A SURVEY OF LITERATURE ON HIGH-FREQUENCY TRADING

Khairul Zharif Zaharudin
PhD student at Massey University, Palmerston North
Tutor at Universiti Utara Malaysia

1.0 Introduction

Technological advancement has shaped the financial world. Prior to the invention of the telegraph in 1844 and the telephone in 1876, communication in securities markets had been primitive – using humans and carrier pigeons to transmit information across the markets (Markham, 2002). For nearly a century, telegraph and telephone are used as the main channel for financial communication – data is received via telegraphic stock ticker, and orders are transmitted via phone calls.

However, in recent years, fiber-optic cables and microwave towers are used as the medium to transfer trading information, traveling at lightning speed. A group of traders, armed with complex algorithms, are willing to spend a large amount of money to gain access to these state-of-the-art facilities and pay to collocate their server within stock exchanges, as those services give them the speed advantage that they need for their trading strategies that banks on being the fastest. In addition, they hire mathematicians and statisticians to work as quantitative analysts or “quants”, to develop the various trading algorithms. This unique group of traders is commonly referred to as “high-frequency traders”, or HFT in short.

In the U.S., HFT’s market share in total equity trading peaked at around 60% in 2009, from around 20% in 2005. The percentage gradually decrease to approximately 50% in 2013 and has been stable until 2016 (Avramovic, Lin, & Krishnan, 2017). In Europe, HFT’s contribution to total equity trading was almost 0% back in 2005, before reaching its highest point at around 40% in 2010, and settles at around 35% in 2014 (Kaya, 2016). Australian Securities and Investments Commission (ASIC) states that HFT accounts for approximately 27% of all equity market turnover in S&P/ASX 200 securities, from January 2012 until March 2015. Despite the stable market share, there is more concentration in the HFT-driven volume – 10 largest HFT account for 21% of all trading turnover in 2015, compared to 17% three years earlier (ASIC, 2015).

This paper is a survey of literature on HFT, which cover the various HFT definitions. We present how HFT works, and what makes HFT different than other groups of investors, and proceeds with a discussion on beneficial HFT strategies (e.g. market-making, directional trading, and statistical arbitrage) and harmful HFT strategies (e.g. front-running, spoofing, and quote-stuffing). The paper continues with an argument on the effects of HFT activities on market quality (e.g. price discovery, liquidity, and adverse selection cost), and wrapped with a discussion on several critical issues associated with HFT (the Flash Crash, the arms race, and market-making obligation).

1.1 Defining HFT

High-frequency trading, or HFT, as noted by the U.S. Securities and Exchange Commission (SEC) has no clear definition (SEC, 2010, 2014). As there is no standard definition of HFT to date, regulators, researchers, and market participants have different ways to describe HFT. The term “high-frequency trading” is typically associated with "trading that utilizes computer technology", "the use of technology to execute orders", "electronic trading", “electronic markets", or “automated trading”. While the terms are indeed closely related to HFT, they are not the same thing, and only portray an incomplete picture of HFT.

The Netherlands' Authority for the Financial Markets (AFM) states that the absence of a unanimous definition of HFT also makes classification difficult, which leads to other problems such as inaccurate estimation of HFT' market shares, and inability to estimate the reach and influence of HFT in their markets. This lack of consensus on HFT definition complicates research conducted in this area and contributes to the various conclusion on the effect of HFT's activity in the market. The inexistence of precise definition of HFT also leads to confusion of HFT with other forms of activities, and consequently, blamed for things that have nothing to do with it (Moosa & Ramiah, 2015).

The incomplete definition of who or what is HFT is a problem to both HFT-existed and HFT-free markets alike. Financial authorities need to meticulously analyze and consider the costs and benefits of having HFT in their market. However, before they can effectively tackle the issue, first and foremost, they need to have a sound definition of HFT. An inaccurate definition would be too costly to the market – any microstructural changes introduced will likely involve a huge sum of money and may affect all class of market participants, from the smallest individual investors to the mutual fund giants.

Zhang (2010) broadly defines HFT as all short-term trading activities by hedge funds and other institutional traders not captured in the 13f database. Kirilenko, Kyle, Samadi, and Tuzun (2017) describe HFT as traders with high volume and low inventory, and Baron et al. (2012) added low overnight inventory to the list. Moosa and Ramiah (2015) define HFT based on its characteristics i.e. data-intensive, latency-sensitive, high-volume, low-margin activity, extremely short holding periods, and rarely held positions overnight. Other scholars define HFT as a large number of small-quantities orders, high-speed order cancellations, and have short position-holding periods (Aldridge, 2009; Brogaard, 2010; Gomber, Arndt, Lutat, & Uhle, 2011).

SEC (2010) refers to HFT as "...professional traders acting in a proprietary capacity that engage in strategies that generate a large number of trades on a daily basis" (p. 45). SEC (2010) also lists down five characteristics commonly attributed to HFT: (1) use of extraordinarily high-speed and sophisticated computer programs for generating, routing, and executing orders; (2) use of co-location services and individual data feeds offered by exchanges and others to minimize network and other types of latencies; (3) very short time-frames for establishing and liquidating positions; (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. Regardless, SEC never suggests that all of the aforementioned characteristics should be met for a firm to be categorized as HFT. By doing so, a broader range of proprietary firms can be classified as HFT (SEC, 2014).

In 2010, AFM produced a report on HFT to shed some lights on the new phenomenon. AFM (2010) defines HFT as a form of automated trading based on mathematical algorithms that implement certain short-term trading strategies by utilizing advanced technology, and not as a separate trading strategy by itself. The main characteristics of HFT according to AFM (2010) are: (1) use trading strategy that involves rapid calculation and execution speeds; (2) use sophisticated systems and efficient infrastructures; (3) use earnings model with very small profit margins in very large volumes; (4) usually take market-neutral (non-directional) and delta neutral (hedged) position, thus in many cases close out their positions with flat position at the end of the day; (5) have a really short average holding period, ranging from seconds to several minutes; and (6) have a very high order-to-transaction ratio.

Similarly, ASIC also agrees to the notion that there is no unanimous definition of HFT. ASIC (2010) characterised HFT as a specialised form of algorithmic trading that (1) generate large numbers of small size orders with high rate of amendment and cancellation; (2) typically have to hold positions with very short time horizons; and (3) use variety trading strategies, but the most common strategy is electronic liquidity provision. HFT also employ high-speed, low-latency technology infrastructures which requires them to: (1) process direct market feeds to have access to the fastest market information available; (2) co-locate their servers in the data centres with the exchange market's matching engine to reduce access times; (3) develop their own sophisticated trading strategies to trade on a short-term basis; and (4) typically end the trading day with no carry-over positions that use capital (ASIC, 2010).

In 2010, the Committee of European Securities Regulators (CESR) conducted a survey to call for evidence on micro-structural issues of the European equity markets. The survey is intended to assess the impact of technology-driven developments such as HFT, sponsored access, and co-location services that have intensified following the implementation of the Markets in Financial Instruments Directive (MiFID) on November 1, 2007. In the survey, CESR (2010) describes HFT: (1) as a form of automated trading that uses sophisticated computers and IT programs; (2) execute trades in matters of milliseconds; (3) hold new equity positions possibly down to a "sub-second"; (4) ends their day with a flat position; (5) use their own capital and do not act on behalf of clients; and (6) and employ trading strategies that are generally geared towards extracting very small margins from hyper fast speed trading. In a response to the survey, London Stock Exchange Group (LSE) refers to HFT as a wide variety of different strategies utilizing ultra-fast technology to conduct electronic market-making and/or to seek arbitrage opportunities. LSE (2010) also noted that HFT is very fast and requires low-latency connection to exchanges' trading systems.

The introduction of Directive 2014/65/EU of the European Parliament and of The Council of May 15, 2014, on markets in financial instruments, sees the amendment of the MiFID. The new directive (commonly referred to as MiFID II) provides a legal definition of HFT. Article 4(1)(40) of MiFID II describes a HFT technique as "an algorithmic trading technique characterised by: (a) infrastructure intended to minimise network and other types of latencies, including at least one of the following facilities for algorithmic order entry: co-location, proximity hosting or high-speed direct electronic access; (b) system determination of order initiation, generation, routing or execution without human intervention for individual trades or orders; and (c) high message intraday rates which constitute orders, quotes or cancellations".

Brogaard, Hendershott, and Riordan (2014) state "one of the difficulties in empirically studying HFT is the availability of data identifying HFT. Markets and regulators are the only sources of these and HFT and other traders often oppose releasing identifying data" (p. 2270). An example of such dataset is the one provided by NASDAQ, which covers 120 U.S. equities over the 2008-2009 period, timestamped to the milliseconds. NASDAQ used its access to order-level information on its market to identify the firms submitting orders, and manually classified 26 of the firms as HFT. The dataset also categorizes whether the execution is either aggressive (liquidity taking) or passive (liquidity providing), and further grouped them into either HFT or non-HFT, resulting in four types of order execution: "HH": HFT take liquidity from other HFT; "HN": HFT take liquidity from non-HFT; "NH": non-HFT take liquidity from HFT; and "NN": non-HFT take liquidity from other non-HFT.

