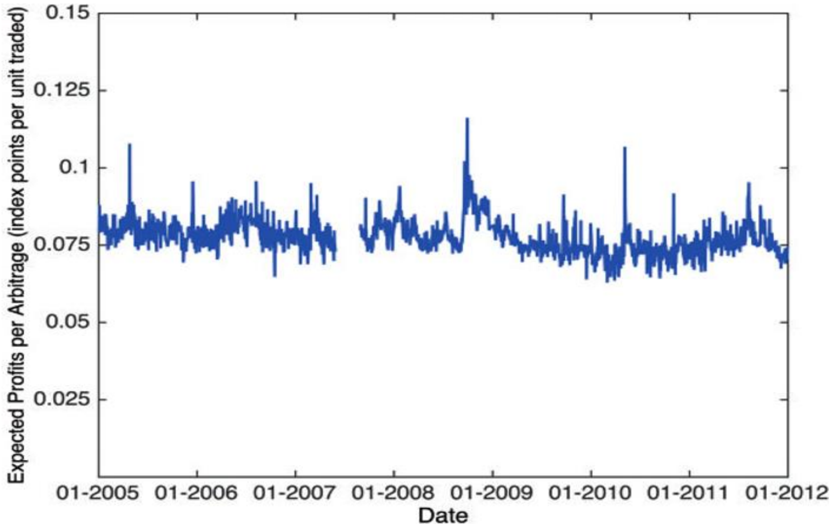


Figure 16. Median profitability of the ES and SPY arbitrage opportunities, per unit traded (2005-2011). Source: Frequent Batch Auctions (Budish et al., 2015).



In this situation, is not unreasonable to think that the firms that provide liquidity to these two securities consider the cost of being snipped, and in consequence, increase the bid ask spread. Furthermore, following the same intuition, snipers invest in speed technology to win the speed race for the stale quote, and for the opposite reason, to avoid snipping, liquidity providers also invest in speed, reaching a prisoner's dilemma. Is in each firm's best interest to make the investment, but collectively, each one of them will be better off without further expenditures in speed technology.

In this regard, as an answer to the prisoner's dilemma, the authors argued that the discrete time market design reduces the tiny speed advantages that are necessary to successfully execute these market opportunities, transforming competition on speed into competition on prices. The logic works as following: If all markets participants have the information and the ability to execute the orders at the same time, they must compete on prices instead on speed.

7.2 The frequent batch auctions market design

Frequent batch auctions are very similar to the continuous limit order book, with two main distinctive features. First, as it has been mentioned, the time is treated as discrete, not continuous. Second, orders are executed using a uniform-price auction, instead of a serial execution (Budish et al., 2015).

So, in the frequent batch auctions market design, the trading day is divided into equal-length discrete time intervals, where investors have the possibility to submit market or limit, buy or sell orders. As in the continuous market design, the market participants also have the possibility to modify or cancel their orders whenever they want, and the orders remain active until they are executed or cancelled. In that way, at the end of each interval, all the outstanding orders are batched, aggregating all the supply and demand functions (bid-ask). If an intersect exists between the demand and supply functions, the market clears at the price where the supply equals demand (uniform price and maximum possible quantity q). If not, the orders remain outstanding for the next batch. The orders are displayed publicly after the batch (not during the time interval), together with the price and quantity.

In that way, orders with bids higher than p, or asks lower than p, are completely executed at p, whereas for orders exactly equal to the clearing price, it may be necessary to ration one side of the book to enable market clearing. Also, a time priority is established, prioritizing orders that have been in the order book for more than one batch.

With these changes, the speed leverage is only profitable if the jump in y (take y as an asset that replicates the fundamental value of asset x) occurs at a specific time in the interval, the time interval in which the fast traders see the jump in y that the slow traders do not see. This time interval depends positively on the fast trader speed advantage, and negatively, on the batch interval. In order to prove these affirmations and the roots problem with the continuous market design, the authors constructed their own theoretical model.

7.3 The model

There is a security x that is traded on a continuous limit order book, and at the same time, it exist a publicly observable signal y that has the same value of security x (movements in y have an arrival rate of λ_{jump}). In other words, the fundamental value of x is perfectly correlated with y, as it happens with SPY and ES. Moreover, x can always be liquidated at its fundamental value, and the authors distinguished between two types of players.