Even so, the data has its limitations. NASDAQ cannot identify all HFT in the market and possibly has excluded HFT firms that also act as brokers while engaging in proprietary lower-frequency trading strategies (e.g. Goldman Sachs, Morgan Stanley). Thus, the orders from HFT firms routed through those large integrated firms might be excluded as well (Brogaard et al., 2014). In similar note, Hagstromer and Norden (2013) assert that the use of mediation trading services such as sponsored access and/or trading desks of banks, which consist a mixture of HFT and non-HFT, makes it difficult to distinguish the origins of the trading activity, and to interpret the results obtained from this group. According to AFM (2010), even with an accurate definition of HFT, trading platforms would still be unable to properly distinguish HFT from other forms of AT. To do so, the trading platforms need to establish market share of the various trading strategies that employed AT, in which based on today's technology, is not yet possible.

Albeit not having a conclusive definition of HFT, certain characteristics distinguishing HFT from other forms of trading can be specified. In general, the majority of the regulatory body agree that

HFT is: (1) a specialised form of algorithmic trading; (2) use high-speed, sophisticated computer programs and systems; (3) have a very high order-to-transaction ratio; (4) have extremely short average holding periods; and (5) end their trading day with flat positions.

2.0 HFT mechanics and Strategy

There is nothing new in the way HFT works. The short-term trading strategies employed by HFT has long existed (AFM, 2010). The way HFT profit from the market, in general, is similar to other traders' strategy. For instance, they will buy stocks at a lower price, then sell them at a higher price. For stocks with short selling option, they have more choices – they are able to make money from either bearish or bullish stocks. For stocks with options, they could make a profit from price disparities between the parent stocks and their option securities. They might as well assume the role of a market-maker, in which they stand ready to buy and sell securities, and profit from the market-making spread. Moreover, market-making HFT might receive rebates from certain trading venues for providing liquidity to their market.

What makes HFT a unique class of investor lies in their speed; to observe for profitable trading opportunities, to quickly process new information and execute the appropriate action, to analysis textual context of news flow, at a much higher-frequencies and shorter time-frames than a human being capable to. This is also an important advantage that the machines (i.e. HFT) have over humans (Menkveld, 2014). The infrastructural and technological advantages that they possess allow for the optimization of a wide-array of complex trading strategies, from the beneficial market-making strategies to the harmful and devious strategies such as quote stuffing (AFM, 2010; O'Hara, 2015). According to Angel (2014), the trading speed nowadays is so fast that it almost reached the theoretical speed of light – approximately 300,000 km/s. This superhuman speed also makes certain trading strategies exclusive to HFT, especially the ones that rely on speed, as other types of market participants unable to replicate such strategies (Harris, 2013), which further stressing HFT's need for superior calculation and execution speeds (AFM, 2010).

Hasbrouck and Saar (2013) find that some algorithms are so fast that the time it takes to complete a trading cycle starting from the detection of a market event, analyses it, and send an order appears to be 2–3 milliseconds. This intense activity within the "millisecond environment" is also where computer algorithms react to each other (Hasbrouck & Saar, 2013). O'Hara (2015) states that order latencies are now measured in milliseconds (one-thousandth of a second), microseconds (one-millionth of a second), and even nanoseconds (one-billionth of a second). For comparison purpose, it takes the human eye 400–500 milliseconds to respond to visual stimuli, and human reaction times are generally thought to be around 200 milliseconds, which in both cases is far behind the HFT's speed (Kosinski, 2013; O'Hara, 2015). At such speeds, human traders cannot accurately follow the low-latency activity on their trading screen, and the market dynamics that may be driven by the interactions between algorithms (Chordia, Goyal, Lehmann, & Saar, 2013; Hasbrouck & Saar, 2013). Due to HFT strategies depends heavily on speed, latency issues such as the speed of cross-market information flow, and transmission speed across geographical locations are of their concern. Therefore, to utilize their trading strategies, many HFT would have multiple locations across several cities such as New York, Chicago, and London.

The earnings model for HFT consists of executing many transactions with very small profit margins in very large volumes (AFM, 2010). Using fully automated trading strategies, HFT attempt to identify and profit from short-term irregularities, and earn small amounts of money from every trade. Even though the profit per trade is often as small as a few basis points only, it is amplified by the high trading volume (Zhang, 2010). The ability to trade at low latency allows HFT to profit from the trading environment itself, rather than from investing in financial securities (Hasbrouck & Saar, 2013). Budish, Cramton, and Shim (2015) state two common characteristics used by HFT in their trading strategies, which are (1) often cancel their orders soon after placing them, and (2) high-ratio of messages to completed trades.[1]

HFT exhibit variability in their trading strategies by documenting differences in liquidity provision, end-of-day and maximum intra-day positions, trading revenues, etc. The variability in strategies also translates into different sensitivities of HFT' position changes to inventory levels and to recent price changes (Benos & Sagade, 2016). Brogaard et al. (2014) find that the direction of HFT trading is correlated with publicly available information, such as macroeconomic news announcements and limit-order book imbalance. They also find that HFT followed contrarian trading strategies, evidenced by the negative correlation between HFT overall trading with past returns. Goldstein, Kumar, and Graves (2014) state that naturally, the HFT strategies are employed by proprietary firms, in which the majority are either broker-dealer proprietary trading desks,[2] hedge funds,[3] and proprietary trading groups.[4] This is only logical due to the high cost involved in employing sophisticated technology and obtaining the big data to execute HFT strategy (Moosa & Ramiah, 2015; Kauffman, Hu, & Ma, 2015).

Aldridge (2013) generally categorized HFT trading strategies into three groups, which are (1) statistical arbitrage, also known as value-motivated strategies; (2) directional strategies, also known as informed trading; and (3) market-making, also known as liquidity trading. The algorithms employed by HFT may determine their order execution style, such as either being aggressive or passive or to send the orders in either one trade or split them into smaller trades.[5] Similarly, AFM (2010) also divides HFT strategies into market-making, statistical arbitrage, and low-latency. While the first two groups are similar to Aldridge's (2013), the third group classification, i.e. low-latency, has a broader scope. AFM (2010) states that the success factor of the latter group is determined by the sheer speed of the users, hence, creating the need to have the fastest systems and the best connection to trading venues. Harris (2013) on the other hand, grouped HFT trading strategies into three (3) groups based on their effect on the market. The first group, Valuable, is a group of trading strategies that are acceptable to the market in general, such as market-making and statistical arbitrage. The other two groups, namely Harmful and Very Harmful, are a group of trading strategies that is intolerable, with the latter worse than the former. The strategies belong to these groups benefits the HFT at the cost of other market participants, such as front-running and quote stuffing.

Most HFT-based strategies such as market-making promote market liquidity, while the arbitrage strategies have a positive contribution to price discovery and market efficiency. Therefore, the action to prevent or hamper these strategies by inadequate regulation, or imposing specific constraints for this group of strategies, may trigger counterproductive effects to market quality. Regardless, regulatory bodies should always combat any predatory strategies that go against market integrity or create disruptive or confusing effects on other market participants (Gomber et al., 2011). Harris (2013) highlights that the financial authorities should be meticulous in regulating the market, to avoid from unintentionally harming friendly HFT strategies. Cooper, Davis, and Vliet (2016) examine regulatory

[1] Regulatory bodies intend to introduce minimum resting time and impose maximum order-to-trade ratio, in which each rule is aimed to address the aforementioned characteristics respectively.

[2] Proprietary trading desks is a trading unit in a firm such as banks, that trade using the firms' own money to make profit, instead of relying on commissions from their clients. In the U.S., the Volcker Rule prohibits banks from engaging in high-risk, speculative trading activity on their own account, such as the short-term proprietary trading.

[3] Example of hedge funds that employ HFT strategies are Renaissance Technologies, Worldquant, DE Shaw, and Millennium.

[4] Example of proprietary trading groups that use HFT strategies are Getco LLC, Allston Trading LLC, Infinium Capital Management LLC, Hudson River Trading LLC, Quantlab Financial LLC.

[5] An aggressive order is an order that is placed on the current market price, a.k.a. market order, or a limit-order with price near to the current market price. A passive order on the hand is a limit-price placed far from the current market price.

efforts related to HFT, particularly on the issue of HFT's deception in the market. They conclude that the action to treat a deception, or even an intentional deception, as a misconduct in a financial market, is a mistake. They outlined three (3) acceptable criteria for algorithm trading strategies, which are; (1) the trading strategy should be prudent, in which it would not be harmful to the market should they behave unexpectedly; (2) the trading strategy should not block price discovery, i.e. it should not interfere with the ability of other market participants to reflect their private information on the price; and (3) the trading strategy should not circumvent transparent price discovery, and therefore, strategies that conceal information from being discovered, such as using dark pools or hidden orders, should be prohibited.

2.1 Beneficial strategies

The following sections provide a brief a discussion on acceptable HFT trading strategies, namely statistical arbitrage, directional trading strategies, and market-making. These strategies are deemed as acceptable as they do not harm the market and have positive effects on market quality.

2.1.1 Statistical arbitrage

Statistical arbitrage, also commonly known as "stat arb", is a trading strategy that is based on the theory that two similar instruments should share similar behavior, and therefore, any short-term divergence between their relative prices are likely to converge again. The temporary divergence is more likely to be driven by momentary order imbalance in the market, rather than by any meaningful fundamental change (Narang, 2013).[6] This trading strategy is designed to make a profit from price disparity, and temporary deviations of statistically significant relationships,[7] while considering tens or hundreds of stocks to utilize this strategy (Lhabitant & Gregoriou, 2015; Golub, Dupuis, & Olsen, 2013; Moosa & Ramiah, 2015). Accordingly, HFT will hunt for the opportunities that arise during the temporary deviations period and exploit them before the phenomenon disappears (Moosa & Ramiah, 2015).

Wissner-Gross and Freer (2010) highlight the importance of minimizing information transmission delay in modern-day securities trading. In their paper on relativistic statistical arbitrage, they demonstrated that there exist optimal intermediate locations between trading centers that host cointegrated securities, which minimizes transmission delays and maximizes profit potential. As traders continue to aim at being the fastest, the importance of having optimal locations is even more pronounced (Donner, 2010; Wissner-Gross & Freer, 2010). Regardless, Kozhan and Tham (2012.) argue that while competition is commonly associated with improved price discovery, competition among arbitrageurs might inflict negative externalities on each other due to the crowding effect, which in turn will limit efficiency.

The opportunities for statistical arbitrage might surge from long-term investors' strategic decision. For instance, their action to buy or to sell certain securities might create a price impact on the securities' price, which consequently create a ripple of price impact across the markets, especially in correlated securities (Goldstein et al., 2014). The fastest trader who first notices such opportunities and trades on them will make the most, if not take all, of the profits from the phenomenon. Therefore, speed is essential to successfully execute this trading strategy, and HFT that implement this strategy are willing to spend a lot to keep their technological capabilities up-to-date (Chung & Lee, 2016;

[6] This trading strategy is also commonly known as "pairs trading". Although in theory it is possible to find directly comparable instruments, however, very few assets can be compared precisely with another instrument, rendering the potential benefits from this strategy to be infeasible (Narang, 2013).
[7] HFT might use statistical approach that measures the relationship between two or more instruments such as cointegration or correlation analysis.

MacIntosh, 2015). This strategy plays a key role in the market in terms of liquidity provision, as well as in price discovery and information dissemination process (Goldstein et al., 2014). Nonetheless, Hasbrouck and Saar (2013) argue that even though HFT helps in eliminating momentary price distortions but given that the improvement is only within millisecond environment, the effect is deemed insubstantial.

2.1.2 Directional trading

Directional trading strategies is a group of high-frequency trading strategies based on the identification of short-term trend or momentum, which includes event-driven strategies and short-term price movements prediction strategies. Directional strategies are time-sensitive (Aldridge, 2013), as they need to anticipate an intraday price movement, which involves taking un-hedged positions based on forecasted price changes, such as exploiting the divergence between fundamental values and actual market prices. Benos and Sagade (2016) find that HFT with neutral liquidity taking/making behavior is trend chasers. They trade in the direction of short-term price changes, i.e. they buy when the price is rising and sell when the price is falling, which is suggestive of momentum strategies.

Directional strategies are based on the theory that the price movement has directions and they are predictable, which might be following a trend (momentum strategies) or reversal of a trend (mean reversion strategies). Under the momentum strategy, HFT will identify a trend or a significant move, and bet that the price will continue to move in the same direction, driven by the idea of there is a growing consensus among market participants (Narang, 2013). The mean reversion strategy, on the other hand, is built on the notion that any deviations in price, such as a trend or a consistent direction, may be temporary in nature. Thus, price movements do not persistently move in one direction, and will eventually revert and bounce back (Easley, Prado, & O'Hara, 2012).

To be successful in implementing directional trading strategies, HFT needs to have superior access to information (e.g.: information from paid-for news sources such as Bloomberg) and able to immediately assess and analyze market condition. Foucault, Hombert, and Rosu (2016) suggest that the contribution of news trading to the directional trader's profit increases with news informativeness, and the fastest traders will gain the most profit. Furthermore, the competitive edge that directional traders have from the early access to new information will not last long, as the information will soon be available to the public. Thus, the directional traders are normally aggressive, as they use market orders or post limit prices close to market (Aldridge, 2013).

2.1.3 Market-making

Market-making, in general, can be described as the placement of limit orders on both sides of the market price, in which limit buy (sell) orders are placed just below (above) the market price, which provides liquidity to the market. HFT market-making strategies help the market to be more efficient and have stabilizing effects to the market as they (the HFT) provide buying power when others want to sell and selling power when others want to buy (Angel, 2014). Despite the financial landscape has developed so much as technology evolves, the general mechanics of market-making still hold even in a high-frequency world. Goldstein et al. (2014) state that market-making HFT uses automated liquidity provision, a strategy which rapidly places, cancel, and replace bid (buy) and ask (sell) limit orders, and profit from the resulting spreads. The high-frequency updating process involves in the market-making process resulted in enormous orders volume and a high cancellation rate of 90% or more (SEC, 2010).

Unlike HFT that uses directional trading strategies, HFT market-makers do not seek to make a directional bet, but instead, take a position on both sides of the order book to maximize their inventory turnover. HFT market-makers typically would turn over their inventory more than five times in a day, which explains their high share of volume traded in the market. They also hold minimum or even zero inventory positions at the end of a trading day. Since they have very small inventories and short holding period, essentially, they could perform their market-making activities with very low capital, while using high-speed trading to control their position risk (Easley, Prado, & O'Hara, 2011). Benos and Sagade (2016) find evidence that passive HFT is consistent with market-making activity, in which they trade in the opposite direction (i.e. contrarian trading) of the most recent price changes, post limit orders, and use aggressive trade to make quick inventory adjustments.[8] Regardless, they also find that passive HFT has a high information-to-volume ratio, suggesting that the HFT might use various market-making strategies, rather than solely using aggressive orders to make the market.

Aldridge (2013) states that a market-maker is exposed to two types of risk once his market limit orders are placed, which are (1) inventory risk and (2) adverse selection risk. Inventory risk is the risk that the inventory that a market-maker is holding decline in value due to natural market movements, while adverse selection risk is the possibility of the market-maker is trading against a party that is better-informed about the true price of the stock. Thus, it is only natural that the market-maker to be compensated not only for the liquidity-providing service that they provided, but also the risk they have to bear from their role as a market-maker (Aldridge, 2013; Golub et al., 2013).

Some electronic exchanges use maker-taker pricing model to price their order-matching service (Harris, 2015).[9] Durbin (2010) defines the model as "a pricing policy of some exchanges where active traders pay a fee, some of which is distributed to the associated passive trader" (p. 206). The maker-taker pricing model is used to encourage market-making instead of market-taking activity in the market through incentives in the form of rebates or reduced transaction costs to market-makers. The rebate is indeed important for market-making HFT. The absence of rebate would put HFT in a loss position (Hendershott & Riordan, 2013), and their revenue from supplying liquidity would be negative (Brogaard et al., 2014), which in turn may discourage HFT's liquidity provision activities.

2.2 Harmful strategies

The controversial strategies are the strategies that profit at the expense of others through "dirty" means such as front-running, order anticipation, quote-matching, quote-stuffing, spoofing, and layering. Moreover, HFT' ability to rapidly enter and cancel orders faster than other traders makes it difficult to identify where liquidity exists across fragmented markets and this uncertainty creates even more profitable opportunity for HFT (O'Hara, 2015).

2.2.1 Front-running

Harris (2013, 2015) describe front-running as "very harmful" trading strategies, and further categorized them into "order-anticipating" and "quote-matching" strategy. Order-anticipation works by examining trades and quotes to detect algorithms used by traders that intend to move large orders.[10] The HFT would then trade ahead of (i.e. front-run) the incoming large orders and profit from the anticipated direction of the price changes. This will make the price higher (lower) for incoming large buy (sell) orders, which increase the transaction costs for traders intended to execute a large order. HFT that apply

[8] Benos and Sagade (2016) categorized HFT based on their liquidity taking/making behaviour. For computational details of the measure, kindly refer to their paper at https://doi.org/10.1016/j.finmar.2016.03.004

[9] The maker-taker pricing model is criticized for causing distortion in the market (Angel, 2014, *FR*), providing unfair advantages to high-speed traders. This issue is further discussed in the *Section 2.7.4: Issues on maker-taker pricing model*.