On the one hand, the end users or the liquidity takers (pension funds, hedge funds, ordinary investors…) categorized as investors, with a positive time preference and an arrival rate of λ_{invest}. On the other hand, "Trading Firms", which are classified as the players that do not have any necessity to buy or sell, and in consequence, they act when they seize a market opportunity (possibility to buy low and sell high), being liquidity makers or takers (HFT traders, Algorithmic traders or other type of market makers). The analysis was made considering two different scenarios:

7.3.1 Exogenous entry

In both scenarios, there is a trading firm that takes the role of the liquidity provider, while the other $N - 1$ firms are considered stale quote snipers. The liquidity provider has the obligation to send bid and ask limit orders of x at $y - spread/2$, and it obtains a profit equal to the spread. In that way, if there is a jump in y, or a trade in x is completed at price y, the liquidity provider modifies its quote on x, and it cancels the stale quote. On the other hand, if the change in y $(y' - y)$ is sufficiently high to compensate $s/2$, the stale quote sniper tries to complete the transaction on x at price y, before the liquidity provider has time to cancel its order, and with the intention of getting an arbitrage gain later.

Hereby, in the exogenous trading model, where all trading firms observe changes in the signal y at exactly the same time (trading firms cannot invest in speed technology), serial processing becomes a problem, as stale quote snipers are more abundant (N −1), and in consequence, the liquidity provider will lose the race the $(N - 1)/ N$ of the times.

So, with these assumptions, the authors assumed that the profits of the liquidity provider can be described as following:

$$\lambda_{invest}.\left(\frac{s}{2}\right) - \lambda_{jump}.\Pr(J{>}\left(\frac{s}{2}\right)).E\left(J - \left(\frac{s}{2}\right)\middle| J > \left(\frac{s}{2}\right)\right).\frac{N-1}{N}$$

The liquidity provider makes a gain when investors arrive to the market $(\lambda_{invest} * s/2)$, and loses when the jump (that occurs at λ_{jump}) is sufficiently high to compensate for the spread and it loses the race. When that occurs, the expected loss is the conditional expectation of the jump size (J) minus the bid-ask spread.

On the other hand, each stale quote sniper obtains a gain when it wins the race, which occurs with a probability of $1/N$ and when the jump is high enough for compensating the spread. Furthermore, its conditional expected profit is equal to the conditional expected loss of the liquidity provider.

$$\lambda_{jump}.\Pr(J{>}\left(\frac{s}{2}\right)).E\left(J - \left(\frac{s}{2}\right)\middle| J > \left(\frac{s}{2}\right)\right).\frac{1}{N}$$

So, in equilibrium[8]:

$$\lambda_{invest}.\left(\frac{s}{2}\right) = \lambda_{jump}.\Pr(J{>}\left(\frac{s}{2}\right)).E\left(J - \left(\frac{s}{2}\right)\middle| J > \left(\frac{s}{2}\right)\right)$$

[8] The left-hand side represents the revenue from the positive bid ask spread, whereas the right-hand side represents the rents from sniping.

The results were obtained making the strong assumption that the players only transact one unit of x. However, intuitively, it can be though that the costs of providing liquidity increase with the quantity offered (wider bid-ask spreads for each additional unit offered), as snipers will try to transact for the whole depth of the market, while benefits of increasing the quantity offered will not be such as substantial, as investors rarely transact for the whole quantity offered (see Budish et al., 2015 for further proves).

So, the models exhibit evidence that suggests that the serial execution of orders has increased the price of liquidity beyond zero, which is the theoretical cost of trading that a model with no inventory costs, search costs, or information asymmetries should have due to the Bertrand competition between the trading firms (see Copeland and Galai, 1983; Glosten and Milgrom, 1985; Kyle, 1985; Stoll, 1978; Duffie et al., 2005 for further explanations). In other words, the continuous limit order is a source of costly liquidity provision in itself, as liquidity providers recover the expense of being sniped widening the bid-ask spread and making the markets thinner. Why? Because in the continuous market design, symmetrically observed public information seems to be processed by markets as if it was asymmetrically observed private information (Budish et al., 2015). In this extent, Copeland and Galai (1983) and Glosten and Milgrom (1985) already predicted that liquidity providers are exploited by players that take advantage of the stale quotes when asymmetric information exists.