[10] A trader will split their large orders into "smaller packages" to conceal their private information and reduce the impact on the market. In this aspect, this is quite similar to the reason traders use dark pools to trade their large orders.

order anticipation strategy cleverly design their algorithms to play by the book, without violation of a duty, misappropriation of information, or other misconduct (SEC, 2010). Regardless, the strategy that they use is "parasitic" – not only it does not contribute to price discovery or liquidity, but it also preys on other traders and jeopardize the large traders the most (Harris, 2015). Some institutional investors even claim that the order-anticipation strategy may adversely affect their trading strategy, which impacts costs for these institutional investors (Agarwal, 2012).

Quote matching, on the other hand, make profits by posting slightly better limit order, e.g. one-tick higher (lower) than slow traders' limit buy (sell) orders, which gives them price-priority. In the case of the market is moving against their position, quote-matching HFT would trade with the slower traders' quotes (which has become the best quotes) to minimize their loss. The problem of quote matching is not something new to the large buy-side traders. It was an important source of profit for exchange specialists before the era of HFT. The main difference between now and then is the identity of the quote matchers (Harris, 2013). Unlike the order anticipation strategy that requires high-quality pattern-recognition algorithms, the success of quote matching strategy highly depends on HFT's low-latency communication. Speed is crucial to quote-matchers to get their orders be the first to fill the large orders, also to revise their unexecuted orders should the large orders are canceled or filled before they can be matched, thus, it is dominated by the faster HFT (Harris, 2013). Nevertheless, both strategies unnecessarily increased the large traders' transaction costs (Chung & Lee, 2016), and may impede the process of impounding fundamental information into the price (Jarnecic & Snape, 2014).

Aquilina and Ysusi (2016) empirically examine HFT order anticipation activity using data from LSE and find no evidence that HFT systematically anticipate orders sent to different venues by non-HFT, and try to front-run the orders. However, they do find trading patterns consistent with HFT anticipate non-HFT' order flow when analyzing longer time periods. Regardless, the result can also mean that the HFT able to react faster to news and other public information than non-HFT. They conclude that "HFT appear not to anticipate near-simultaneous orders…but they could be predicting the flow over longer time periods" (p. 26).

2.2.2 Spoofing

Spoofing and layering are defined as a strategy that:

> Submitting multiple orders at different prices on one side of the order book slightly away from the touch, submitting an order to the other side of the order book (which reflects the true intention to trade) and, following the execution of the latter, rapidly removing the multiple initial orders from the book (European Securities & Markets Authority, 2011, p. 27).

Financial Industry Regulatory Authority (FINRA, 2012) in general describe spoofing as a form of market manipulation intended for "triggering another market participant(s) to join or improve the NBBO, followed by canceling the non-bona fide order and entering an order on the opposite side of the market" (para. 5). Dodd-Frank Wall Street Reform and Consumer Protection Act, on the other hand, outlined a broader definition of spoofing, defined as a disruptive practice that involves "bidding or offering with the intent to cancel the bid or offer before execution" (p. 364), which makes it unlawful to practice such strategy.

Spoofing is executed with the intention to attract liquidity by posting fake market or limit orders to mislead other investors, especially algorithms specialize in tape reading (Serbera & Paumard, 2016), by forming an illusion that the market is moving soon due to a great demand in the order book. For example, HFT may create such situation by posting large displayed limit orders just below the best bid price, leaving others under the impression that the price will soon move upwards. This situation encourages other traders to quickly buy the stock by quoting the stock at a higher bid price or even execute market orders. In the meantime, the HFT might already own the stocks beforehand, and can now sell them at a higher price in a bigger volume, thanks to the artificially inflated price that was driven by the fake limit buy orders.

Layering is a form of spoofing, which involves placement of a large number of fake orders on several different price limits on one side of the order book (AFM, 2010). This creates an appearance of changing levels of supply and demand in the affected securities (FINRA, 2012). Others may falsely interpret this pattern as a signal of an increasing directional pressure on the price and act accordingly. The HFT will then make a profit from the price move they have initiated and cancel the fake limit orders before they are executed. Both spoofing and layering convey an impression that a security is more liquid than it actually is, or suggest that the security is currently under- or overpriced (Harris, 2015). Regardless, Cooper et al. (2016) claim that spoofing and layering is just another form of bluffing, and just like poker, bluffing is allowed. They conclude that the regulators should not treat all deception in the financial market as a misconduct and proposed a set of criteria in deciding which trading strategy should be regulated, and which should not.[11]

2.2.3 Quote-stuffing

Quote-stuffing is another form of a market manipulation strategy that utilizes HFT's ability to rapidly send and cancel orders. Easley et al. (2012) describe quote stuffing as a strategy that "involves sending and canceling massive numbers of orders with the intent of taking all available bandwidth and thereby preventing other traders from being able to submit orders" (p. 228). In similar notes, the U.K. Government Office for Science (2012, p. 168) defines quote-stuffing as "entering large numbers of orders and/or cancellations/updates to orders so as to create uncertainty for other participants, slowing down their process and to camouflage the manipulator's own strategy". The high rate of orders entering and canceling involves in quote stuffing is viewed as a way to manipulate markets, and luring other traders into making mistakes (Narang, 2013).

Unlike spoofing and layering that use limit order near the best bid and ask price, quote stuffing involves placing large amounts of nonexecutable orders – i.e. limit orders that are far from the best quote, aimed to congest the market and slow down other competitors (Lhabitant & Gregoriou, 2015). An exchange's network bandwidth might be congested from receiving unusually large numbers of trade messages (e.g. rapid orders and cancellations), thus impairing other traders' access to the market (Angel & McCabe, 2013). The impairment leaves the slower traders with an unclear picture of the actual market situation and affected their ability to execute trades. The faster traders on the other hand, able to get a better understanding of what is happening in the market, allowing them to profit at the expense of slower traders (Biais & Woolley, 2011). Since quote stuffing strategy seeks to make a profit by preventing others from adding their private information into the security, it lacks the criteria of an acceptable HFT strategy, and thus, should be prohibited (Cooper et al., 2016).

3.0 The effect of HFT on market quality

A large and growing body of literature has investigated the effect of HFT on market quality. The term "market quality" in itself is broadly defined, but it is commonly associated with price discovery and efficiency, liquidity, and volatility (e.g.: Harris, 2003; The U.K. Government Office for Science, 2012). Based on HFT's characteristics (see section 2.2 – Defining HFT), it can be thought as a new breed of

[11] Cooper et al. (2016) states that an acceptable trading strategy (1) should be prudent, (2) should not block price discovery, and (3) should not circumvent transparent price discovery.

intermediary, which may improve or harm the market. The issue of whether HFT is beneficial or detrimental to the market is a still a hot topic, debated among market participants, regulators, media, as well as academics (Menkveld, 2014). The many perspectives on HFT may have stemmed from the lack of consensus on the mechanics of HFT, which may act as market-makers, arbitrageurs, predators, or some combination (Carrion, 2013).

3.1.1 Price discovery

In his seminal paper, Fama (1970) states "a market in which prices always *fully reflect* available information is called *efficient*" (p. 383). The EMH postulates that in an informationally efficient market, security prices will adjust rapidly to the arrival of new information, and thus, the prevailing prices reflect all existing information about the security. There are three assumptions underlying the hypothesis: (1) an efficient market requires that a large number of profit-maximizing participants analyse and value securities, each independently of the others; (2) new information regarding securities comes to the market in a random fashion, and the timing of one announcement is generally independent of others; and (3) the competition between the many profit-maximizing investors to profit from the new information causes the security prices to adjust rapidly, and thus, the impact of new information is reflected in the security prices. Thus, the price changes are hypothesized to be independent and random and require a certain minimum amount of trading by the numerous competing investors in making the market more efficient.

EMH asserts that the existing securities prices in an efficient market should unbiased, and able to reflect all currently available information. Thus, should EMH holds, once an information is publicly disclosed, it is quickly reflected in prices (Fox, Glosten, & Rauterberg, 2017), and any mispricing and associated arbitrage opportunities should be rapidly eliminated (Goodhart & O'Hara, 1995). Furthermore, in the era of HFT, the term "immediately", "rapidly", or "current" need to be refined, as their (HFT) definition and perception on these terms are very much different than ordinary human traders. Comparatively, it takes 400–500 milliseconds for a human being to blink an eye, while HFT might have traded hundreds or thousands of times during a similar period (O'Hara, 2015).

There are two main functions of a financial market, i.e. to provide liquidity and to promote price discovery; in which both are vital for asset pricing. The process of incorporating new information into asset prices is known as price discovery, and together with liquidity, they play an important role for an efficient capital allocation in the economy. An efficient market allows individuals to reallocate their asset holdings, resulting in risk sharing among investors (O'Hara, 2003). The market is deemed as efficient when the price of a security fully reflects all currently available information about its economic value, both current and historical information. Since financial market is not naturally efficient, the market will move towards efficiency through price discovery (Cooper et al., 2016), and trading activities by informed traders, either through market or limit orders, will incorporate their private information about a security on its price (Cao, Hansch, & Wang, 2009). Therefore, the maximization of price discovery is seen as an important objective by regulators and academics alike (Cespa & Foucault, 2014).

In addition, as noted by Aldridge (2013), the process of impounding information from news to price is hardly instantaneous. The price will first swing due to the implicit "negotiation" among the many buyers and sellers which can be seen in the order flow before eventually finds its optimal post-announcement price range. This process is commonly referred to as tâtonnement – a French word for "trial and error". The price fluctuation gives HFT an opportunity to profit from the arbitraging surrounding news release and bring the market one step closer to its efficient state – as per EMH. Using directional event-based strategies, HFT will place its trades based on forecasted market reaction towards an event.

Froot, Scharfstein, and Stein (1992) show that in theory, short-term traders may bank on short-term information too much, and less concern on fundamentals value of a firm, which in turn, dampened market efficiency. Vives (1995) suggests that short-horizon traders reduce price informativeness with the concentrated arrival of information, which is likely to be the case around earnings news events. Zhang (2010) estimated the volume of HFT in the U.S. capital market for the year 2009 and find that HFT accounts for 78% of the total trading volume, which is very close to Tabb Group's estimate at 73%. He finds that HFT is positively correlated with price volatility even after controlling for stock's fundamentals and explanatory variables for volatility. The result is stronger especially in 3,000 largest stocks by market capitalization, in stocks with high institutional holdings, and during high market uncertainty periods.

Froot et al. (1992) and Zhang (2010) demonstrate that on account of their relative emphasis on the short-term horizon, the HFT firms hamper the price discovery process in the market. In fact, the HFT activities may cause the markets to be "too efficient" (overshooting fundamental values) and therefore, need to be restrained. Zhang (2010) shows that in the short run, HFT activity causes stock prices to move excessively in the direction of the news about fundamentals making it detrimental to the price discovery process. For instance, after positive fundamental news about a stock is released, HFT firms will rapidly enter a long position in the stock, and consequently, raising its price. At a later time, fundamental investors make their moves to buy the stock too, causing the stock price to rise more than the news about the fundamentals warranted, and thus, leads to "overshooting". Another reason for this phenomenon could be that HFT firms try to front run fundamental investors by anticipating the general direction of the subsequent trades. These firms will buy/sell the stock before the fundamental investors can do so, and when they (fundamental investors) eventually execute their trades it causes the price to move excessively.

3.1.2 Liquidity

A vast majority of the empirical study on HFT and automated trading find a positive influence on the market quality, in the sense that it reduces the bid-ask spreads, improves market liquidity, and makes stock prices more efficient (Jones, 2013). Hasbrouck and Saar (2013) study the effect of low-latency activities on market quality using the NASDAQ HFT dataset and find that an increase in HFT activities reduce quoted spreads, reduce price impact, increase depth, and lowers short-term volatility. They also test the relationship between normal and heightened uncertainty periods in the U.S. and find evidence that higher low-latency activities improve market quality in both periods. This is also consistent with Conrad, Wahal, and Xiang (2015) that uses the full cross-section of securities in the U.S. equity markets and three hundred largest stocks on the Tokyo Stock Exchange (TSE).[12,13] They find that high-frequency quotation activity not only has no detrimental effect on market quality but in fact, the presence of high-frequency quotes improves the efficiency of the price discovery process and reduce the trading costs. These findings are further supported by the evidence from Boehmer, Fong, and Wu (2015) that find co-location services facilitate HFT, which causally improves market quality.[14]

Market-making HFT provides liquidity by matching buyer and seller orders, or by buying and selling securities from their own inventories should they failed to immediately match buyers and sellers (Shorter & Miller, 2014). HFT that engage as market-maker use their speed advantage to quickly update

[12] Conrad et al. (2015) sample is from 2009 – 2011 for the U.S. markets, and from 2010 and 2011 for the TSE.
[13] The three hundred largest stocks are from the First Section of the TSE by beginning-of-month market capitalization.
[14] Boehmer et al. (2015) use data from 42 markets to study the effect co-location on AT and HFT. The first implementation date of co-location in each country is used to capture the effect on latency prompted by the co-location service.

quotes, and they profit from the difference between the price buyers are willing to pay and the ask prices sellers are willing to accept for a security. Since this activity requires HFT to maintain limit orders on both sides of the trades, it provides liquidity to the market (Chung & Lee, 2016). Hagstromer and Norden (2013) studied the event of changes in minimum tick size to examine the effect of HFT activities on market quality using 30 Swedish large-cap stocks traded on the NASDAQ-OMX Stockholm Exchange. Their findings suggest that HFT market-making activities reduce short-term volatility, which is healthy for the overall market quality. Similarly, Riordan and Storkenmaier (2012) study the effect of decreasing in latency on market quality following the release of Xetra 8.0 by Deutsche Boerse in 2007 find significant improvement in the market quality post upgrade, determined by narrower spread measures and higher relative quotes contribution to price discovery.[15]

Even though the increasing competition of market-making in general benefits the market, the fact that HFT does not have affirmative obligation to make market unlike the traditional market-maker or specialists, raised concern that they might cause disruptions by fleeing the market at their will (Carrion, 2013), e.g. when it is no longer profitable to do so (Anand & Venkataraman, 2013). The absence of the constraining obligations also gives HFT more flexibility to formulate market-making strategies beyond the traditional means (Brogaard et al., 2014). To gain more volume, certain trading venues offer liquidity rebates to market-making HFT, which benefits both the HFT and the exchanges themselves, as the HFT has the motivation to route the orders to their exchanges (Harris, 2015). The aim for such rebates is to encourage and reward the liquidity supply provided by the market-makers (The U.K. Government Office for Science, 2012). This is justified by the finding of Hendershott and Riordan (2013) which suggest that HFT market-makers would lose money in the absence of rebate. Similarly, Brogaard et al. (2014) find that HFT liquidity supplying revenues are negative without the fee rebates, especially during transactions with tighter spreads.

Theoretically, HFT could have both positive and negative effects on liquidity. The light-speed trading activity by HFT is claimed to promote liquidity through rapid price adjustments, allowing for narrower bid-ask spreads within a market, strengthening the inter-market linkage and activity (Goldstein et al., 2014), and lowering the cost of intermediation (Jones, 2013). However, the higher level of trading activity by HFT cannot simply be the indicator of better liquidity in the market, as the HFT could be in either side of the trades. A dominance in the supply side would lead to higher liquidity and narrower spread, while a greater number of trading activity in the demand side would take liquidity away from the market and widened spreads (Goldstein et al., 2014). For instance, CFTC-SEC (2010b) report suggest that even though HFT usually provide liquidity, during the Flash Crash, they turned to consume liquidity. Easley et al. (2011) suggest that the action produces toxic order flow and has exacerbated the ongoing liquidity crisis. This behavior of HFT has called for regulatory discussion and debate on whether to impose HFT with quotation obligation and/or prevent them from doing high-speed quotation entering/deleting (Gomber et al., 2011).

3.1.3 Adverse selection cost

Brogaard et al. (2014) find that HFT, in general, has a positive role in the price discovery process, especially contributing to the speed of price adjustment to new information, and smaller pricing errors. However, they also contest that even though the price informativeness is commonly viewed as something positive for the economy, the information that HFT used is short-lived, lasting for only 3-4 seconds. Should the information eventually become public without HFT's intermediation, the adverse

[15] The new trading platform was introduced with a sole reason to reduce the system latency, with no other meaningful microstructure change. Following the introduction, system latency is reduced from 50 millisecond to 10 millisecond.

selection costs that slower traders have to bear might cause the potential welfare gains from the faster price discovery becomes trivial or even negative.

Biais and Woolley (2011) argue that while the development of sophisticated and rapid trading algorithms may benefit the markets and investors through better price discovery and liquidity, they might be detrimental to the slower traders due to adverse selection problem. In a similar note, Biais, Foucault, and Moinas (2015) claim that even though the investment in fast trading does help to deal with the issue of market fragmentation, it also comes with the risk of adverse selection to the slow-traders, which lowers the social welfare. Scholtus, van Dijk, and Frijns (2014) find evidence of deterioration of market quality around the U.S. macroeconomic announcements. Using 60 seconds event window from the release of the news [0, 60], they find that higher algorithmic trading activity leads to lower depth, and higher quoted spreads, adverse selection costs, and volatility measures.[16]

In fact, some argue that it is the sheer speed of HFT that cause other slower investors bearing the cost of adverse selection (Jones, 2013). In a theoretical paper, Budish et al. (2015) develop a model in which market-makers or traders that invest in speed will be the first to react and make a profit from the newly arrived public information. In the event of the traders receive and react to a news before the market-makers do, they (the fast traders) will trade with the stale quotes, which impose adverse selection cost on the market-makers. This situation will discourage liquidity provision, and consequently, the market-makers include the cost of them being adversely selected in their quote, resulting in the wider spread and higher cost for other slower investors. This could be made possible due to HFT's speed and resources, which allow them to quickly process and take appropriate action whenever a new publicly available information arise. Slower traders on the hand, take a longer time to revise their orders, allowing HFT adversely select other participants' orders (Brogaard et al., 2014).