7.3.2 Endogenous entry

In the endogenous entry scenario, the authors incorporated latency to the equation and allowed trading firms to invest in speed technology, initiating what they called "the arms race" or the race to be the fastest. In that way, if no investment in speed technology takes place, trading firms observe the change in y with a latency of ∂_{slow}, while on the other hand, with a cost of c_{speed}, trading firms have the possibility to reduce the latency period to $\partial_{fast} < \partial_{slow}$. Hence, in the new model, if a liquidity provider or a trading firm decides not to invest in speed, it loses the race for not being sniped the 100% of the times, while the rest of the participants that undertake the investment will be successful the $\frac{1}{N_{invest}}$ of the times. So, following a zero-profit equilibrium condition, the equilibrium equation for the liquidity provider is the following:

$$\lambda_{invest}.\left(\frac{s}{2}\right) - \lambda_{jump}.\Pr(J > \left(\frac{s}{2}\right)).E\left(J - \left(\frac{s}{2}\right) \middle| J > \left(\frac{s}{2}\right)\right).\frac{N-1}{N} = c_{speed}$$

The only difference with the previous model is that with the new assumptions, as there is no place in equilibrium for slow players, profits are equal to the cost of investing in speed. The exact same case happens in the zero-profit equilibrium equation for trading firms:

$$\lambda_{jump}.\Pr(J>\left(\frac{s}{2}\right)).E\left(J-\left(\frac{s}{2}\right)\middle|J>\left(\frac{s}{2}\right)\right).\frac{1}{N}=c_{speed}$$

So, together, both zero-profit conditions determine the equilibrium quantity of entry (N^*) and the equilibrium bid-ask spread (s^*). The equilibrium bid-ask spread is identical to the one obtained in the exogenous model[9], and considering that there is only one liquidity provider and $N-1$ trading firms[10], the quantity of entry, or N^* can be calculated:

$$\lambda_{invest}.\left(\frac{s^*}{2}\right)=N^*.c_{speed}$$

The equation above shows that expenditure by trading firms on speed technology is borne by investors. Also, as equilibrium is now obtained with a zero-profit condition, the rents created by the continuous limit order book are reduced to zero due to the arms race (Budish et al., 2015). Theoretically, in the authors' words, all trading firms will be better off with a truce, but every one of them has reasons to invest in speed technology in order to maximize its profits (Nash equilibrium). In short, it can be said that the authors were able of obtaining evidence that suggests that continuous limit order books increase bid-ask spreads and reduce market depth, and that the speed race leads to a wasteful prisoner's dilemma.

7.4 How frequent batch auctions solve the problem

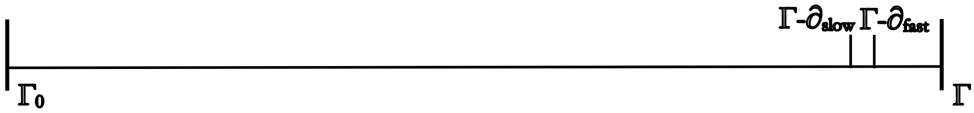
After bringing to light the complications that may arise with a continuous market design and excessive investments in speed technology, the authors explained why the frequent batch model can be an appropriate alternative to solve the problems presented above.

First, in the frequent batch auctions design, any jump in y that occurs during $(0,\Gamma-\partial_{slow})$, being Γ the length of the discrete time interval, can be observed by both the slow and the fast market participants in time to react before the next batch auction. Furthermore, if the movement in the price of y occurs at $(\Gamma-\partial_{fast},\Gamma)$, neither the fast nor the slow trader has time to react. So, in the frequent batch auctions market design, the only time interval where the speed advantage is relevant is reduced to $(\Gamma-\partial_{slow},\Gamma-\partial_{fast})$. This is the unique time where the fast trader has an economically meaningful advantage against a slow trader, in

[9] A subtraction is made between the zero profit conditions

[10] 1 * Zero-profit equilibrium of the liquidity provider + (N-1) * Trading firms' zero profit equilibrium = 0

comparison to the 100% of the times in the continuous market design. Put another way, if the batch interval is 100 milliseconds and the speed advantage is 100 microseconds, the proportion of the trading day where the speed leverage is relevant is $((0.0001/0.1) * 100)$ 0.1% or $(\partial / \Gamma) * 100$.



Second, the use of frequent batch auctions eliminates snipping. This is easily understood with an example: Imagine that as it has been described in the first model, no speed difference exists among trading firms. In this situation, every time there is a large enough jump in y, the fast trader is vulnerable to be snipped, and in accordance, it cancels the stale quote on x and it sends a new limit order. However, if the batch interval is sufficiently small, trading firms do not have the adequate time to react, and even if the new orders manage to enter in the batch, the liquidity provider has time to cancel its quote (no speed advantage exists, and orders are not processed serially).

On the other hand, in order to see what happens when speed differences are taken into account, the worst-case scenario is analyzed, where a slow trading firm and $N-1$ fast quote snippers make up the market, and the jump in y takes place at the critical time interval ($\Gamma - \partial_{slow}, \Gamma - \partial_{fast}$). In other words, the liquidity provider is the only market participant that is not able to see the jump in y in time to avoid being snipped. Nevertheless, even in this scenario, as orders are not serially executed, and all the $N-1$ trading firms are able of submitting the orders in time, market participants do not have any other choice than to compete on prices. The trade goes to the trader that offers the best price on x.