It is also possible for the algorithm to be fed with false information – either intentionally or accidentally. For instance, the United Airlines (UAL) stock price suddenly crash from US$12 to US$3 on September 8, 2008, in a mere 12 minutes, in which the shareholders lost (in value) of approximately US$1 billion (New York Times, 2008). An investigation later revealed that the rapid drop was mainly due to the interplay between algorithms that reacted to a six-year-old headline that mistakenly hit the news feed since human traders might not be deceived by the headline blunder (Donefer, 2010). Similarly, two weeks prior to the UAL's unfortunate event, on August 27, 2008, Bloomberg News accidentally published an obituary for Steve Jobs – the CEO of Apple (APPL) (Fortune.com, 2008). Luckily, the blunder happened during off-trading hours and was quickly retracted. Should it be otherwise, then Apple's stock price might suffer the same fate that befell United Airlines' stocks two weeks later (Donefer, 2010). This leads to another question – does immediacy of information dissemination always a good thing?

In a nutshell, scholars' understanding of the impact of HFT on market quality is still lacking due to its young literature and the lack of high-quality data (Carrion, 2013; Boehmer et al., 2015). However, in general, there is mixed empirical evidence on the impact of HFT on market quality. Despite the majority of empirical studies find positive effects of HFT's participation in the market, they cannot rule out the possibility that HFT, in theory, may harm the market through their predatory trading strategy (Manahov, Hudson, & Victor, 2014). Regulatory bodies around the world are either implemented or mulling over rules to contain and mitigate any HFT activity that may potentially detriment market quality (Benos & Sagade, 2016). It is agreed that any abusive or predatory trading activity which goes against market integrity should be eradicated. Nonetheless, regulators must be extra careful in formulating their arrangement to avoid any excessive regulations and constraints that may be

[16] To measure market quality, Scholtus et al. (2014) examine liquidity and volatility in the market. Liquidity is measured using depth, volume, and spread. Volatility is measured using two realized measures calculated over intervals of one (s = 60) and five (s = 300) minutes.

counterproductive and have unanticipated effects on market quality. For instance, the newly formulated regulation should not prevent beneficial HFT strategies that have positive effects on liquidity (e.g.: market-making strategies) or price discovery and market efficiency (e.g.: arbitrage strategies) from taking place (Gomber et al., 2011).

4.0 Controversies on HFT

This section is aimed to highlight the negative sentiment and controversies surrounding the HFT. The identified controversies are (1) the flash crash of May 6, 2010; (2) the economic welfare of the arms race; and (3) HFT's market-making obligation.

4.1.1 Flash crash of May 6, 2010

On May 6, 2010, the US financial markets were shocked with a short-lived, yet severe drop in prices, all happened within minutes. The sudden market crash of May 6, 2010 is later dubbed as the "flash crash", given the brief moment of the event. The U.S. Commodity Futures Trading Commission (CFTC) and U.S. Securities & Exchange Commission (SEC) released joint preliminary findings with regards to the event on May 18, 2010 (CFTC-SEC, 2010a), and full findings were released later on September 30, 2010 (CFTC-SEC, 2010b).

The US market opened on May 6 with unsettling political and economic issues surrounding the European debt crisis. The concern over the future direction of the European market has heightened the level of uncertainties in the US market, evidenced by high volatility, a flight to quality, and rise in premiums for buying protection against default by the Greek government on their sovereign debt. Consequently, the Euro experienced a sharp decline against the U.S. Dollar and Japanese Yen around midday. In the U.S., the financial market was shrouded by negative market sentiment, causing the S&P500 volatility index (VIX) to rise by 22.5 percent at around 2.30 p.m. (Central Time, CT) from its opening level. This has triggered investors to engage in flight to quality, created a selling pressure which has pushed down the Dow Jones Industrial Average (DJIA) by 2.5%.

At 2.32 p.m., Waddell & Reed (a large fundamental trader) initiated a sell order algorithm (Sell Algorithm) to sell 75,000 E-mini (S&P500 futures) contracts to hedge its existing equity position (Reuters, 2010; CFTC-SEC, 2010a). The Sell Algorithm is programmed to target an execution rate set to 9% of the trading volume calculated over the previous minute, without regard to price or time. Normally orders at such scale (valued at approximately US$4.1 billion) are fed in multiple stages to avoid shocks to the market, but apparently, this time it was not. Initially, the selling pressure was absorbed by HFT and other intermediaries in the futures market, followed by the fundamental buyer and cross-market arbitrageurs, in which the latter transferred the selling pressure to the equities market. Within 13 minutes of execution (between 2:32 p.m. and 2:45 p.m.), 35,000 E-mini contracts (valued at approximately US$1.9 billion) out of the intended 75,000 were sold.

From the Sell Algorithm order, HFT has accumulated a net long position of about 3,300 contracts.[17] Between 2:41 p.m. to 2:44 p.m., HFT aggressively sold about 2,200 E-mini contracts they held to reduce their inventories. Nearly 140,000 E-mini contracts (over 33% of total trading volume) were traded by HFT.[18] The dramatic increased in trading volume increased volatility in the market, which in turn shied long-term traders away from the market. The lack of demand in the market caused HFT to buy and sell from one another, generating a "hot-potato" volume effect. Enormous selling

---

[17] 16 out of over 15,000 trading accounts are classified as HFT, and traded over 1,455,000 contracts on May 6, equivalent to almost one-third of the total daily trading volume (CFTC-SEC, 2010b).

[18] This is consistent with the HFT' typical practice of trading a very large number of contracts, but not accumulating an aggregate inventory beyond three to four thousand contracts in either direction.

pressure from the combination of the Sell Algorithm, HFT, and other traders drove the price of the E-mini down by 3% within this 4-minutes period. At the same time, cross-market arbitrageurs who bought the E-mini simultaneously sold equivalent amounts in the equities markets, driving the price of S&P 500 SPDR (SPY) also down by approximately 3%.[19]

The combined selling pressure was so tremendous it almost wiped clean the entire buy-side orders of the E-mini, creating an order imbalance in the market. At that moment, there were less than 1,050 buy-side orders unmatched, and still, more than 50% of the Sell Algorithm's orders yet to be matched. This severe liquidity absence pushed the E-mini prices down by another 1.7% in a mere 15 seconds, reaching its intraday low of 1,056 points. At 2:45:28 p.m., the Chicago Mercantile Exchange's (CME) Stop Logic Functionality was triggered due to the rapid prices decline of the E-mini, causing all trading on the E-mini to be halted for five seconds. After the trading resumed at 2:45:33 p.m., the E-mini prices stabilized and starting to recover, thanks to opportunistic and longer-term traders who re-entered the market and rapidly accumulated long positions (Kirilenko et al., 2017). Subsequently, SPY also recovered.

Despite the E-mini recovering, the prices of other affected securities continued to decline. The sell orders placed on some individual securities and ETFs experienced reduced buying interest, mainly due to a high level of uncertainty among market participants in the market. Accordingly, some market-makers and other liquidity providers either widened their spreads and/or reduced offered liquidity, while others simply withdrew their position off the market. HFT in the equity markets traded proportionally more as volume increased, and overall were net sellers in the fast-declining market. Some of the HFT continued their trading and tap on the opportunities arose from the severe price dislocations in individual securities as the market started to recover, while some others just stopped trading completely.

There were approximately 2 billion shares with a total volume of more than US$56 billion traded between 2:40 p.m. and 3:00 p.m. on that day. During the 20-minutes window, more than 98% of all shares were traded within 10% of their value at 2:40 p.m. Due to the unusually high level of uncertainty in the market, orders sent to the market found no immediate interest, caused trades being executed at irrational prices. For instance, Accenture plc (ACN) rapidly declined in 7 seconds from about US$30 at 2:47:47 p.m., to US$0.01 by 2:47:54 p.m., and recovered within a matter of seconds. An ETF, iShares Russell 1000 Growth Index Fund's (IWF) share price plummeted from about US$45 just before 2:46 p.m. to the lowest price of US$0.0001 at 2:47:28 p.m., and slowly recovered to its prior level by 2:56 p.m. On the contrary, Sotheby's (BID) was traded at an extremely high price of US$99,999.9999 at 2:57:08 p.m., from around US$30 only minutes before that (CFTC-SEC, 2010a). These extreme cases were caused by orders executed against stub quotes, which was triggered due to the sudden loss of liquidity during the flash crash (Gomber et al., 2011).[20]

Overall, over 20,000 trades (amounting to 5.5 million shares) across 300 separate securities and ETFs have executed at prices 60% or more away from their 2:40 p.m. prices. By 3:00 p.m., prices for most securities had reverted back to trading at their rational values. After the market closed, the SEC and FINRA have met and agreed to adopt the “clearly erroneous" trade rules, and thus all trades

[19] The E-mini and SPY are the two most active stock index instruments traded in the electronic futures and equity markets. Both are derivative products designed to track stocks in the S&P 500 Index, which in turn represents approximately 75% of the market capitalization of U.S.-listed equities. Since the E-mini and SPY both track the same set of S&P 500 stocks, cross-market arbitrage between these two products kept their prices closely aligned during their rapid declines.