This happens because trading firms have incentives to deviate and offer a bid (ask) slightly higher (lower) than the one that they would have offered in the continuous market design. So, due to Bertrand competition, the auction price of x will be equal to the price of y at time $\Gamma - \partial_{fast}$. From a critical point of view, it has to be mentioned that actually, latency differences exist among "Trading Firms" in markets, and in consequence, it may happen that only some of them make it on time to get the order in the next batch. This could reduce competition among them, and as a result, we may see a price on x different to $y_{\Gamma - \partial fast}$. However, considering the actual latency periods of HF and Algo traders, a batch interval of approximately 0.1 seconds should be high enough to create fierce competition among them.

To sum up, the main benefit that frequent batch auctions provide is that they stop both snipping and the arms race in speed, a socially wasteful prisoner's dilemma. Thanks to that, the danger of market destabilization is contained, and markets are in principle less vulnerable to events like the Flash Crash. Why? Because even though frequent batch auctions cannot fight against the reality that trading firms can occasionally make programming errors (see Strasburg and Bunge, 2012), they do reduce their incentive to give up robustness and precision for speed.

Finally, the discrete time market design also simplifies the work for regulators, as it makes the markets more predictable. With the actual market design, the sequence of timestamps in the exchanges does not reflect the real sequence of events with a 100% accuracy, and thereby, examining market events could be an extremely hard job, as it was clear in the Flash Crash investigation (see SEC, 2010). For all these reasons, the frequent batch auction market design could be the most appropriate alternative to minimize the risk that HFT activity creates on average investors. However, even this new approach is not exempt from problems, and in the section [7.5], a review of them is made.

7.5 Frequent Batch Auctions: Implementation problems

The frequent batch auctions market design could be too idealistic. Academic theory that is just not practical. At first sight, it may look like the perfect solution against the arms race and the structural problems that the continuous markets may present, but things are much more complicated in reality.

First, in order to be effective, the frequent batch auction market design has to be legally imposed by worldwide regulators. Regulation is key to ensure coordination among markets, and to avoid massive investors migrations to other markets. Put another way, the discrete time market design has to be implemented all the way down in every exchange/trading venue. If not, new market opportunities will be created between continuous and discrete markets, or between different assets. Continuous markets will provide leading signals, and market participants will try to exploit the new market opportunities in discrete time markets. Because of that, worldwide authorities, or at least the ones that have a low latency between their markets, will have the obligation to impose very homogeneous regulations. Otherwise, the remedy will be worse than the disease.

In other words, if the frequent batch auctions market design is not mandatory, a new prisoner's dilemma will exist, as each exchange would have an incentive to deviate. If some exchanges do not impose frequent and discrete batch auctions, they will be able to save the technological expense that the new market design will provoke, and at the same time, they will not be affected in terms of market share, as low latency investors will love to have a continuous market when the rest of the markets are batching (arbitrage opportunities). Furthermore, if HFT firms decide to invest in speed technology, it is reasonable to think that they profit from the investment, and in consequence, they will not be willing to finish the "speed race".

So, as Hemsley, CEO of the BATS Chi-X Europe stock exchange mentioned at the European Capital Markets Institute's annual conference in Brussels: "In order to avoid the asynchronicity of price discovery between exchanges, all exchanges would have to become frequent batch auctions and have the auctions perfectly synchronized" (Hemsley, 2014). This is an extremely complicated task in today's world, because even if governments are able to settle things with lobbyist, they need to get every country on board to avoid massive migrations of investors to other markets. Furthermore, actual legislations in the more advance countries of the world implicitly assume continuous trading, so massive changes will be necessary.

On the other hand, the work of Budish et al. (2015) does not consider that if investors must wait to trade, they could suffer greater delay costs, increasing their opportunity costs. If these costs are insignificant or not (investors may be impatient) may be subject to discussion, but as it has been mentioned in section [6], with higher execution delays within season, Brown and Yang (2016) reported an increase in bid-ask spreads and a decrease in betting volume at the betting exchange Betfair. Furthermore, another important aspect for investor's protection, transparency, may also be affected, as the information concerning the respective batches will only be displayed after the execution. Finally, under such a market design, market participants (especially liquidity providers) would be obliged to postpone hedging until the next batch auction takes place. In this extent, even if the delay lasts for milliseconds, inventory risk will be higher for market markers.