[20] Stub quotes are quotes generated by market-makers at levels far away from the current market in order to comply with its obligation to maintain a continuous two-sided quoting obligations. However, the stub quotes are not intended to be executed (CFTC-SEC, 2010a).

classified as "clearly erroneous" were canceled (broken).[21,22] Almost two-thirds of shares in the broken trades were executed at prices of less than US$1.00, and approximately five percent were executed at prices of greater than US$100 (CFTC-SEC, 2010b). From the joint report, it is evident that HFT did not trigger the Flash Crash. However, the repeated buying and selling of contracts executed by the automated systems created the hot-potato effect as HFT competed for liquidity. Thus, their trading behavior during the unusually large selling pressure on May 6 is perceived to have exacerbated the price decline and market volatility (Kirilenko et al., 2017). Due to this event, HFT has received considerable critical attention from both the CFTC and SEC for creating "excessive" short-term volatility (CFTC-SEC, 2010b, 36-37).

4.1.2 HFT arms race and social welfare issues

HFT contribution in the process of price discovery is indeed beneficial, as more informative stock prices might lead to better resource allocation in the economy. Nonetheless, Brogaard et al. (2014) find that the information used by HFT are short-lived, lasted for less than 3 to 4 seconds. Should the information will eventually become public without HFT' intermediation, then the potential welfare contribution by HFT might be minuscule, or even negative in the situation where longer-term investors are significantly affected by the adverse selection costs from trading with HFT. In a similar note, Menkveld (2014) agree that the presence of market-making HFT in electronic markets does improve welfare by reducing informational frictions from non-simultaneous orders arrival in the market. However, the net welfare from HFT is questionable – the positive contribution from market-making activity might be destroyed when HFT pick off investors' quotes at lightning speed on information that will surely arrive at the slower investors at a lower frequency.

HFT acknowledge the importance of investing in hardware, software and network capabilities to reduce latency in an automated trading process, motivated by the nature of the game where winner-takes-all. The upgrades allow them to continuously refine their trading algorithms, and emerge victorious in the arms race (Kauffman, Liu, & Ma, 2017). Regardless, the technology arms race to shave-off several seconds raised concerns about the excessive spending of money without meaningful progress in market quality (Chung & Lee, 2016). The race among institutions to be the fastest is deemed as unproductive, and the unwarranted investments in technological infrastructure to reduce trading latency creates doubts of whether HFT adds value overall (Chordia et al., 2013; Jones, 2013). In addition, Menkveld (2014) asserts that the technology investment itself may as well be the source of negative externality through the relative speed disadvantage it creates for others.

From another point of view, Budish et al. (2015) claim that the arms race is indeed socially wasteful, but their existence is actually a symptom, stemmed from a flaw in the architecture of modern financial exchanges that use continuous-time trading, which also creates adverse selection rents that

---

[21] Under the "clearly erroneous" trade rules, the regulatory body may declare a trade to be null and void, should the trade in question was considered to be "clearly erroneous" (CFTC-SEC, 2010a). On September 10, 2010, the SEC approved new rules submitted by the national exchanges and FINRA that clarify the process for breaking erroneous trades (https://www.sec.gov/rules/sro/bats/2010/34-62886.pdf).

[22] Following the wide-scale disruption of May 6, 2010, the exchanges and FINRA settled on the relatively high 60% standard for breaking trades (CFTC-SEC, 2010a, 2010b); (1) For stocks priced US$25 or less, trades will be broken if the trades are at least 10% away from the circuit breaker trigger price; (2) For stocks priced more than US$25 to US$50, trades will be broken if they are 5% away from the circuit breaker trigger price; (3) For stocks priced more than US$50, the trades will be broken if they are 3% away from the circuit breaker trigger price; (4) Where circuit breakers are not applicable, the exchanges and FINRA will break trades at specified levels for events involving multiple stocks depending on how many stocks are involved; (5) For events involving between five and 20 stocks, trades will be broken that are at least 10% away from the "reference price," typically the last sale before pricing was disrupted; and (6) For events involving more than 20 stocks, trades will be broken that are at least 30% away from the reference price.

attract HFT. Budish et al. (2015) suggest that the problem can be addressed using a frequent batch auction, which will create a discrete-time market to replace the current market design that is based on the continuous limit order book. This will make the tiny speed advantage less valuable, which intuitively put an end to the arms race. In a similar notion, Yao and Ye (*forthcoming*) find evidence that even with discrete timing, HFT might continue to race each other – this time to compete for rents from the queuing channel, originated from yet another microstructure design – tick size. Either way, both types of rents are lucrative by-products of market's imperfections and can be dominated by being the fastest, which leads to an arms race in speed.

Regardless, even without the issue of arms race, HFT still pose a threat to many as they may use high-speed predatory trading strategies (see section 2.6 – Detrimental HFT strategies), such as introducing "microstructure noise" that generates an unnecessary extra layer of intermediation between buyers and sellers, leading to increased price volatility and worsened market quality (Cartea & Penalva, 2012).

4.1.3 Market-making obligations

Anand and Venkataraman (2013) study the trades of two types of market-makers, the Designated Market-makers (DMMs) and Endogenous Liquidity Providers (ELPs). The main difference between DMMs and ELPs lies in their obligation to make a market. DMM or Specialists are bounded by specific obligations imposed by the exchange, i.e. to maintain a market presence by continuously posting quotes with reasonable depth. ELP on the other hand, employs market-making strategies because of its profitability, with no affirmative obligations to maintain markets. Anand and Venkataraman (2013) states the HFT are the most active market-makers in financial markets today, in which some position themselves as ELP – meaning that they are likely to supply liquidity whenever it is profitable for them to do so and cease from providing liquidity when facing large adverse selection risks (Chung & Chuwonganant, 2018), or whenever the market conditions are unfavourable for them to make profits, which is more likely to happen in times of high market uncertainty (Zhang, 2010).

The lack of commitment to make market especially in times of market stress and in thinly traded securities raised concern among practitioners and regulators. HFT's optional market-making may exacerbate execution uncertainty, and thus, the liquidity supplied by HFT are deemed unreliable, which might reduce investors' confidence and participation. Liquidity withdrawal by HFT might thin out the order book, which may induce extreme market movements (Gomber et al., 2011). This might also be the underlying reason for the heightened sensitivity of liquidity and returns to market volatility in the era of HFT. Furthermore, the non-HFT are playing at an uneven playing field due to their technological inferiority to HFT, and they might find that the market is unfair, and consequently, stop participating altogether (Anand & Venkataraman, 2013). In response to this potential problem, regulators consider imposing quotation obligations on HFT, and/or preventing them from engaging in high-speed order entering and cancellation (Gomber et al., 2011).

Bibliography

Agarwal, A. (2012). *High frequency trading: Evolution and the future.* London, UK.

Aldridge, I. (2013). *High-frequency trading: A practical guide to algorithmic.* New Jersey: John Wiley & Sons, Inc.

Anand, A., & Venkataraman, K. (2013). Should exchanges impose market maker obligations? *Working paper.*

Angel, J. J. (2014). When finance meets physics: The impact of the speed of light on financial markets and their regulation. *The Financial Review, 49*, 271–281.

Angel, J. J., & McCabe, D. (2013). Fairness in financial markets: The case of high frequency trading. *Journal of Business Ethics*, *112*(4), 585–595 |.

Aquilina, M., & Ysusi, C. (2016). Are high-frequency traders anticipating the order flow? Cross-venue evidence from the UK market. *Working paper*.

Australian Securities and Investments Commission. (2010). *Australian equity market structure.* Victoria.

Australian Securities and Investments Commission. (2013). *Dark liquidity and high-frequency trading.* Report 331.

Australian Securities and Investments Commission. (2015). *Review of high-frequency trading and dark liquidity.* Report 452.

Avramovic, A., Lin, V., & Krishnan, M. (2017). *We're all high frequency traders now.* Credit Suisse.

Benos, E., & Sagade, S. (2016). Price discovery and the cross-section. *Journal of Financial Markets, 30*, 54–77.

Biais, B., & Woolley, P. (2011). High Frequency Trading. *Working paper*.

Biais, B., Foucault, T., & Moinas, S. (2015). Equilibrium fast trading. *Journal of Financial Economics, 116*, 292–313.

Boehmer, E., Fong, K., & Wu, J. (2015). International evidence of algorithmic trading. *Working paper*.

Brogaard, J. (2010). High frequency trading and its impact on market quality. *Northwestern University Kellogg School of Management Working Paper*, 66.

Brogaard, J., Hendershott, T., & Riordan, R. (2014). High-frequency trading and price discovery. *The Review of Financial Studies*, *27*(8), 2267–2306.

Budish, E., Cramton, P., & Shim, J. (2015). The high-frequency trading arms race: Frequent batch auctions as a market design response. *Quarterly Journal of Economics, 130*(4), 1547–1621.

Cao, C., Hansch, O., & Wang, X. (2009). The information content of an open limit-order book. *Journal of Futures Markets, 29*(1), 16-41.

Carrion, A. (2013). Very fast money: High-frequency trading on the NASDAQ. *Journal of Financial Markets*, *16*, 680–711.

Cartea, A., & Penalva, J. (2012). Where is the Value in High Frequency Trading? *Quarterly Journal of Finance*, *2*(3), 1-46.

Cespa, G., & Foucault, T. (2014). Sale of Price Information by Exchanges: Does It Promote Price Discovery? *Management Science, 60*(1), 148-165.

Chordia, T., Goyal, A., Lehmann, B. N., & Saar, G. (2013). High-frequency trading. *Journal of Financial Markets, 16*, 637-645.

Chung, K. H., & Chuwonganant, C. (2018). Market volatility and stock returns: The role of liquidity providers. *Journal of Financial Markets, 37*, 17-34.

Chung, K. H., & Lee, A. J. (2016). High-frequency trading: Review of the literature and regulatory initiatives around the world. *Asia-Pacific Journal of Financial Studies, 45*, 7–33.

Committee of European Securities Regulators. (2010). *Call for evidence on micro-structural issues of the European equity markets.* Paris.

Conrad, J., Wahal, S., & Xiang, J. (2015). High-frequency quoting, trading, and the efficiency of prices. *Journal of Financial Economics, 116*, 271-291.

Cooper, R., Davis, M., & Vliet, B. V. (2016). The mysterious ethics of high-frequency trading. *Business Ethics Quarterly, 26*, 1-22.

Donefer, B. S. (2010). Algos gone wild: Risk in the world of automated trading strategies. *Journal of Trading*, *5*(2), 31-34.

Durbin, M. (2010). *All about high-frequency trading: The easy way to get started.* New York: McGraw-Hill Companies, Inc.

Easley, D., Prado, M. M., & O'Hara, M. (2011). The microstructure of the "Flash Crash": Flow toxicity, liquidity crashes, and the probability of informed trading. *Jounal of Portfolio Management, 37*(2), 119-128.

Easley, D., Prado, M. M., & O'Hara, M. (2012). The volume clock: Insights into the high-frequency paradigm. *Journal of portfolio management*, 19-29.

European Securities and Markets Authority. (2011). *Guidelines on systems and controls in a highly automated trading environment for trading platforms, investment firms and competent authorites.* Paris.

Financial Industry Regulatory Authority. (2012). *FINRA Joins Exchanges and the SEC in fining Hold Brothers more than $5.9 million for manipulative trading, anti-money laundering, and other violations.* Washington, D.C.

Foucault, T., Hombert, J., & Rosu, I. (2016). News trading and speed. *Journal of Finance, 71*(1), 335–382.

Froot, K. A., Scharfstein, D. S., & Stein, J. C. (1992). Herd on the street: Informational inefficiencies in a market with short-term speculation. *Journal of Finance, 47*(4), 1461-1484.

Goldstein, M. A., Kumar, P., & Graves, F. C. (2014). Computerized and high-frequency trading. *Financial Review, 49*, 177–202.

Golub, A., Dupuis, A., & Olsen, R. B. (2013). High-frequency trading in FX markets. In D. Easley, M. L. Prado, & M. O'Hara, *High-Frequency Trading New Realities for Traders, Markets and Regulators* (pp. 65-90). London: Risk Books.

Gomber, P., Arndt, B., Lutat, M., & Uhle, T. (2011). High-frequency trading. Available at SSRN: https://ssrn.com/abstract=1858626 or http://dx.doi.org/10.2139/ssrn.1858626

Goodhart, C. A., & O'Hara, M. (1997). High frequency data in financial markets: Issues and applications. *Journal of Empirical Finance*, *4*(2-3), 73-114.

Hagströmer, B., & Nordén, L. (2013). The diversity of high-frequency traders. *Journal of Financial Markets*, *16*, 741–770.

Harris, L. (2003). *Trading and exchanges: Market microstructure for practitioners.* New York: Oxford University Press.

Harris, L. (2013). What to do about high-frequency. *Financial Analysts Journal, 69*(2), 6-9.

Harris, L. (2015). *Trading and Electronic Markets: What Investment Professionals Need to Know.* Research Foundation of CFA Institute.

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

Hendershott, T., & Riordan, R. (2013). Algorithmic trading and the market for liquidity. *Journal of Financial and Quantitative Analysis, 48*(4), 1001–1024.

Jarnecic, E., & Snape, M. (2014). The provision of liquidity by high-frequency participants. *The Financial Review, 49,* 371-394.

Jones, C. M. (2013, March 20). What do we know about high-frequency trading? *Columbia Business School Research Paper No. 13-11*.

Kauffman, R. J., Hu, Y., & Ma, D. (2015). Will high-frequency trading practices transform the financial markets in the Asia Pacific Region? *Financial Innovation, 1*, 1-27.

Kauffman, R. J., Liu, J., & Ma, D. (2015). Innovations in financial IS and technology ecosystems: High-frequency trading in the equity market. *Technological Forecasting and Social Change, 99,* 339–354.

Kaya, O. (2016). *High-frequency trading reaching the limits.* Deutsche Bank Research.

Kirilenko, A., Kyle, A. S., Samadi, M., & Tuzun, T. (2017). The flash crash: High-frequency trading. *The Journal of Finance, 72*(3), 967–998.

Kosinski, R. J. (2013). A literature review on reaction time. *Working paper*.

Kozhan, R., & Tham, W. W. (2012). Execution risk in high-frequency arbitrage. *Management Science, 58*(11), 2131-2149.

Lhabitant, F.-S., & Gregoriou, G. N. (2015). High-frequency trading: Past, Present and future. In G. N. Gregoriou, *Handbook of high frequency trading* (pp. 155-165). Academic Press.

Manahov, V., Hudson, R., & Viktor, B. G. (2014). Does high frequency trading affect technical analysis and market efficiency? And if so, how? *Journal of International Financial Market, Institutions and Money, 28,* 131– 157.

Markham, J. W. (2002). *A financial history of the United States. 1. From Christopher Columbus to the Robber Barons (1492 -1900).* New York: M. E. Sharpe.

Menkveld, A. J. (2014). High-frequency traders and market structure. *The Financial Review, 49*, 333-344.

Moosa, I., & Ramiah, V. (2015). The profitability of high-frequency trading: Is it for real? In *The Handbook of high-frequency trading* (pp. 25-45). Academic Press.

Narang, R. K. (2013). *Inside the black box: A Simple Guide to Quantitative* (Second Edition ed.). New Jersey: John Wiley & Sons, Inc.

Netherlands Authority for the Financial Markets. (2010). *High frequency trading: The application of advanced trading technology in the European marketplace.* Amsterdam.

O'Hara, M. (2015). High frequency market microstructure. *Journal of Financial Economics, 116*, 257–270.

O'Hara, M. (2003). Presidential Address: Liquidity and price discovery. *Journal of Finance, 58*(4), 1335-1354.

Riordan, R., & Storkenmaier, A. (2012). Latency, liquidity and price discovery. *Journal of Financial Markets, 15*, 416-437.

Scholtus, M. n., Dijk, D. v., & Frijns, B. (2014). Speed, algorithmic trading, and market quality around macroeconomic news announcements. *Journal of Banking and Finance, 38,* 89-105.

Securities and Exchange Commission. (2010). Concept release on equity market structure. *Federal Register*, *75*(13), 3594-3614.

Securities and Exchange Commission. (2014). Equity market structure literature review Part II: High frequency trading. *Staff of the Division of Trading and Markets*.

Serbera, J.-P., & Paumard, P. (2016). The fall of high-frequency trading: A survey of competition and profits. *Research in International Business and Finance, 36*, 271–287.

Shorter, G., & Miller, R. S. (2014). *High-frequency trading: Background, concern and regulatory developement.* Congressional Research Service.

The U.K. Government Office for Science. (2012). *Foresight: The future of computer trading in financial markets - Final Project Report.* London.

U.S. Commodity Futures Trading Commission and U.S. Securities & Exchange Commission. (2010a). *Preliminary findings regarding the market events of May 6, 2010.* Washington, D.C.

U.S. Commodity Futures Trading Commission and U.S. Securities & Exchange Commission. (2010b). *Findings regarding the market events of May 6, 2010.* Washington, D.C.

Vives, X. (1995). Short-term investment and the information efficiency of the market. *Review of Financial Studies, 8*(1), 125-160.

Wissner-Gross, A. D., & Freer, C. E. (2010). Relativistic statistical arbitrage. *Physical Review, 82*(5), 1-7.

Yao, C., & Ye, M. (2018). Why trading speed matters: A tale of queue rationing under price controls. *The Review of Financial Studies, 31*(6), 2157-2183.

Zhang, X. F. (2010). High-frequency trading, stock volatility, and price discovery. *Working paper*.