In short, although Frequent Batch Auctions introduce an innovative and interesting alternative to combat the problems that HFT may provoke in our markets, they also present evident practical problems. So, we cannot affirm if social welfare increases or decreases with this new market design.

8. CONCLUSION

During the lines of this paper, an extensive overview of High Frequency Trading, and its consequences in financial markets has been done. By describing the different strategies that ultra-fast trading firms employ, analyzing the positive and negative consequences of the activity, and examining the possible alternatives that may exist to mitigate the problems that HFT provokes, hopefully, the reader has been able to obtain a more accurate insight of the industry.

First, it is convenient to consider that HFT is not a trading strategy as such, and in consequence, the future regulatory talks should be oriented to regulate the different trading strategies that HF traders use, and not HFT as itself. In this regard, is not unreasonable to think that HFT, in principle, provide positive externalities to other market participants. However, the ignorance that the main public, other trading professionals, and even regulators have had about the activity for years, has clearly contributed to the development of questionable HFT practices, that when uncovered, have tarnished the name of HFT, even for great financial minds as Buffet or Bogle.

In theory, HFT can be considered as a natural evolution of computer trading, and HF traders as sophisticated market participants that obtain a profit as a compensation for their risk exposure. However, far from arbitrage and market making strategies, the continuous use of predatory strategies, together with the propagation of dark pools and co-location, have enabled ultra-fast trading companies to obtain a profit without any risk exposure, something that is inadmissible, and that questions the whole legitimacy of the activity.

In this respect, one thing is clear; experience has shown that HFT needs an up-to-date and a constant supervision to ensure the appropriate functioning of the industry. Financial authorities have to make certain that HFT firms act in good faith, and that their strategies do not steel value from other market participants. For that, although an adequate regulatory framework is essential, the most important aspect is to ensure that regulators are able to understand the complex financial system that they need to control, and that they are constantly in evolution. This will prevent trading firms from undertaking legally questionable activities, and the prevailing feeling that market regulators are always one step behind the more advance market participants will be vanished.

It is true that with the appropriate supervision HFT still presents volatility or liquidity withdrawal problems, but these are without any doubt, more acceptable complications. Why? Because these difficulties do not have their roots in any illegal or ethically questionable trading strategy. So, as it has been highlighted in previous lines of the work, worldwide market participants need to decide if their social benefits are higher or lower with HFT firms in the market. In this sense, financial studies have been indicating that overall, the benefits that HFT provides in terms of liquidity, price discovery, and market efficiency, seem to be higher than the drawbacks. What is more, few are the papers that do not find any positive effect on HFT activity. So, with the appropriate precaution, it could be sustained that completely preventing these strategies by inadequate regulation or excessive barriers, may not be the best idea, as detrimental effects to market quality and efficiency could be triggered. Nevertheless, due to the lack of reliable empirical data, the research work becomes much more complicated, and at the same time, makes everyone more cautious at the time of presenting their conclusions. For the future, collaboration between regulatory agencies, prestigious researches, and HFT firms ranks as an absolute priority.

In short, the underlying conclusion of this paper is that regulators need to work to preserve the benefits of HFT while mitigating the problems. HFT will always generate debate, but ensuring a level playing field between all market's participants, stablishing appropriate risk management mechanisms (as measures to combat liquidity problems at market stress periods), ensuring supervision and communication (in order to increase trust), and preserving the economic rationale behind financial markets, the "beast" can be tamed. In this sense, various scholars and other market experts have proposed interesting alternatives that row in the appropriate direction, as the frequent batch auction market design presented by Budish et al. (2015), the deregulation of the minimum tick size in financial markets proposed by Gai et al. (2013), or the new Investors Exchange (IEX) created by Katsuyama and his team, but a lot of work still needs to be done.

REFERENCES

- Angel, J. J., Harris, L. E., and Spatt, C. S. (2015). Equity trading in the 21st century: An update. Quarterly Journal of Finance, 5.
- Arnuk, S., and Saluzzi, J. (2012). Brokern markets. Financial Times Pren Hall.
- Baldauf, M., and Mollner, J. (2015). High-Frequency Trading and Market Performance. Stanford Institute for Economic Policy Research, Stanford University (SIEPR Discussion Paper No. 15-017).
- Baron, M., Brogaard, J., and Kirilenko, A. (2014). Risk and Return in High Frequency Trading. SSRN Electronic Journal. doi: 10.2139/ssrn.2433118
- Battalio, R., Greene, J., and Jennings, R. (1997). Do Competing Specialists and Preferencing Dealers Affect Market Quality?. Review Of Financial Studies, 10(4), 969-993. doi: 10.1093/rfs/10.4.969
- Beckhart, B., and Keynes, J. (1936). The General Theory of Employment, Interest and Money. Political Science Quarterly, 51(4), 600. doi: 10.2307/2143949
- Bernales, A., & Daoud, J. (2013). Algorithmic and High Frequency Trading in Dynamic Limit Order Markets. SSRN Electronic Journal. doi: 10.2139/ssrn.2352409
- Biais, B., Foucault, T., and Moinas, S. (2015). Equilibrium fast trading, Journal of Financial Economics, 116 (2), 292–313
- Bogle, J. (2012). The clash of the cultures "investment vs. speculation". New York: John Wiley and Sons, Inc.
- Borkovec, M., Cochrane, J., Domowitz, I., and Escobar, C. (2014). The Cost of Liquidity in the FX Market. The Journal Of Trading, 10(1), 6-19. doi: 10.3905/jot.2014.10.1.006
- Breuer, A. (2013). An empirical analysis of order dynamics in a high-frequency trading environment. Sternenfels: Verl. Wissenschaft und Praxis.
- Brogaard, J. (2010). High Frequency Trading and Market Quality. SSRN Electronic Journal. doi: 10.2139/ssrn.1970072
- Brogaard, J., Garriott, C., and Pomeranets, A. (2014). High-Frequency Trading Competition. Bank Of Canada Working Paper 2014-19. doi: 10.2139/ssrn.2435999
- Brogaard, J., Hendershott T., Riordan R. (2014). High-Frequency Trading and Price Discovery. The Review of Financial Studies, Volume 27, 2267–2306.doi: 10.1093/rfs/hhu032
- Brown, A. and Yang, F. (2016). Slowing Down Fast Traders: Evidence from the Betfair Speed Bump. SSRN Electronic Journal. doi: 10.2139/ssrn.2668617
- Brunnermeier, M., and Pedersen, L. (2005). Predatory Trading. The Journal Of Finance, 60(4), 1825-1863. doi: 10.1111/j.1540-6261.2005.00781.x
- Budish, E., Cramton, P., and Shim, J. (2015). The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response. The Quarterly Journal Of Economics, 130(4), 1547-1621. doi: 10.1093/qje/qjv027
- Cao, C., Hansch, O., and Wang, X. (2009). The information content of an open limit-order book. Journal Of Futures Markets, 29(1), 16-41. doi: 10.1002/fut.20334
- Cartea, Á., and Penalva, J. (2012). Where is the Value in High Frequency Trading?. Quarterly Journal Of Finance, 02(03), 1250014. doi: 10.1142/s2010139212500140
- Chaboud, A., Chiquoine, B., Hjalmarsson, E., and Vega, C. (2014). Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market. The Journal Of Finance, 69(5), 2045-2084. doi: 10.1111/jofi.12186

- Cookson, C. (2013). Time is money when it comes to microwaves | Financial Times. Retrieved from https://www.ft.com/content/2bf37898-b775-11e2-841e-00144feabdc0
- Copeland, T., and Galai, D. (1983). Information Effects on the Bid-Ask Spread. The Journal Of Finance, 38(5), 1457. doi: 10.2307/2327580
- Crippen, A. (2014). Buffett, Gates and Munger criticize high-frequency trading. Retrieved from https://www.cnbc.com/2014/05/05/buffett-gates-and-munger-criticize-high-frequency-trading.html
- Duffie, D., Gârleanu, N., and Pedersen, L. (2005). Over-the-Counter Markets. Econometrica, 2005, v73, 1815-1848
- European Securities and Markets Authority. (2014). ESMA's Technical Advice to the Commission on MiFID II and MiFIR.
- European Securities and Markets Authority. (2018). MiFID II Supervisory briefing - Directive 2014/65/EU. Retrieved from https://www.esma.europa.eu/sites/default/files/library/2019-esma35-36-1640-MiFID_ii_supervisory_briefing_on_appropriateness.pdf
- Fama, E. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. Journal of Finance, 25, 2, p. 383-417.
- Foucault, T. (1999). Order flow composition and trading costs in a dynamic limit order market. Journal Of Financial Markets, 2(2), 99-134. doi: 10.1016/s1386-4181(98)00012-3
- Frazzini, A., Israel, R., and Moskowitz, T. (2012). Trading Costs of Asset Pricing Anomalies. SSRN Electronic Journal. doi: 10.2139/ssrn.2294498
- Gai, J., Yao, C., and Ye M. (2013). The Externalities of High Frequency Trading. SSRN Electronic Journal.
- Glosten, L., and Milgrom, P. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. Journal Of Financial Economics, 14(1), 71-100. doi: 10.1016/0304-405x(85)90044-3
- Gomber, P., Arndt, B., Lutat, M., and Uhle, T. (2011). High-Frequency Trading. SSRN Electronic Journal. doi: 10.2139/ssrn.1858626
- Gomber, P., Sagade, S., Theissen, E., Weber, M., and Westheide, C. (2013). Competition/Fragmentation in Equities Markets: A Literature Survey. SSRN Electronic Journal. doi: 10.2139/ssrn.2362216
- Jarnecic, E., Snape, M. (2010). An Analysis of Trades by High Frequency ParticipantsOn the London Stock Exchange. Working paper University of Sydney. Sydney, Australia
- Hagströmer, B., and Nordén, L. (2013). The diversity of high-frequency traders. Journal Of Financial Markets, 16(4), 741-770. doi: 10.1016/j.finmar.2013.05.009
- Harris, L. (2013) What to do about high-frequency trading. Financial Analysts Journal, 69, 6-9.
- Hasbrouck J., and Sofianos, G. (1993). The Trades of Market Makers: An Empirical Analysis of NYSE Specialists. The Journal Of Finance, 48(5), 1565-1593. doi: 10.1111/j.1540-6261.1993.tb05121.x
- Hasbrouck, J., and Saar, G. (2013). Low-latency trading. Journal Of Financial Markets, 16(4), 646-679. doi: 10.1016/j.finmar.2013.05.003
- Hemsley, M. (2014). CEO of the BATS Chi-X Europe. In European Capital Markets Institute's Annual Conference. Brussels.
- Hendershott, T., and Riordan, R. (2011). Algorithmic Trading and Information. SSRN Electronic Journal. doi: 10.2139/ssrn.1472050

- Hendershott, T., Jones, C., and Menkveld, A. (2011). Does algorithmic trading increase liquidity? Journal of Finance 66:1–33
- Hoffmann, P. (2014). "A dynamic limit order market with fast and slow traders," Journal of Financial Economics, 113 (1), 156–169
- Huang, R. D.(2002). The Quality of ECN and NASDAQ Market Maker Quotes. Journal of Finance, 57, issue 3, p. 1285-1319.
- Investment Industry Regulatory Organization of Canada. (2011). Retrieved from http://www.iiroc.ca/Documents/2011/83e897a0-84b4-40f4-bcb1-106062612c13_en.pdf#search=getco
- Jones, C. 2013. What do we know about high-frequency trading? : Columbia Business Schoo.Research Paper No. 13-11.
- Kervel, V., and Menkveld, A. (2015). High-Frequency Trading around Large Institutional Orders. SSRN Electronic Journal. doi: 10.2139/ssrn.2619686
- Kirilenko, A., Kyle, A., Samadi, M. and Tuzun, T. (2016) The flash crash: High frequency trading in an electronic market. The Journal Of Finance, 72(3), 967-998. doi: 10.1111/jofi.12498
- Klock, M., and McCormick, D. (1999). The Impact of Market Maker Competition on NASDAQ Spreads. The Financial Review, 34(4), 55-73. doi: 10.1111/j.1540-6288.1999.tb00469.x
- Kyle, A. (1985). Continuous Auctions and Insider Trading. Econometrica, 53(6), 1315. doi: 10.2307/1913210
- Lattemann, C., Loos, P., Gomolka, J., Burghof, H., Breuer, A., and Gomber, P. et al. (2012). High Frequency Trading. Business and Information Systems Engineering, 4(2), 93-108. doi: 10.1007/s12599-012-0205-9
- Lewis, M. (2014). Flash boys. Norton and Company.
- Malinova, K., Park, A., and Riordan, R. (2012). Do Retail Traders Suffer from High Frequency Traders?. SSRN Electronic Journal. doi: 10.2139/ssrn.2183806
- Mayhew, S. (2002). Competition, Market Structure, and Bid-Ask Spreads in Stock Option Markets. The Journal Of Finance, 57(2), 931-958. doi: 10.1111/1540-6261.00447
- Menkveld, A. J. (2013) High frequency trading and the new market makers. Journal of Financial Markets, 16, 712-740.
- Menkveld, A., and Zoican, M. (2017). Need for Speed? Exchange Latency and Liquidity. The Review Of Financial Studies, 30(4), 1188-1228. doi: 10.1093/rfs/hhx006
- Meyer, G., Bullock, N., and Rennison, J. (2009). How high-frequency trading hit a speed bump – The Financial Times. Retrieved from https://www.ft.com/content/d81f96ea-d43c-11e7-a303-9060cb1e5f44
- MiFID II / MiFIR series. (2014). Retrieved from https://www.nortonrosefulbright.com/en/knowledge/publications/6d7b8497/MiFID-ii-mifir-series
- O'Hara, M. (2015). High frequency market microstructure. Journal Of Financial Economics, 116(2), 257-270. doi: 10.1016/j.jfineco.2015.01.003
- Panayides, M. (2007). Affirmative obligations and market making with inventory. Journal of Financial Economics 8:513–42.
- Perez, E. (2011). The speed traders. New York: McGraw-Hill.
- Riordan R., and Zhang, S. (2011). Technology and market quality: The case of high frequency trading. In: Proceedings of the ECIS 2011.

- Sannikov, Y., and Skrzypacz, A. (2016). Dynamic Trading: Price Inertia and Front-Running. SSRN Electronic Journal. doi: 10.2139/ssrn.2882809
- Scholtus, M., van Dijk, D., and Frijns, B. (2014). Speed, algorithmic trading, and market quality around macroeconomic news announcements. Journal Of Banking and Finance, 38, 89-105. doi: 10.1016/j.jbankfin.2013.09.016
- Schwartz, R., and Francioni, R. (2004). Equity markets in action. Hoboken, N.J.: J. Wiley and Sons.
- SEC, 2016. Disclosure of order handling information. Securities Exchange Act Release No. 78309 (July 13, 2016), 81 FR 49432, 49434 (July 27, 2016).
- Securities and Exchange Commission. (2010). Concept release on equity market structure, Release No. 34-61358. File No. S7-02-10.
- Securities and Exchange Commission. (2016). The New FINRA Registration Requirement for Algorithmic Traders: Implications for Broker-Dealers and Investment Advisers. The Financial Industry Regulatory Authority, Inc.
- Stiglitz, J. (2013). Inequality Is Holding Back the Recovery. The New York Times.
- Stoll, H. (1978). The Pricing of Security Dealer Services: An Empirical Study of NASDAQ Stocks. The Journal Of Finance, 33(4), 1153. doi: 10.2307/2326946
- Summers, L. (1986). Does the Stock Market Rationally Reflect Fundamental Values?. The Journal Of Finance, 41(3), 591. doi: 10.2307/2328487
- Tapia, M., and Alubakundi, B. (2014). ¿Qué sabemos de la negociación de alta frequencia?. Bolsas y Mercados, innovación y tecnología.
- Technology Advisory Committee | U.S. Commodity Futures Trading Commission. (2018). Retrieved from https://www.cftc.gov/About/CFTCCommittees/TechnologyAdvisory/index.htm
- Techstars. (2017). History of High Frequency Trading (HFT) – Infographic. Retrieved from https://www.techstars.com/content/startup-digest/history-high-frequency-trading-hft-infographic/
- Thaler, R. (1988). Anomalies: The Winner's Curse. Journal Of Economic Perspectives, 2(1), 191-202. doi: 10.1257/jep.2.1.191
- The Netherlands Authority for the Financial Markets. (2016). A case analysis of critiques on high-frequency trading.
- Tobin, J. (1978). A Proposal for International Monetary Reform. Eastern Economic Journal, 4, 153-159.
- Twaronite, L. (2017). Japan passes law to tighten regulations on high-frequency trading. Retrieved from https://uk.reuters.com/article/uk-japan-regulations-hft/japan-passes-law-to-tighten-regulations-on-high-frequency-trading-idUKKCN18F0TU
- U.S. Commodity Futures Trading Commission, the U.S. Securities, and Exchange Commission. (2010). Findings regarding the market events of may 6, Report, SEC, September 2010.
- Securities and Exchange Commission (2014). Equity Market Structure Literature Review, Part II: High Frequency Trading.
- UK Home | Investopedia. (2019). Retrieved from https://www.investopedia.com/
- Van Ness, B., Van Ness, R., and Warr, R. (2005). The impact of market maker concentration on adverse-selection costs for NASDAQ stocks. Journal Of Financial Research, 28(3), 461-485. doi: 10.1111/j.1475-6803.2005.00134.x

- Vayanos, D. (1999). Strategic Trading and Welfare in a Dynamic Market. Review Of Economic Studies, 66(2), 219-254. doi: 10.1111/1467-937x.00086
- Weller, B. (2013). Intermediation chains and specialization by speed: Evidence from commodity futures markets. Chicago: University of Chicago, Division of the Social Sciences, Department of Business and Economics, 2013.
- Weston, J. (2000). Competition on the NASDAQ and the Impact of Recent Market Reforms. The Journal Of Finance, 55(6), 2565-2598. doi: 10.1111/0022-1082.00300